Spatio-temporal transcriptomic divergence across human and macaque brain development

Ying Zhu[1,2,†], André M. M. Sousa[1,†], Tianliuyun Gao[1,†], Mario Skarica[1,†], Mingfeng Li[1,†], Gabriel Santpere[1], Paula Esteller-Cucala[3], David Juan[3], Luis Ferrández-Peral[3], Forrest O. Gulden[1], Mo Yang[1], Daniel J. Miller[1], Tomas Marques-Bonet[3,4,5,6], Yuka Imamura Kawasawa[7], Hongyu Zhao[2], Nenad Sestan[1,8,*]

Introduction: Better understanding of how developing human nervous system differs from that of closely related non-human primates is fundamental for teasing out human-specific aspects of behavior, cognition, and disorders.

Rationale: The shared and unique functional properties of the human nervous system are rooted in the complex transcriptional programs governing the development of distinct cell types, neural circuits, and regions. However, the precise molecular mechanisms underlying shared and unique features of the developing human nervous system have been only minimally characterized.

Results: We generated complementary tissue-level and single-cell transcriptomic datasets from up to sixteen brain regions covering prenatal and postnatal development in human and rhesus macaque (*Macaca mulatta*), a closely related species and the most commonly studied non-human primate. We created and applied *TranscriptomeAge* and *TempShift* algorithms to age-match developing specimens between the species and more rigorously identify temporal differences in gene expression within and across the species.

By analyzing regional and temporal patterns of gene expression in both the developing human and macaque brain, and comparing these patterns to a complementary dataset that included transcriptomic information from the adult chimpanzee, we identified shared and divergent transcriptomic features of human brain development. Furthermore, integration with single-cell and single-nucleus transcriptomic data covering prenatal and adult periods of both species revealed that the developmental divergence between human and macaque can be traced to distinct cell types enriched in different developmental times and brain regions, including the prefrontal cortex, a region of the brain associated with distinctly human aspects of cognition and behavior.

We found two phases of prominent species differences: embryonic to late midfetal development and adolescence/young adulthood. This evolutionary cup-shaped or hourglass-like pattern with highly divergent prenatal development and adolescence/young adulthood and lower divergence in early postnatal development resembles the developmental cup-shaped pattern described in the accompanying study by Li et al. Even though the developmental (ontogenetic) and evolutionary (phylogenetic) patterns have similar profiles, the overlap of genes driving these two patterns is not substantial, indicating the existence of different molecular mechanisms and constraints for regional specification and species divergence.

Notably, we also identified numerous genes and gene co-expression modules exhibiting human-distinct patterns in either temporal (heterochronic) or spatial (heterotopic) gene expression, as well as genes with human-distinct developmental expression, linked to autism spectrum disorder, schizophrenia, and other neurological or

psychiatric diseases or traits, potentially suggesting mechanistic underpinnings of these disorders and traits.

Conclusion: Our study reveals novel insights into the evolution of gene expression in the developing human brain and may shed some light on potentially human-specific underpinnings of certain neuropsychiatric disorders.

Figure legend: Concerted ontogenetic and phylogenetic transcriptomic divergence in human and macaque brain. (Left) Human and macaque brain regions spanning both prenatal and postnatal development were age-matched using a novel algorithm. (Right) Phylogenetic transcriptomic divergence between human and macaque resembles the developmental (ontogenetic) cup-shaped pattern of each species, with highly divergent prenatal development and adolescence/young adulthood and lower early postnatal (perinatal to adolescence) divergence. Single-cell transcriptomics revealed shared and divergent transcriptomic features of distinct cell types.

Title:

Spatio-temporal transcriptomic divergence across human and macaque brain development

Short Title:

Evolution of human brain development

One Sentence Summary:

Spatio-temporal human and macaque brain transcriptomes display concerted ontogenetic and phylogenetic cup-shaped divergence patterns.

Authors: Ying Zhu[1,2,†], André M. M. Sousa[1,†], Tianliuyun Gao[1,†], Mario Skarica[1,†], Mingfeng Li[†], Gabriel Santpere[1], Paula Esteller-Cucala[3], David Juan[3], Luis Ferrández-Peral[3], Mo Yang[1], Daniel J. Miller[1], Tomas Marques-Bonet[3,4,5,6], Yuka Imamura Kawasawa[7], Hongyu Zhao[2], Nenad Sestan[1,8,*]

Affiliations:

[1]Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT, USA
[2]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
[3]Institute of Evolutionary Biology (UPF-CSIC), PRBB, Barcelona, Spain
[4]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain
[5]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain
[6]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain
[7]Departments of Pharmacology and Biochemistry and Molecular Biology, Institute for Personalized Medicine, Penn State University College of Medicine, Hershey, Pennsylvania, USA
[8]Departments of Genetics, of Psychiatry, and of Comparative Medicine, Program in Cellular Neuroscience, Neurodegeneration and Repair, and Yale Child Study Center, Yale School of Medicine, New Haven, CT, USA

*Correspondence to: Nenad Sestan (nenad.sestan@yale.edu)

†These authors contributed equally to this work.

Abstract: Human nervous system development is an intricate and protracted process that requires precise spatio-temporal transcriptional regulation. Here we generated tissue-level and single-cell transcriptomic data from up to sixteen brain regions covering prenatal and postnatal rhesus macaque development. Integrative analysis with complementary human data revealed that global intra-species (ontogenetic) and inter-species (phylogenetic) regional transcriptomic differences exhibit concerted cup-shaped patterns, with a late fetal-to-infancy (perinatal) convergence. Prenatal neocortical transcriptomic patterns revealed transient topographic gradients, whereas postnatal patterns largely reflected functional hierarchy. Genes exhibiting heterotopic and heterochronic divergence included those transiently enriched in the prenatal prefrontal cortex or linked to autism spectrum disorder and schizophrenia. Our findings shed light on transcriptomic programs underlying the evolution of human brain development and the pathogenesis of neuropsychiatric disorders.

Main Text:

Introduction

The development of the human nervous system is an intricate process that unfolds over a prolonged time course, ranging from years to decades, depending on the region (*1-6*). Precise spatial and temporal regulation of gene expression is crucial for all aspects of human nervous system development, evolution, and function (*6-13*). Consequently, alterations in this process have been linked to psychiatric and neurological disorders, some of which may exhibit primate or human-specific manifestations (*11, 14-18*). However, our ability to explain many aspects of human nervous system development and disorders at a mechanistic level has been limited by our evolutionary distance from genetically tractable model organisms, such as mouse (*15, 16, 19-22*), and by a lack of contextual and functional interpretations of polymorphisms and disease-associated variations in the human and non-human primate (NHP) genomes (*11, 17, 21, 23*). Moreover, neither the extent of molecular changes underlying human-specific differences nor the specific developmental programs affected by these changes have been thoroughly studied.

The rhesus macaque (*Macaca mulatta*) is the most widely studied NHP in neuroscience and medicine (*24-26*). The macaque nervous system parallels the human nervous system with its complex cellular architecture and extended development, providing a unique opportunity to study features of neurodevelopment that are shared and divergent between the two closely related primates. Furthermore, studies of postmortem NHP tissues provide a unique opportunity to validate results obtained using postmortem human tissue, especially those from critical developmental periods that can be confounded by ante-mortem and post-mortem factors and tissue quality. Finally,

significant advances in transgenic and genome editing technologies now allow the possibility to create more precise genetic models for human disorders in macaque (*24-26*). This will facilitate the interrogation of the effects of specific gene mutations in a model that is closer to the human brain than any other experimental animal.

Comparative transcriptomic profiling offers unbiased insight into conserved and clade- or species-specific molecular programs underlying cellular and functional development of the human nervous system (*27-31*). However, a systematic characterization of the spatial and temporal transcriptomic landscapes of the macaque brain at the region-specific and single-cell levels, as well as the identification of shared and divergence features between human and macaque, are lacking. Such data and analyses presented here should provide both retrospective and prospective benefits to the fields of neuroscience, evolutionary biology, genomics, and medicine.

Study design and data generation

Here, we describe the generation and integrated analysis of RNA sequencing (RNA-Seq) data obtained using bulk tissue (366 samples from 26 prenatal and postnatal brains) or single cells/nuclei (113,274 cells/nuclei from 2 fetal and 3 adult brains) from post-mortem rhesus macaque specimens. Both tissue and single cell/nucleus datasets were subjected to multiple quality control measures (figs. S1-6; tables S1-2; Methods (*32*)). Tissue-level samples covered the entire span of both prenatal and postnatal neurodevelopment (Fig. 1A and B; table S1) and included 11 areas of the cerebral neocortex (NCX), hippocampus (HIP), amygdala (AMY), striatum (STR), mediodorsal nucleus of thalamus (MD), and cerebellar cortex (CBC). Subject ages ranged from 60

postconception days (PCD) to 11 postnatal years (PY) and were matched by age and brain region to 36 human brains from an accompanying study (*33*) and 5 adult chimpanzee brains from a previous study (*34*) (Fig. 1A). To investigate the contribution of different factors to the global transcriptome dynamics, we applied unsupervised clustering and principal component analysis, which revealed that age, species, and regions contribute more to the global transcriptomic differences than other tested variables (figs. S3 and S4).

To explore cell-type origins of tissue-level inter-species difference, we conducted single-cell RNA-Seq (scRNA-Seq) on 86,341 cells from six matching regions of two 110 PCD fetal macaque brains (i.e., the dorsolateral prefrontal neocortex [DFC, aka DLPFC], HIP, AMY, STR, MD, and CBC) and single-nucleus RNA-Seq (snRNA-Seq) of 26,933 nuclei from three adult macaque DFC (8, 11 and 11 PY; tables S2-3; (*32*)). These data were complemented by 17,093 snRNA-Seq samples from the adult human (see (*33*)) as well as two scRNA-Seq datasets from embryonic and fetal human neocortex (*33, 35*). In the six fetal macaque brain regions, we identified 129 transcriptomically distinct clusters of cell types (i.e., 19 in DFC, 20 in HIP, 25 in AMY, 22 in STR, 20 in MD, and 23 in CBC) (figs. S7-12; tables S3-4). In the adult human DFC (fig. S13) and adult macaque DFC (fig. S14), we identified 29 and 21 transcriptomically distinct cell types, respectively (table S3, S5, and S6). Alignment of our macaque fetal data with the adult single-nuclei data revealed hierarchical relationships and similarities between major cell classes, reflecting their ontogenetic origins and functional properties (fig. S15). Cell clusters were categorized by their gene expression patterns and assigned identities commensurate with their predicted cell type and, in the case of human adult neocortical

excitatory neurons, their putative laminar identity. While the majority of cell clusters were composed of cells derived from all brains, we found a few clusters in subcortical regions (AMY: 2 out of 25 clusters; CBC: 1 out of 23 clusters; and STR: 1 out of 22 clusters) that included cells from a single donor brain. This might be due to variations in dissection, age (even though both fetal macaques were 110 PCD, a 3 to 4 days variation remains), individual differences, and other technical bias. We used the single-cell datasets in this and the accompanying study (*33*) to deconvolve tissue-level RNA-Seq data, identify temporal changes in cell type specific signatures, analyze differences in cell types and their transcriptomic profiles, and conduct cell type enrichment analyses.

Similarities and differences in the spatio-temporal dynamics of the human and macaque brain transcriptomes

Unsupervised hierarchical clustering and principal component analysis of bulk tissue revealed common principles of transcriptomic regional architecture across development in macaque and human (figs. S3 and S4). Among macaque regions, these analyses showed distinct and developmentally regulated clustering of NCX (combination of 11 areas), HIP, and AMY, with CBC exhibiting the most distinctive transcriptional profile, an observation shared with our complementary study in humans (*27, 29, 30, 33, 36*). A hierarchical clustering of both fetal and postnatal NCX areal samples revealed their grouping by topographical proximity and functional overlap, similar to those relationships we observed in human brain (fig. S3). Thus, these results show that the transcriptomic architecture of the macaque brain is regionally and temporally specified

and reflects conserved global patterns of ontogenetic and functional differences that are also found in humans.

To explore species similarities and differences in the spatio-temporal dynamics of the brain transcriptome, we used the *XSAnno* computational framework (*37*) to minimize biases in comparative data analyses due to the disparate quality of gene annotation for the two species. We created common annotation sets of 27,932 and 26,514 orthologous protein coding and non-coding mRNA genes for human-macaque and human-chimpanzee-macaque comparisons, respectively (fig. S2 and (*32*)). Next, we developed *TranscriptomeAge*, an algorithm to unbiasedly predict the equivalent ages of human and macaque samples on the basis of temporal transcriptomic changes (*32*). We chose to optimize this model for age-matching the aforementioned eleven neocortical areas, which are highly similar in terms of their transcriptomes, cellular composition, and developmental trajectories when compared to other brain regions (see (*33*)). *TranscriptomeAge* confirmed transcriptomic similarities in both species coinciding with major prenatal and postnatal developmental phases, including fetal development, infancy, childhood, and adulthood (Figs. 1, A and B, S16-18). However, we identified two human developmental periods where alignment suggested that they are transcriptomically distinct from macaque and/or especially protracted. First, 60 PCD macaque specimens (which correspond to human early fetal period (*29*) according to *Translating Time* model (*38*)), most closely aligned with mid-fetal human samples (102 - 115 PCD or 14.5 - 16.5 PCW). This suggests that, transcriptomically, human brain development is protracted even at early fetal periods. Second, we found that 2, 3.5, 4, 5, and 7 PY macaque specimens, of which at least the youngest should chronologically match to human

childhood (*39*), did not align with any of our human specimens from early or late childhood (1 -12 PY or periods 9 and 10 according to (*29*)), but aligned with adolescent and adult human (Fig. 1A and B), indicating that, consistent with previous morphophysiological and behavioral studies (*5*), macaques lack global transcriptomic signatures of late childhood and/or that humans have a prolonged childhood relative to macaque (Fig. 1A and B).

Species differences in the timing of concerted neurodevelopmental processes

We hypothesized that the observed developmental differences between humans and macaques might be grounded on transcriptomic changes in concerted biological processes in developmental timing (i.e., heterochrony). By decomposing the gene expression matrix of human neocortical samples, we identified five transcriptomic signatures underlying neocortical development (*32*). Using top cell type-specific genes derived from our prenatal single-cell and adult single-nucleus data, we analyzed cell-type enrichment of each of the five signatures, and ascribed them to neurogenesis, neuronal differentiation, astrogliogenesis, synaptogenesis, and oligodendrocyte differentiation and myelination (Figs. 1C and D, S19). To determine whether the transcriptomic signatures we identified were correctly assigned, we compared their developmental patterns to the timing of major human neurodevelopmental processes, expression trajectories of key genes previously implicated in those processes, and to the trajectories of cell-type proportions identified by the deconvolution of tissue level data (figs. S19-20). We found that the developmental trajectories of genes associated with neuronal differentiation, synaptogenesis, and myelination, as well as the cell-type proportions of fetal human or macaque excitatory

neurons, astrocytes, and oligodendrocytes, matched those of the corresponding transcriptomic signatures (fig. S20). Moreover, comparison of transcriptomic signatures to independently generated non-transcriptomic data predicting the start and end of human neocortical neurogenesis (for neurogenesis) (*40*), and data measuring the number of DCX-immunopositive nascent neurons in the human hippocampus throughout development and adulthood (for neuronal migration and initial differentiation) (*41*), the developmental variation in synaptic density in the human cortex (for synaptogenesis) (*42*), and the myelinated fiber length density (for myelination) (*43*) also confirmed the identities we assigned to these transcriptomic signatures (fig. S19).

Next, we analyzed how the shape of the five transcriptomic trajectories was conserved across the eleven neocortical areas within each species and between species. Analysis of their trajectories within each species revealed that the shape of a given trajectory is similar across neocortical areas (Fig. 1C and S17). However, the transcriptomic trajectories associated with oligodendrocyte differentiation and myelination exhibited a prominent temporal shift (asynchrony) across neocortical areas in both species (fig. S17). Between species, myelination and, to a lesser extent, synaptogenesis exhibited species differences in the shapes of these trajectories; the myelination transcriptomic signature progressively increases in the human neocortex beginning from late fetal development through adulthood without reaching an obvious plateau until 40 PY, but in the macaque neocortex the myelination signature reaches a plateau around the first postnatal year (Fig. 1C). This corresponds to early childhood in human neurodevelopment (window 6 or period 10 according to (*33*) or (*29*), respectively) and is consistent with histological studies and reflective of previously reported

hierarchical maturation of neocortical areas (*43-47*). Similarly, we corroborated synchronous or concurrent transcriptomic patterns of neocortical synaptogenesis by analyzing previously collected data on synaptic density in multiple areas of the macaque neocortex (*48*) (Fig. S19). However, we observed that the synaptogenesis transcriptomic trajectory peaked earlier in human than macaque, at the transition between late infancy and early childhood (Fig. 1C). In addition, expression trajectories of genes induced by neuronal activity, a process critical for synaptogenesis, also showed dramatic increases during late fetal development and infancy, and like the synaptogenesis trajectory displayed a concurrent or synchronous shape across neocortical areas (see (*33*)). Interestingly, the developmental transcriptomic profile of *DCX*, a marker of nascent, migrating neurons, showed that macaques maintain higher expression in the hippocampus throughout postnatal development and adulthood, suggesting that postnatal neurogenesis is more prominent in macaque hippocampus than in human, as recently shown (fig. S19) (*49*). Thus, both species exhibited distinct transcriptomic signatures of neoteny, such as prolonged myelination in humans and prolonged postnatal hippocampal neurogenesis in macaques. Together, these data suggest the temporal staging of major neurodevelopmental processes, in particular with myelination beginning in primary areas before association neocortical areas, is a conserved feature of primate development, but the temporal progression of certain processes is heterochronic.

Concerted ontogenetic and phylogenetic transcriptomic divergence

After matching the global transcriptome by age between the two species, we analyzed regional differences in gene expression (heterotopy) within each species. By adopting

Gaussian process models to accommodate the spatio-temporal correlations of gene expression (*32*), we found that the developmental cup-shaped or hourglass-like pattern of transcriptomic inter-regional differences we observed in humans (*33*) is also present in macaque neocortices and other brain regions (Figs. 2, A and B, S21) with greater differential expression between regions observed during early and mid-fetal ages preceding this period and subsequent young adulthood. Notably, two brain regions – CBC and STR – exhibited greater differences, compared to other brain regions, beginning immediately after birth, rather than beginning during childhood/adolescence (fig. S21). This suggests that the development of the primate forebrain may be constrained by unique developmental or evolutionary influences, leading us to query the gene expression patterns, developmental processes, and cell types underlying this transcriptomic phenomenon.

In order to do that, we considered three phases of brain development mirroring major transitions in the cup-shaped pattern: prenatal development, early postnatal development, and adulthood. Between these three phases are two transitional periods: a steep late-fetal transition (*33*) and a more moderate transition between childhood/adolescence and adulthood. We performed weighted gene co-expression network analysis (WGCNA) independently for each phase and species, resulting in *Homo sapiens* (HS) and macaque (*Macaca mulatta*, MM) modules ((*32*); table S7), with analyses conducted on eleven neocortical areas; this allowed us to identify discrete spatio-temporal expression patterns that otherwise might be co-mingled due to the highly disparate nature of CBC and other non-neocortical regions. Within the prenatal phase, we found twelve modules consisting of genes exhibiting spatial expression gradients along

the anterior/posterior (8 modules), and medial/lateral (1 module) axes of the neocortex and broadly reflecting prospective neocortical areal topography (Fig. 2C). For example, prenatal modules HS85 and HS87 exhibited prefrontal/frontal-enriched graded expression in the human brain tapering to lowest expression in the temporal and occipital lobes (Fig. 2D). Furthermore, prenatal modules, such as HS15 and MM57, had their highest expression restricted to the temporal lobe (fig. S22-23; table S8) during prenatal development.

In contrast to the prenatal phase, modules identified from early postnatal development (i.e., infancy, childhood, and adolescence) in either species did not exhibit anterior/posterior or medial/lateral expression gradients. Rather, the greater regional synchrony characterizing gene expression in this phase yielded differences organized not around topography but between primary and association areas of the neocortex (Fig. 2C, S24-25; table S9). This suggests that the gradient-like transcriptomic patterns arising during prenatal development are superseded by myelination and neuronal activity-related processes postnatally, which may differentiate the separation between primary and association areas. Early postnatal modules such as MM42, MM24, and MM23, among others, exhibited greater expression in primary areas such as M1C, A1C, and V1C as compared to association areas such as DFC and VFC (Fig. 2D).

The transition to young adulthood was marked by another decrease in inter-regional differences, but this reduction was not as pronounced as in the late-fetal transition, nor were inter-regional patterns of gene expression markedly different in the adult. Thus, gene expression differences between primary and association areas continued to drive regional variation in both the adult human and macaque (Figs. 2, C and D, S26-

27; table S10). Gene Ontology (GO) enrichment analysis using the top variant genes in each period, with all genes expressed in each period as background, indicated that there is differential enrichment of biological processes associated with different cell populations across areas and time. As observed in the accompanying human study (*33*) and commensurate with the developmental trajectories of the observed transcriptomic signatures, the functional terms enriched prenatally were generally related to neurogenesis and neuronal differentiation, whereas early postnatal and adult functional terms were enriched for processes related to synaptogenesis and myelination (fig. S28).

We next sought to determine whether the regional-specific expression patterns of co-expression modules detected in human correlated with their expression patterns in macaque, and vice versa (*32*). We found that genes in two human prenatal modules exhibiting a pronounced anterior/posterior gradient in the human neocortex, HS85 and HS87, did not contain genes with enriched expression in the macaque prefrontal cortex (Fig. 2D; table S8). Among genes in these modules were *RGMA* and *SLIT3*, two genes encoding axon guidance molecules (*50*), and *BRINP2* and *CXXC5*, which encode proteins involved in retinoic acid signaling (*51*), potentially implicating this signaling pathway, critical for early brain development and neuronal differentiation (*51*), in the patterning of the human prefrontal cortex. We also observed several modules in macaque postnatal development (MM23, MM24, MM26, and MM42) that did not correlate well with human modules and were enriched for genes that are expressed in oligodendrocytes (Fig. 2D; fig. S24; table S9) and were up-regulated in all primary areas of macaque NCX relative to association areas. Conversely, genes in these modules were up-regulated in humans only in M1C and A1C, but not in S1C or V1C (fig. S24; table S9). Integration with our multi-

regional database of the adult chimpanzee transcriptome (*34*) indicates that the macaque gene expression pattern, rather than the human gene expression pattern, may be unique among these species (fig. S29). Many of the species-specific patterns of diversification between primary and association areas that we observed during early postnatal development were preserved in adult modules of both species (fig. S26), with some notable exceptions. For example, the adult macaque module MM25 exhibited up-regulation in association areas in both species but prominent up-regulation in medial prefrontal cortex (MFC) and down-regulation in V1C only in macaque (Fig. 2D; fig. S26; table S10). These findings reaffirm a conserved framework in primate neocortical development and function (*21*), including a topographic basis for transcriptomic differences during prenatal development and functional relationships postnatally. Our analyses also suggest that inter-regional and inter-species differences in oligodendrocyte development and myelination, particularly during early postnatal development, mediate key aspects of transcriptomic variation both within and among species.

Heterotopic changes in human and macaque brain transcriptome

We next investigated the transcriptomic divergence between human and macaque for each brain region across development. We found that the developmental phases exhibiting high levels of inter-regional differences within each species (i.e. prenatal development and young adulthood) also displayed greater transcriptomic divergence between the two species, revealing a concerted phylogenetic (evolutionary) cup-shaped pattern (Fig. 3A). This phylogenetic cup-shaped pattern divided neurodevelopment into the same three phases as the regional ontogenetic (developmental) cup-shape (Fig. 3A).

However, unlike the ontogenetic (developmental) cup-shaped pattern, where CBC, MD and STR disproportionally exhibited more intra-species differences than NCX, HIP, and AMY, all regions appeared to exhibit a relatively similar amount of inter-species differences (Fig. 3A). Interestingly, inter-species differences among neocortical areas were distinct enough to provide clear clustering of topographically and functionally related prefrontal areas (i.e., MFC, OFC, DFC, and VFC), particularly during prenatal development, or topographically distributed non-visual primary areas (i.e., M1C, S1C, and A1C) in adulthood. Prospective areas of the prefrontal cortex, which underlie some of the most distinctly human aspects of cognition, were more phylogenetically distinct than other neocortical areas during early prenatal development (Fig. 3A and fig. S30). Together, these findings suggest that the evolutionary and developmental constraints acting on the brain transcriptome, in particular NCX, may share some overlapping features.

To get insights into the transcriptomic programs driving phylogenetic divergence across neocortical areas, we conducted a functional annotation of the top 100 genes driving the observed variation along the first principal component (PC1). We found that inter-species divergence in the prenatal prefrontal cortex could be explained by an enrichment of genes related to cell proliferation (FDR < 10^{-5}). This indicated that the observed inter-species divergence in prefrontal cortex was likely due to a different proportion of progenitor cells in the early fetal human prefrontal tissue samples (fig. S30). In contrast, during postnatal development, PC1 separated prefrontal areas and inferior temporal cortex (ITC) from the other neocortical areas. This pattern was mainly driven by genes associated with myelination-associated categories (FDR < 0.05; fig. S30)

and genes associated with synaptic transmission (FDR < 0.05; fig. S29). While speculative, these observations potentially link the expansion of human prefrontal cortex, the wealth of human-specific connectivity made possible by that extension, and the altered patterns of myelination we observe between human and macaque. Confirming the observed regional diversification in each species, postnatal development displayed the lowest number of differentially expressed genes between species and most of these (89.3%) were also differentially expressed in adulthood, the phase where we observed the greatest number of inter-species differentially expressed genes (Fig. 3B; table S11). Genes differentially expressed between human and macaque exhibited distinct patterns of spatio-temporal divergence (Fig. 3C) and showed diverse functional enrichment (table S12). While 229 genes (2.6%) displayed up- or down-regulation in all the sampled brain regions throughout development and adulthood, others were specifically up- or down-regulated in a subset of brain regions and/or in a particular developmental phase.

To test whether genes with differential expression between human and macaque showed distinct conservation profiles, we compared dN/dS values for the whole set of genes differentially expressed in any of the 16 brain regions in at least one of the three developmental phases (*32*). We found that the differentially expressed genes between human and macaque also show significantly higher dN/dS values associated with higher evolutionary rates than the remaining protein coding genes (Wilcoxon-Mann-Whitney $P = 2.2\text{x}10^{-8}$, n = 4,429). This result was also observed when we focused on the genes differentially expressed in prenatal development ($P = 3.7\text{x}10^{-11}$, n = 2,380 genes), early postnatal development ($P = 4.5\text{x}10^{-24}$, n = 1,765 genes) or adulthood ($P = 1.0\text{x}10^{-6}$, n = 3,837 genes) separately. Moreover, these higher dN/dS values for differentially expressed

genes remained highly significant in all the brain regions and developmental phases analyzed, highlighting the consistent association between inter-species transcriptional variation and gene evolution.

Integration with our complementary dataset generated on adult chimpanzee brains (*34*), revealed that 531 (10.6%), 507 (12.9%), and 1079 (13.9%) genes differentially expressed between species in prenatal development, early postnatal development and adult, respectively, show human-specific expression in the same brain region in the adult brain. Several genes among those exhibiting species- or human-specific patterns of gene expression were developmentally and regionally regulated. *PKD2L1*, a gene that encodes an ion channel (*52*), exhibited human-specific up-regulation only postnatally (Fig. 3C). Conversely, *TWIST1*, a gene encoding a transcriptional factor implicated in Saethre-Chotzen syndrome (*53*), showed human-specific down-regulation only postnatally (Fig. 3C). In contrast, *MET*, a gene linked to autism spectrum disorders (*54*), showed human-specific up-regulation in the prefrontal cortex and STR postnatally (Fig. 3C). *PTH2R*, a gene encoding the parathyroid hormone 2 receptor, exhibited macaque-distinct up-regulation in prenatal NCX, but human-distinct up-regulation in the adult NCX, and is enriched, as evidenced by immunohistochemistry, in excitatory neurons (fig. S31). These results show that at least some of the tissue level inter-species differences we observed are due to changes at the level of specific cell types. Furthermore, even though the ontogenetic and phylogenetic patterns have similar profiles, the overlap of genes driving these two patterns is not substantial (Fig. 3D), indicating the existence of different molecular mechanisms and constraints for regional specification and species divergence.

To gain a more complete understanding of the inter-species transcriptomic differences, we performed inter-species differential exon usage as a conservative way of exploring the impact of putative differential alternative splicing. We detected largely similar numbers of genes containing clear differentially used exons between species in all developmental phases ((*32*); table S13), with 1,924 genes showing inter-species differential exon usage in at least one brain region during the prenatal phase, 1,952 during the early postnatal phase, and 1,728 during adulthood (Figs. 3B and S32). In our set of differentially used exonic elements, non-protein coding regions were overrepresented ($P < 2.2x10^{-16}$, Chi-squared independence test), with 4,705 out of the 5,372 differentially used exonic elements in non-coding regions. This enrichment was especially strong for non-UTR exonic elements belonging to non-protein coding transcripts from protein-coding genes and 5' UTR regions ($P < 2.2x10^{-16}$), but also significant for 3' UTR regions ($P = 1.81x10^{-11}$) and non-UTR exonic elements from non-protein coding genes ($P = 0.02364$), suggesting post-transcriptional regulation may contribute to species differences at the exon level.

Phylogenetic divergence in transcriptional heterotopic regulation

Because transcription factors can regulate the expression of multiple genes, the differential expression we observe between species in different brain regions might be mediated in part by differential expression of a relatively small number of transcription factors. To assess this possibility, we searched for transcription factor binding sites (TFBS) that were enriched in the annotated promoters of inter-species differentially expressed genes for each brain region and developmental stage in our analysis (*32*). We

found that the binding sites for 86 transcription factors were enriched among inter-species differentially expressed genes, 7 of which were inter-species differentially expressed genes (table S14). *RUNX2* was differentially expressed between human and macaque in the prenatal HIP, *PAX7* in the early postnatal AMY, *STAT6* in the prenatal NCX, *STAT4* in the early postnatal and adult NCX, *SNAI2* in the adult CBC, and *EWSR1* and *NEUROD1* in the adult NCX. Although these enriched motifs were found in only a relatively small proportion of the promoters of the inter-species differentially expressed genes (table S15), expression changes of almost 30% of the differentially expressed genes in NCX can be explained solely by *STAT4*, *EWSR1*, and *NEUROD1*, transcription factors that have been previously implicated in neuronal development (*55*) and brain disorders (*56*, *57*). This suggests that species differences in the expression levels of influential transcription factors could be phenotypically relevant.

To substantiate the possibility that these transcription factors might regulate inter-species differences in gene expression, we next conducted an independent analysis utilizing epigenomic data of regulatory elements (including both active promoters and enhancers) showing macaque-human differential signal of H3K27ac binding in different regions of adult brains (*58*). Using region-matched (i.e., NCX, STR, MD, and CBC) aspects of this dataset, we performed TFBS enrichments for both the regions defined as up-regulated in human as well as those down-regulated in human compared to macaque ((*32*), table S16-18). As before, we then compared TFBS enriched among regulatory elements differentially detected in human and macaque with the transcription factors differentially expressed in a given area/region between species. We observed a higher number of differentially expressed transcription factors associated with binding sites

selective for epigenetic loci down-regulated in human (17, 6, 6, and 1 for NCX, CBC, MD, and STR, respectively) than for loci up-regulated in human (3, 1, and 1, for NCX, CBC, and MD). Moreover, 86% of promoters associated with inter-species differentially expressed genes in the NCX contained TFBS for transcription factors that were differentially expressed between species in NCX. The same was true for 33% of all differentially expressed genes retrieved from CBC, 29% for the differentially expressed genes in MD, and 8.5% of the differentially expressed genes in STR.

We found that all enriched DEX-TFBS were also found in the TFBS enrichment using epigenomic data (*58*) in matched brain regions and developmental stages. The good agreement between both independent datasets supports the regulatory relevance of these DEX-TFBSs in driving the expression changes of other DEX genes.

Diversity and cell type specificity of species differences

To explore whether cell type-specific transcriptomic changes account for the inter-species divergence observed at the tissue level, we tested the enrichment of human up-regulated genes in human single cells and human down-regulated genes in macaque single cells. Furthermore, we used prenatal scRNA-Seq data for prenatal differentially expressed genes and adult snRNA-Seq data for the early postnatal and adulthood periods (Figs. 4, A and B; fig. S33). In all prenatal neocortical areas, human up-regulated genes were enriched in neural progenitors, indicating that human NCX may possess more neural progenitors at matched timepoints compared with macaque counterparts, although we cannot completely exclude the possibility that a lack of macaque samples matching human early fetal samples (Fig. 1, A and B) might contribute to this observation, despite

the efforts we made to minimize the effects of sampling bias between species by fitting a Gaussian process model. In contrast, macaque up-regulated genes were enriched in multiple subtypes of excitatory and inhibitory neurons in all neocortical areas (Fig. 4A). Interestingly, a specific subtype of excitatory neurons (i.e., ExN2) was enriched for the macaque up-regulated genes only in prefrontal areas. In the postnatal and adult NCX, human up-regulated genes were enriched in a single population of likely upper-layer excitatory neurons (ExN2b), which was not described in a recent snRNA-Seq study of the adult human neocortex (*59*). Conversely, postnatally up-regulated macaque genes were enriched in multiple subtypes of excitatory neurons (Fig. 4B). Inter-species differentially expressed genes in non-neocortical brain regions of the prenatal brain were also enriched in specific cell types (fig. S33). For example, genes displaying inter-species differentially expressed genes in HIP, STR, and CBC were enriched in a population of oligodendrocyte progenitor cells (OPC), medium spiny neurons (MSN), and external granular layer transition to granule neuron cells (EGL-TransGraN), respectively. Furthermore, genes showing inter-species differential expression in hippocampus, amygdala, striatum, and cerebellum were enriched in a population of microglia (fig. S34).

By integrating our single-cell data sets with a tissue-level transcriptomic dataset of adult human, chimpanzee, and macaque brains (*34*), we identified the cell-type enrichment of several genes showing human-specific up- or down-regulation in NCX or all brain regions compared with chimpanzee and macaque. For example, *CD38* was found to be down-regulated in all human brain regions and enriched in astrocytes (Fig. 4C). This gene encodes a glycoprotein that is important in the regulation of intracellular calcium and its deletion leads to impaired development of astrocytes and

oligodendrocytes in mice (*60*). *CLUL1*, a gene reported to be specifically expressed in cone photoreceptor cells (*61*), showed human-specific up-regulation in all brain regions and was enriched in oligodendrocytes and astrocytes. *TWIST1* exhibited human-specific down-regulation in all neocortical areas postnatally, and is enriched in putative upper-layer excitatory neurons (Fig. 4C). Conversely, *PKD2L1* is up-regulated in NCX postnatally and enriched in putative deep-layer excitatory neurons (Fig. 4C). *MET* exhibited human-specific up-regulation in the prefrontal cortex and STR postnatally, and is enriched in upper-layer excitatory neurons (Fig. 4C).

Shared and divergent transcriptomic features of homologous cell types

To test whether the observed differential expression between human and macaque was due to differences in cell type composition or due to transcriptomic differences between homologous cell types, we performed a comparative analysis between human and macaque cell types of prenatal and adult dorsolateral prefrontal cortices. The correlation between human and macaque cell types showed that all human cell types had a close homologue in macaque, and vice versa (Fig. 5A and B). Nevertheless, we identified genes showing inter-species differential expression in homologous cell types (Fig. 5C). To avoid biases inherent to high variation in sc- or sn-RNA-Seq, we filtered out genes that did not display differential expression between species at the tissue-level and only included genes that exhibited enrichment in cell types where they show inter-species differential expression (Preferential expression measure > 0.3; (*32*)). We identified 14 differentially expressed genes in prenatal development and 41 differentially expressed genes in adulthood (Fig. 5C). For example, *TRIM54*, which encodes a protein implicated

in axonal growth (*62*), was down-regulated in human prenatal neocortical excitatory neurons (Fig. 5C). *VW2CL*, which encodes a protein associated with AMPA-type glutamate receptors (*63*), was down-regulated in prenatal human neocortical interneurons. *SLC17A8* (aka *VGLUT3*), which encodes the vesicular glutamate transporter 3, is up-regulated in human postnatal somatostatin-positive interneurons (InN8). Overall, we found that human DFC cell types showed high correlation with macaque DFC cell types and that only a small set of genes displayed differential expression between these homologous cell types (Fig. 5C). Thus, the inter-species differences identified at the tissue level are likely to be the results of variations in cellular diversity, abundance, and, to a lesser extent, transcriptional divergence between cell types.

Heterochronic changes in human and macaque brain transcriptome

The observed heterotopic differences may result, in part, from changes in the timing of gene expression, or heterochrony. To identify such heterochronic differences, we created a Gaussian-process based model (*TempShift*; (*32*)) and applied this model independently to human and macaque gene expression datasets. Due to the similar nature of the transcriptomic signature of the eleven neocortical areas relative to other brain regions, and to maintain consistency with earlier analyses, we focused our analysis on eleven neocortical areas (see *(33)*). We identified genes with inter-regional temporal differences within neocortical areas of each species and aggregated them into 36 regional clusters (RC; fig. S35; table S19). For both human and macaque brain, analysis of all heterochronic genes revealed greater inter-areal differences during prenatal periods than early postnatal or adult ages (fig. S36). In addition, although we observed differences in

inter-areal heterochrony between the early postnatal phase and the adult phase in human, we did not observe these differences in macaque (fig. S36). This suggests that inter-regional synchrony in the macaque precedes that in age-matched human, possibly reflecting the protracted development of the human brain during childhood and the earlier plateauing of myelination-associated processes in macaque postnatal development (Figs. 1C and S19). Analysis of the regional clusters revealed further insights into shared and species-distinct aspects of neurodevelopment. For example, we identified five regional clusters (RC 4, 21, 26, 29, and 34) enriched for genes expressed selectively by neural progenitors that exhibited temporal differences between human neocortical areas (fig. S35). Each of these clusters exhibited a gradient whereby a decrease in expression in central regions of the prenatal neocortex preceded a decrease at the anterior and posterior poles, suggesting increased progenitor populations or a prolonged neurogenic period in prefrontal cortex as well as STC, ITC, and V1C. However, although we observed similar temporal gradients in macaque for RC 4, 26, and 29, neither RC21 nor RC34, the modules exhibiting the sharpest delay in posterior neocortex, exhibited a similar central-to-polar gradient in macaque (Fig. 6A). Conversely, RC10 and RC12 exhibited an inverse gradient in humans, with decreased expression in prefrontal neocortex and STC, ITC, and V1C preceding a decrease in central cortex. These modules, which are enriched in astrocytes, did not exhibit a similar gradient in macaque (Fig. 6A; fig. S35). This indicates that even though the transcriptomic signature associated with astrogliogenesis showed a global synchronicity between species (Figs. 1C and S19), a smaller group of genes enriched in astrocytes displayed heterochrony between species.

Despite the global enrichment of heterochronic genes in prenatal development (fig. S36), we also identified clusters exhibiting higher inter-regional differences in postnatal development and adult. One example is RC 25, a cluster enriched for oligodendrocyte markers that exhibited a pattern of early expression in primary motor and somatosensory areas in macaque but not in human NCX (Fig. 6A). This corroborates myelination-related regional asynchrony, as primary areas myelinate earlier, and that there is inter-species heterochrony in oligodendrocyte maturation and myelination-associated processes. Reflective of the cup-shaped pattern of regional variation in global development, the regional clusters also suggest the asynchronous maturation of prenatal areas, a gradual synchronization during early postnatal development in both species, and additional postnatal and adult differences driven in part by myelination.

We next applied *TempShift* to identify genes exhibiting inter-species heterochronic divergence. Among eleven neocortical areas, we identified approximately 3.9% of coding and non-coding mRNA genes (1,100 out of 27,932 analyzed orthologous genes) exhibiting inter-species heterochronic expression in at least one neocortical area. We then used Ingenuity Pathway Analysis to assess upstream transcriptional regulation of heterochronic genes. We found that the differential expression of 139 inter-species heterochronic genes could be explained by as few as 8 co-regulated heterochronic transcriptional regulators (Fig. 6B; (*32*)), plus one transcription factor with heterotopic expression (down-regulated in the postnatal human neocortex) between species, *TWIST1* (fig. S37). Interestingly, a majority (90 out of 139) of these putative target genes of the nine transcriptional regulators exhibited accelerated expression in the human neocortex. As aforementioned, humans exhibit an accelerated heterochronic pattern for the

synaptogenesis transcriptomic signature and the presence of *FOS*, a neuronal activity-regulated gene, as one of the hubs of this transcriptional network, indicates that this accelerated synaptogenesis likely drives the accelerated expression of several genes in the human neocortex. Furthermore, an ontological analysis of the genes with heterochronic expression revealed an enrichment for functional categories such as “axonal guidance signaling”, “glutamate receptor signaling” and “CREB signaling in neurons” (Fig. 6C), suggesting that heterochronic processes include molecular pathways related to axon guidance and synaptic activity.

We next identified 15 evolutionary clusters (EC) on the basis of the 1,100 heterochronic genes displaying inter-species neocortical heterochronic expression patterns (table S20). Among the evolutionary clusters, EC 14 exhibited a delayed expression in the human dorsolateral prefrontal cortex and was enriched for intermediate progenitor cell (IPC) markers (Figs. 6D and S38) in agreement with the progenitor cell population differences we observed previously in prefrontal cortex, indicating that this neocortical prefrontal area likely has a protracted neurogenesis when compared to macaque. Similarly, the species-distinct maturation gradients of neural progenitors, astrocytes, and oligodendrocytes also support observations we made concerning inter-species heterotopy. These results were supported by selective validation of the expression profiles of heterochronic genes; using droplet digital PCR, we selected 5 genes with different developmental profiles across regions and species (figs. S39-43), confirming not just the expression profiles of these genes but also that our observations were not the result of biases introduced by *TranscriptomeAge.*

Species difference in spatio-temporal expression of disease genes

Next, we investigated whether genes associated with risk for neuropsychiatric disorders exhibited differences in their spatio-temporal expression between human and macaque. We focused our analysis on genes linked to autism spectrum disorders (ASD) and other neurodevelopmental disorders (NDD), attention deficit hyperactivity disorder (ADHD), schizophrenia (SCZ), bipolar disorder (BD), major depressive disorder (MDD), Alzheimer's disease (AD), and Parkinson's disease (PD) in previous genetic studies or through our integrative analysis from the accompanying study or (see (*33*) and table S21 for details). We next sought to determine whether the expression of genes associated with these neuropsychiatric disorders were enriched in any particular developmental phase. Consistent with previous studies associating the mid-fetal timeframe with specific high confidence ASD (hcASD) genes (see (*64*)), we found that a larger group of hcASD genes were more highly expressed in the prenatal than the early postnatal and adult brains in both species (fig. S44). In contrast, AD-related genes were more highly expressed in the early postnatal and adult than the prenatal brains in both species (fig. S44). Other groups of disease related genes did not show any obvious global difference across development. Potentially suggesting the involvement of species-specific aspects in the etiology of ASD, NDD and SCZ, we identified genes with heterochronic or heterotopic expression between the two species that are associated with ASD (6 and 0, respectively), non-hcASD NDD (56 and 14, respectively) and SCZ (45 and 14, respectively) (Figs. 7A and C). Unsupervised hierarchical clustering of SCZ- associated genes, but not NDD, with heterotopic expression yielded 5 obvious spatio-temporal clusters, three of which exhibited species differences exclusively during prenatal development (fig. S44). Of the

prenatal clusters, cluster 1 showed enrichment in prefrontal cortex, cluster 3 in temporal cortex, and cluster 2 in both frontal and temporal cortices, in human. Cluster 4 displayed an enrichment in the postnatal and adult macaque frontal cortex and cluster 5 exhibited a similar enrichment in adult macaque prefrontal cortex (Fig. 7D).

Further analysis revealed that the ASD-associated genes *SHANK2* and *SHANK3*, which encode synaptic scaffolding proteins at the postsynaptic density of excitatory glutamatergic synapses, exhibited earlier expression in the macaque NCX and other brain regions relative to human (Fig. 7B). Commensurate with a role for these proteins in neural circuit development, and in agreement with analyses suggesting the involvement of neocortical projection neurons in the etiology of ASD, these two genes also became progressively more expressed across prenatal ages in both human and macaque (fig. S45). SCZ-associated genes displaying inter-species heterochrony included *GRIA1*, a glutamate ionotropic receptor AMPA type subunit that has different expression trajectories in MFC and OFC compared with other neocortical areas, and that is expressed earlier in human VFC, M1C, S1C, IPC, and STC. (Fig. 7B and fig. S45). These evolutionary changes in the spatio-temporal expression of certain disease-associated genes might therefore imply transcriptional underpinnings for potential human-specific aspects of neuropsychiatric disorders. For example, the presence of human-distinct heterochrony in synapse-related proteins associated with ASD, coupled with the lack of obvious heterotopic expression in hcASD genes, may suggest conserved neurodevelopmental programs common to primate species are uniquely shifted temporally in some areas in the human brain, potentially implicating key developmental periods, places, and cell types involved in disease etiology. Similarly, the heterochronic and heterotopic changes we associated with SCZ,

in particular those affecting the prenatal prefrontal and temporal cortices, may be involved in human-specific aspect of disease etiology.

Given the importance of untranslated regions and other non-coding regions in the regulation of gene expression as well as disease, we next explored inter-species differences in exon usage between species in genes associated with neuropsychiatric disorders. We observed that 413 genes with differentially expressed exonic elements were linked to the studied diseases. Moreover, we detected 35 disease genes showing differentially used exonic elements with predicted binding sites (*65*) for miRNAs independently associated with central nervous system diseases ((*32*, *66*), table S22). Several of these genes (e.g., *GRIN2B*, *BCL11B*, and *NKPD1*) were potentially targeted by a large number of disease-associated miRNAs (fig. S46), and gene-miRNA interactions have already been experimentally validated for 11 of the 35 genes we identified, according to miRTarBase (*67*) (table S23). For example, we detected differential exon usage of *BCL11B*, a gene involved in the development of medium spiny neurons (*68*), between human and macaque in the adult STR (fig. S46). However, while *BCL11B* shows lower expression in human than in macaque STR, the exonic element containing the 3'UTR of *BCL11B* was itself not differentially expressed. This observation suggests that overexpression in macaque is associated with an isoform containing a shorter 3'UTR region lacking the capacity to be bound by various miRNAs, possibly including, based on predicted binding sites, members of the miR-219 family of brain-specific miRNAs, which have been experimentally shown to interact with *BCL11B* mRNA (*69*). Together, these findings indicate that certain genes associated with neuropsychiatric disorders

exhibit changes in the timing of their expression, location and splicing pattern between human and NHP brains, and thus may lead to species differences in disease pathogenesis.

Discussion

In this study, we present a comprehensive spatio-temporal transcriptomic brain dataset of the macaque brain. Resource description and access are available at *evolution.psychencode.org*. In addition, we present computational methods such as *TranscriptomeAge* to conduct unbiased age matching across species, and *TempShift*, a Gaussian-process based model for identifying transcriptomic changes in the timing of biological processes.

Our integrative and comparative analysis involving complementary human and adult chimpanzee (*33, 34*), revealed similarities and differences in the spatio-temporal transcriptomic architecture of the brain and the progression of major neurodevelopmental processes between the two species. For example, we have identified shared and divergent transcriptomic feature among homologous brain regions and cell types. We found transcriptomic evidence suggesting that human childhood is especially protracted, as compared to macaques. It has long been recognized that the development of the human brain is prolonged compared to that of other NHPs and that this slower rate of maturation expands the period of neural plasticity and capacity for learning activities, memory, and complex sensory perception, all processes necessary for higher order cognition (*1-4, 14, 28*). We also found that the early periods of human fetal neurodevelopment are transcriptomically distinct and protracted as compared to macaque. Interestingly, a similar observation of early neurodevelopmental protraction was recently observed in

vitro, in neural progenitors derived from pluripotent cells of human and NHPs (*70*). However, we also identified cases of neoteny in macaques, such as the protracted postnatal expression of DCX in the hippocampus, likely reflecting differences in neurogenesis between the two species as recently shown (*49*).

We found that global patterns of spatio-temporal transcriptomic dynamics were conserved between humans and macaque, and display a highly convergent cup-like shape. The most dramatic decrease in inter-regional differences occurs during late fetal ages and before birth, likely reorganizational processes at this developmental period rather than extrinsic influences due to birth and subsequent events (i.e., respiratory activity or other developmentally novel stimuli). Interestingly, after this transitional period, diversification of neocortical areas appears to be driven mainly by differences between primary and association areas. In addition to these largely conserved broad developmental patterns of inter-regional differences, we identified numerous genes and gene modules with human-distinct heterochronic or heterotopic expression. These patterns involved brain regions such as the developing prefrontal areas, which are central to the evolution of distinctly human aspects of cognition and behavior (*19-21*). Surprisingly, we also found that developmental phases exhibiting high levels of inter-regional differences (i.e., early to mid-fetal periods and young adulthood) were also less conserved between the two species. The coincident convergence of the ontogenetic and phylogenetic cups during the late fetal period and infancy is strikingly distinct from the previously reported phylogenetic transcriptomic hourglass-like pattern that occurs during the embryonic organogenetic period (*71, 72*).

Genes with divergent spatio-temporal expression patterns included those previously linked to ASD, SCZ and NDD. These species differences in the expression of disease-associated genes linked to synapse formation, neuronal development, and function, as well as regional and species differences in synaptogenesis and myelination, might have implications for the overall development of neural circuitry and consequently human cognition and behavior. These observations are possibly relevant for recent non-human primate models of neuropsychiatric disease, such as the *SHANK3*-deficient macaque model (*73*), which might therefore not be capable of fully capturing human-distinct aspects of *SHANK3* regulation during neurodevelopment.

Together, our study reveals insights into the evolution of gene expression in the developing human brain. Future work on the development patterns and the functional validation of the genes we report to have heterotopic and/or expression patterns between human and macaque will likely shed some light on potentially human-specific underpinnings of certain neuropsychiatric disorders.

Materials and Methods

Sixteen regions of the macaque brain spanning from early prenatal to adulthood were dissected using the same standardized protocol used for human specimens and described in the accompanying study by Li et al ((*33*); see also (*32*)). The macaque brain regions and developmental timepoints matched human brain regions and timepoints analyzed in the study by Li et al ((*33*)). The sampled homologous brain regions were identified using anatomical landmarks provided in the macaque brain atlas (*74*). An overview of dissected brain regions is provided in fig. S1. *Translating Time* model (*38*) was used to identify

equivalent timepoints between macaque and human prenatal development. The list of macaque brains used in this study and relevant metadata are provided in tables S1-2. Macaque studies were carried out in accordance with a protocol approved by Yale University's Committee on Animal Research and NIH guidelines.

We performed tissue-level RNA extraction and sequencing of all 16 regions, single-cell RNA-Seq of dorsolateral prefrontal cortex [DFC], hippocampus [HIP], amygdala [AMY], striatum [STR], mediodorsal nucleus of the thalamus [MD], and cerebellar cortex [CBC] of mid-fetal macaque, and single-nucleus RNA-Seq of DFC of adult macaque. Single cell/nucleus sample processing was done with 10X Genomics and sequencing was done with Illumina platforms.

For tissue-level analysis, we generated annotation of human-macaque orthologs using the *XSAnno* pipeline, and matched the developmental age of human and macaque samples based on their respective transcriptome using our algorithm *TranscriptomeAge.* We also developed *TempShift*, a method based on Gaussian process model, to reveal the inter-regional differences, inter-species divergence, and genes with heterotopic and heterochronic expression. We also queried differentially expressed genes for enrichment in transcription factor binding sites using findMotifs.pl, and analyzed inter-species differential exon usage using the R package DEXSeq.

The single cell/nucleus data was first analysed by cellranger for decoding, alignment, quality filtering, and UMI counting. After that, data was further analyzed with *Seurat* according to its guidelines, and cell types were clustered for classification with *SpecScore.R.* In order to perform direct comparisons between human and macaque at the single-cell level, we focused on the homologous genes between these species and aligned

monkey and human cells together to further analyze inter-species divergence of homologous cell types (fig. S47). We used *MetageneBicorPlot* function to examine the correlation of neuronal and glial cell subtypes, and we employed the correlation analysis to detect the correspondence of excitatory neuron and interneuron subtypes. Finally, we did functional enrichment of disease-associated genes in both tissue-level and single-cell datasets.

References

1. M. F. A. Montagu, Time, Morphology, and Neoteny in the Evolution of Man. *Am. Anthropol.* **57**, 13-27 (1955).
2. J. Dobbing, J. Sands, Quantitative growth and development of human brain. *Arch. Dis. Child* **48**, 757-767 (1973).
3. H. F. Prechtl, New perspectives in early human development. *European journal of obstetrics, gynecology, and reproductive biology* **21**, 347-355 (1986).
4. B. Bogin, Evolutionary perspective on human growth. *Annu. Rev. Anthropol.* **28**, 109-153 (1999).
5. A. H. Schultz, *Age changes in primates and their modification in man.* J. M. Tanner, Ed., Human Growth (Pergamon Press, Oxford, 2009).
6. J. C. Silbereis, S. Pochareddy, Y. Zhu, M. Li, N. Sestan, The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* **89**, 248-268 (2016).
7. R. S. Hill, C. A. Walsh, Molecular insights into human brain evolution. *Nature* **437**, 64-67 (2005).

8. J. L. Fish, C. Dehay, H. Kennedy, W. B. Huttner, Making bigger brains-the evolution of neural-progenitor-cell division. *J. Cell. Sci.* **121**, 2783-2793 (2008).

9. D. H. Geschwind, G. Konopka, Neuroscience in the era of functional genomics and systems biology. *Nature* **461**, 908-915 (2009).

10. J. H. Lui, D. V. Hansen, A. R. Kriegstein, Development and evolution of the human neocortex. *Cell* **146**, 18-36 (2011).

11. S. A. McCarroll, S. E. Hyman, Progress in the genetics of polygenic brain disorders: significant new challenges for neurobiology. *Neuron* **80**, 578-587 (2013).

12. E. S. Lein, T. G. Belgard, M. Hawrylycz, Z. Molnar, Transcriptomic Perspectives on Neocortical Structure, Development, Evolution, and Disease. *Annu Rev Neurosci* **40**, 629-652 (2017).

13. G. E. Hardingham, P. Pruunsild, M. E. Greenberg, H. Bading, Lineage divergence of activity-driven transcription and evolution of cognitive ability. *Nat. Rev. Neurosci.* **19**, 9-15 (2018).

14. T. Paus, M. Keshavan, J. N. Giedd, Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci*. **9**, 947-957 (2008).

15. B. Y. Liao, J. Zhang, Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6987-6992 (2008)

16. T. Barak, *et al.*, Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat. Genet.* **43**, 590-594 (2011).

17. M. O'Bleness, V. B. Searles, A. Varki, P. Gagneux, J. M. Sikela, Evolution of genetic and genomic features unique to the human lineage. *Nature Reviews. Genetics* **13**, 853-866 (2012).

18. K. Mirnics, J. Pevsner, Progress in the use of microarray technology to study the neurobiology of disease. *Nat. Neurosci.* **7**, 434-439 (2004).

19. T. M. Preuss, Human brain evolution: From gene discovery to phenotype discovery. *Proc. Natl. Acad. Sci. U.S.A.* **109 Suppl 1**, 10709-10716 (2012).

20. C. C. Sherwood, A. L. Bauernfeind, S. Bianchi, M. A. Raghanti, P. R. Hof, Human brain evolution writ large and small. *Prog. Brain Res.* **195**, 237-254 (2012).

21. A. M. M. Sousa, K. A. Meyer, G. Santpere, F. O. Gulden, N. Sestan, Evolution of the Human Nervous System Function, Structure, and Development. *Cell* **170**, 226-247 (2017).

22. H. B. van der Worp, D. W. Howells, E. S. Sena, M. J. Porritt, S. Rewell, V. O'Collins, M. R. Macleod, Can animal models of disease reliably inform human studies? *PLoS Med* **7**, e1000245 (2010).

23. T. Marques-Bonet, O. A. Ryder, E. E. Eichler, Sequencing primate genomes: what have we learned? *Annu Rev Genomics Hum Genet* **10**, 355-386 (2009).

24. J. C. Izpisua Belmonte, E. M. Callaway, S. J. Caddick, P. Churchland, G. Feng, G. E. Homanics, et al., Brains, genes, and primates. *Neuron* **86**, 617-631 (2015).

25. C. G. Jennings, R. Landman, Y. Zhou, J. Sharma, J. Hyman, J. A. Movshon, et al., Opportunities and challenges in modeling human brain disorders in transgenic primates. *Nat Neurosci* **19**, 1123-1130 (2016);.

26. K. Sato, E. Sasaki, Genetic engineering in nonhuman primates for human disease modeling. *Journal of human genetics* **63**, 125-131 (2018).

27. M. B. Johnson, Y. I. Kawasawa, C. E. Mason, Z. Krsnik, G. Coppola, D. Bogdanovic, et al., Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494-509 (2009).

28. M. Somel, H. Franz, Z. Yan, A. Lorenc, S. Guo, T. Giger, et al., Transcriptional neoteny in the human brain. *Proc Natl Acad Sci U S A* **106**, 5743-5748 (2009).

29. H. J. Kang, Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, et al., Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-489 (2011).

30. M. Pletikos, A. M. Sousa, G. Sedmak, K. A. Meyer, Y. Zhu, F. Cheng, et al., Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* **81**, 321-332 (2014).

31. X. Liu, M. Somel, L. Tang, Z. Yan, X. Jiang, S. Guo, et al., Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. *Genome Res.* **22**, 611-622 (2012).

32. Materials and Methods.

33. M. Li, et al., Integrative functional genomic analysis of human brain development and neuropsychiatric risk convergence. *Science*, ***submitted***.

34. A. M. M. Sousa, Y. Zhu, M. A. Raghanti, R. R. Kitchen, M. Onorati, A. T. N. Tebbenkamp, et al.,, Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**, 1027-1032 (2017).

35. S. Zhong, S. Zhang, X. Fan, Q. Wu, L. Yan, J. Dong, et al., A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524-528 (2018).

36. J. A. Miller, S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, et al., Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206 (2014).

37. Y. Zhu, M. Li, A. M. Sousa, N. Sestan, XSAnno: a framework for building ortholog models in cross-species transcriptome comparisons. *BMC Genomics* **15**, 343 (2014).

38. A. D. Workman, C. J. Charvet, B. Clancy, R. B. Darlington, B. L. Finlay, Modeling transformations of neurodevelopmental sequences across mammalian species. *J Neurosci* **33**, 7368-7383 (2013).

39. J. Fooden, *Systematic review of the rhesus macaque, Macaca mulatta (Zimmermann, 1780)*. (Field Museum of Natural History, 2000), vol. Fieldiana Zoology New Series no.96.

40. B. Clancy, R. B. Darlington, B. L. Finlay, Translating developmental time across mammalian species. *Neuroscience* **105**, 7-17 (2001).

41. R. Knoth, I. Singec, M. Ditter, G. Pantazis, P. Capetian, R. P. Meyer, V. Horvat, B. Volk, G. Kempermann, Murine features of neurogenesis in the human hippocampus across the lifespan from 0 to 100 years. *PLoS One* **5**, 8809 (2010).

42. P. R. Huttenlocher, A. S. Dabholkar, Regional differences in synaptogenesis in human cerebral cortex. *J Comp Neurol* **387**, 167-178 (1997).

43. D. J. Miller, T. Duka, C. D. Stimpson, S. J. Schapiro, W. B. Baze, M. J. McArthur, et al., Prolonged myelination in human neocortical evolution. *Proc Natl Acad Sci U S A* **109**, 16480-16485 (2012).

44. J. N. Giedd, J. Blumenthal, N. O. Jeffries, F. X. Castellanos, H. Liu, A. Zijdenbos, T. Paus, A. C. Evans, J. L. Rapoport, Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci* **2**, 861-863 (1999).

45. J. L. Rapoport, N. Gogtay, Brain neuroplasticity in healthy, hyperactive and psychotic children: insights from neuroimaging. *Neuropsychopharmacology* **33**, 181-197 (2008).

46. E. R. Sowell, B. S. Peterson, P. M. Thompson, S. E. Welcome, A. L. Henkenius, A. W. Toga, Mapping cortical change across the human life span. *Nat. Neurosci.* **6**, 309-315 (2003).

47. P. Flechsig Of Leipsic, Developmental (myelogenetic) localisation of the cerebral cortex in the human subject. *The Lancet* **158**, 1027-1030 (1901).

48. P. Rakic, J. P. Bourgeois, M. F. Eckenhoff, N. Zecevic, P. S. Goldman-Rakic, Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science* **232**, 232-235 (1986).

49. S. F. Sorrells, M. F. Paredes, A. Cebrian-Silla, K. Sandoval, D. Qi, K. W. Kelley, et al., Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature* **555**, 377-381 (2018).

50. A. Chedotal, L. J. Richards, Wiring the brain: the biology of neuronal guidance. *Cold Spring Harbor perspectives in biology* **2**, a001917 (2010).

51. M. Rhinn, P. Dolle, Retinoic acid signalling during development. *Development* **139**, 843-858 (2012).

52. A. L. Huang, X. Chen, M. A. Hoon, J. Chandrashekar, W. Guo, D. Tränkner, N. J. Ryba, C. S. Zuker, The cells and logic for mammalian sour taste detection. *Nature* **442**, 934-938 (2006).

53. Q. Qin, Y. Xu, T. He, C. Qin, J. Xu, Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms. *Cell Res.* **22**, 90-106 (2012).

54. D. B. Campbell, J. S. Sutcliffe, P. J. Ebert, R. Militerni, C. Bravaccio, S. Trillo, et al., A genetic variant that disrupts MET transcription is associated with autism. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16834-16839 (2006).

55. A. Pataskar, J. Jung, P. Smialowski, F. Noack, F. Calegari, T. Straub, V. K. Tiwari, NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J* **35**, 24-45 (2016).

56. M. Rodriguez, L. Zoecklein, J. D. Gamez, K. D. Pavelko, L. M. Papke, S. Nakane, et al., STAT4- and STAT6-signaling molecules in a murine model of multiple sclerosis. *FASEB J* **20**, 343-345 (2006).

57. D. Sturm, B. A. Orr, U. H. Toprak, V. Hovestadt, D. T. W. Jones, D. Capper, et al., New brain tumor entities emerge from molecular classification of CNS-PNETs. *Cell* **164**, 1060-1072 (2016).

58. M. W. Vermunt, S. C. Tan, B. Castelijns, G. Geeven, P. Reinink, E. de Bruijn, et al., Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat. Neurosci.* **19**, 494-503 (2016).

59. B. B. Lake, R. Ai, G. E. Kaeser, N. S. Salathia, Y. C. Yung, R. Liu, et al., Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590 (2016).

60. N. L. Rosenberg, B. L. Kotzin, M. C. Kemp, J. S. Burks, T. J. Santoro, Coronavirus SD-induced immunoregulatory disturbances in a murine model of demyelination. *Advances in experimental medicine and biology* **218**, 441-447 (1987).

61. G. M. Sturgill, G. J. Pauer, E. Bala, E. Simpson, S. S. Yaniglos, J. W. Crabb, et al., Mutation screen of the cone-specific gene, CLUL1, in 376 patients with age-related macular degeneration. *Ophthalmic genetics* **27**, 151-155 (2006).

62. M. A. Gomez-Ferreria, U. Rath, D. W. Buster, S. K. Chanda, J. S. Caldwell, D. R. Rines, D. J. Sharp, Human Cep192 is required for mitotic centrosome and spindle assembly. *Curr Biol* **17**, 1960-1966 (2007).

63. J. Schwenk, N. Harmel, A. Brechet, G. Zolles, H. Berkefeld, C. S. Muller, Wet al., High-resolution proteomics unravel architecture and molecular diversity of native AMPA receptor complexes. *Neuron* **74**, 621-633 (2012).

64. S. J. Sanders, X. He, A. J. Willsey, A. G. Ercan-Sencicek, K. E. Samocha, A. E. Cicek, et al., Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-1233 (2015).

65. V. Agarwal, G. W. Bell, J. W. Nam, D. P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, (2015).

66. Q. Jiang, Y. Wang, Y. Hao, L. Juan, M. Teng, X. Zhang, et al., miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* **37**, D98-104 (2009).

67. C. H. Chou, S. Shrestha, C. D. Yang, N. W. Chang, Y. L. Lin, K. W. Liao, et al., miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**, D296-D302 (2018).

68. P. Arlotta, B. J. Molyneaux, D. Jabaudon, Y. Yoshida, J. D. Macklis, Ctip2 controls the differentiation of medium spiny neurons and the establishment of the cellular architecture of the striatum. *J. Neurosci.* **28**, 622-632 (2008).

69. C. Zhao, X. Li, B. Han, Z. You, L. Qu, C. Liu, J. Song, L. Lian, N. Yang, Gga-miR-219b targeting BCL11B suppresses proliferation, migration and invasion of Marek's disease tumor cell MSB1. *Scientific reports* **7**, 4247 (2017).

70. T. Otani, M. C. Marchetto, F. H. Gage, B. D. Simons, F. J. Livesey, 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. *Cell Stem Cell* **18**, 467-480 (2016).

71. T. Domazet-Loso, D. Tautz, A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815-818 (2010).

72. A. T. Kalinka, K. M. Varga, D. T. Gerrard, S. Preibisch, D. L. Corcoran, J. Jarrells, U. Ohler, C. M. Bergman, P. Tomancak, Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811-814 (2010).

73. H. Zhao, Z. Tu, H. Xu, S. Yan, H. Yan, Y. Zheng, et al., Altered neurogenesis and disrupted expression of synaptic proteins in prefrontal cortex of SHANK3-deficient non-human primate. *Cell Res.* **27**, 1293-1297 (2017).

74. K. S. Saleem, N. Logothetis, *A combined MRI and histology atlas of the rhesus monkey brain in stereotaxic coordinates*. (Academic, London ; Burlington, MA, 2007), pp. ix, 326 p.

75. L. Jiang, F. Schlesinger, C. A. Davis, Y. Zhang, R. Li, M. Salit, T. R. Gingeras, B. Oliver, Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543-1551 (2011).

76. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

77. J. Harrow, F. Denoeud, A. Frankish, A. Reymond, C.-K. Chen, J. Chrast, J. Lagarde, et al., GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7 Suppl 1**, 4 (2006).

78. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).

79. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014);.

80. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621-628 (2008).

81. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).

82. I. Efroni, P. L. Ip, T. Nawy, A. Mello, K. D. Birnbaum, Quantification of cell identity from single-cell gene expression profiles. *Genome Biol.* **16**, 9 (2015).

83. J. Friedman, T. Hastie, R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software* **33**, 1-22 (2010).

84. Z. He, H. Bammann, D. Han, G. Xie, P. Khaitovich, Conserved expression of lincRNA during human and macaque prefrontal cortex development and maturation. *Rna* **20**, 1103-1111 (2014).

85. R. Gaujoux, C. Seoighe, A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).

86. L. Huminiecki, A. T. Lloyd, K. H. Wolfe, Congruence of tissue expression profiles from Gene Expression Atlas, SAGEmap and TissueInfo databases. *BMC Genomics* **4**, 31 (2003).

87. Y. Benjamini, D. Drai, G. Elmer, N. Kafkafi, I. Golani, Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* **125**, 279-284 (2001).

88. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, et al., Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457 (2015).

89. C. E. Rasmussen, C. K. I. Williams, *Gaussian processes for machine learning*. Adaptive computation and machine learning (MIT Press, Cambridge, Mass., 2006), pp. xviii, 248 p.

90. D. Smedley, S. Haider, S. Durinck, L. Pandini, P. Provero, J. Allen, et al., The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589-598 (2015).

91. I. Ezkurdia, D. Juan, J. M. Rodriguez, A. Frankish, M. Diekhans, J. Harrow, J. Vazquez, A. Valencia, M. L. Tress, Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum. Mol. Genet.* **23**, 5866-5878 (2014).

92. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

93. P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719-720 (2008).

94. G. Csardi, T. Nepusz, The igraph software package for complex newtwork research. *Interjournal* **1965**, (2006).

95. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, et al., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589 (2010).

96. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol.* **11**, 106 (2010).

97. A. Reyes, S. Anders, R. J. Weatheritt, T. J. Gibson, L. M. Steinmetz, W. Huber, Drift and conservation of differential exon usage across tissues in primate species. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15377-15382 (2013).

98. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgements: We thank M. Horn, G. Sedmak, M. Pletikos, D. Singh, G. Terwilliger, and S. Wilson for assistance with tissue acquisition and processing. We also thank Alvaro Duque for using equipment from MacBrainResource (NIH/NIMH R01MH113257). **Funding:** Data were 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, P50MH106934 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), 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). Tomas Marques-Bonet is supported by BFU2017-86471-P (MINECO/FEDER, UE), U01 MH106874 grant, Howard Hughes International Early Career, Obra Social "La Caixa" and Secretaria d'Universitats i Recerca and CERCA Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880). Paula Esteller-Cucala is supported by a Formació de Personal Investigador fellowship from Generalitat de Catalunya (FI_B00122). Luis Ferrández-Peral is supported by La Caixa Foundation. David Juan is supported by a Juan de la Cierva fellowship (FJCI-2016-29558) from MICINN. Additional support was provided by the NIH grant MH109904 and MH106874, the Kavli Foundation, the James S. McDonnell Foundation. **Author contributions:** A.M.M.S. and N.S designed the study, and procured and dissected all samples. A.M.M.S and Y.I.K. performed all tissue-level experiments and validations. T.G. and M.S. performed single-cell experiments. M.S., D.J.M, and M.Y performed single-nucleus experiments. Y.Z. developed *TranscriptomeAge* and *TempShift.* Y.Z., M.L., and G.S. analyzed the data. P.E.-C., D.J., L.F.-P., and T.M.-B. performed transcription factor enrichment, differential exon usage analyses and evolutionary conservation analyses. Y.Z, A.M.M.S., F.O.G. and N.S. wrote

the manuscript with input of the other authors. **Competing interests:** Authors have no conflict of interest. **Data and materials availability:** Data is available at NCBI BioProjects (accession number: PRJNA448973) and via Synapse in www.psycoencode.org. All algorithms, packages, and scripts are available at evolution.psychencode.org. Supplement contains additional data.

List of Supplementary Materials:

Materials and Methods

Figs. S1 to S47

Tables S1 to S27

References (*75*-*97*)

Figure Legends

Fig. 1. **Conserved and divergent transcriptomic features of human and macaque neurodevelopmental processes.** (**A**) Plot depicting the real age (x axis) and the age predicted by *TranscriptomeAge* (y axis) of human (red), chimpanzee (blue), and macaque (green). Macaque (164 PCD) and human (266 PCD) births are labeled with green or red dashed line, respectively. (**B**) Schematic showing the human developmental periods, as described in Kang et al., 2011 (*29*) and the matched macaque developmental and chimpanzee adult data sets. Each line corresponds to one macaque or one chimpanzee specimen and the corresponding predicted age when compared to human neurodevelopment. PCD – post-conception day; PY – postnatal year. (**C**) The weight (W) of five transcriptomic signatures in the developing human and macaque NCX and the respective association with neurodevelopmental processes. In signature 1 (neurogenesis), the arrow indicates the point at which the signature reaches the minimum in human (red) and macaque (green). The asterisk indicates the extension of the early fetal period that is also observed in (B), in which early fetal macaques (E60) cluster with mid-fetal humans. In transcription signatures 2, 3, 4, and 5, arrows indicate the point at which the signatures reach the maximum in human (red) or macaque (green). Note that for transcriptomic signatures 2 and 3 (neuronal differentiation and astrogliogenesis) there is a synchrony between human and macaque, whereas for transcriptomic signature 4 and 5 (synaptogenesis and myelination), there is heterochrony between the species, with acceleration in human synaptogenesis and delay in human myelination. Prefrontal cortical areas are plotted in red, primary motor cortex in orange, parietal areas in green, temporal areas in blue, and primary visual cortex in gray. MFC – medial prefrontal

cortex; OFC – orbital prefrontal cortex; DFC – dorsolateral prefrontal cortex; VFC – ventrolateral prefrontal cortex; M1C – primary motor cortex; S1C – primary somatosensory cortex; IPC – inferior posterior parietal cortex; A1C – primary auditory cortex; STC – superior temporal cortex; ITC – inferior temporal cortex; V1C – primary visual cortex. (**D**) Cell type-enrichment is shown for each signature. *P* values adjusted by Benjamini–Hochberg procedure were plotted (size of dots) and significance was labeled by color (True: red and False: gray). H – human; M – macaque; eNPC – embryonic neuroepithelial progenitor; eIPC – embryonic intermediate progenitor cell; eNasN – embryonic nascent neuron; ExN – excitatory neuron; InN – interneuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell; VSMC – vascular smooth muscle cell.

Fig. 2. Ontogenetic inter-regional transcriptomic differences display a cup-shaped pattern in human and macaque. (**A**-**B**) The inter-regional difference of a given module was measured as the average distance of each neocortical area to all other areas in the (**A**) human and (**B**) macaque neocortices across development. The upper quartile inter-regional difference among all genes was plotted and the magnitude was shown in colors. The gray planes represent the transition from prenatal to early postnatal development (late fetal transition) and from adolescence to adult. (**C**) The number of co-expression modules that display gradient-like expression (anterior-posterior, posterior-anterior, medial-lateral, temporal lobe-enriched), and enrichment in primary areas or enrichment in association areas in each developmental phase. Left panel corresponds to human modules, right panel to macaque modules. (**D**) Donut plots depicting the modules (from

(C)) that exhibited species-distinct inter-regional differences. Red indicates high expression of the genes in the module; blue indicates low expression of the genes in the module. Prenatal modules show a human-distinct anterior-posterior expression gradient (left panel); macaque-distinct early postnatal modules show enrichment in primary or association areas (middle panel); and a macaque-distinct adult module is enriched in association areas, especially in MFC (right panel). The expression pattern of each species-distinct module is shown in human (top) and macaque (bottom). HS – Human (*Homo sapiens*) module; MM – macaque (*Macaca mulatta*) module.

Fig. 3. Transcriptomic divergence between human and macaque throughout neurodevelopment reveals a phylogenetic cup-shaped pattern. (**A**) The inter-species divergence, measured as the absolute difference in gene expression, between human and macaque in each brain region throughout development (coded as in Fig 2A). The upper quartile divergence among all genes was plotted. The gray planes represent the transition from prenatal to early postnatal development (late fetal transition, left) and from adolescence to adult (right). (**B**) Venn diagram displaying the number of differentially expressed genes (DEX, top) or genes with differential exon usage (DEU, bottom) between human and macaque in at least one brain region during prenatal development, early postnatal development, and adulthood. (**C**) Bubble matrix with examples of genes showing global or regional inter-species differential expression. Brain regions displaying significant differential expression between human and macaque are shown with black circumference. Red circles show up-regulation in human; blue shows up-regulation in macaque. Circle size indicates absolute $\log_2$ fold change. (**D**) Percentage of overlap

between genes showing the highest inter-species divergence in each region (driving the evolutionary cup-shaped pattern), and genes with the largest pairwise distance between brain regions in prenatal (red), early postnatal (green) and adult (green) human (solid lines) and macaque brains (dashed lines; driving the developmental cup-shaped pattern). The result was plotted using a variable number of the highest ranked genes based on inter-regional difference and the inter-species divergence. Mean and standard deviation (error bar) across regions were plotted.

Fig. 4. Cell-type specificity of species differences. (**A**) Cell type enrichment for genes up- or down-regulated in human neocortical areas. Enrichment of genes up-regulated in human or macaque was tested using single cells from prenatal human neocortex (*33*) or macaque DFC, respectively. The plot shows -log10-P values adjusted by Benjamini–Hochberg procedure averaged across all neocortical areas (NCX), prefrontal areas (PFC) and non-prefrontal areas (nonPFC). Significance was labeled by color (True: red and False: gray). (**B**) Cell type enrichment for genes up- or down-regulated in human neocortical areas. Enrichment of genes up-regulated in human or macaque was tested using single nuclei from adult human neocortex (*33*) or macaque DFC, respectively. *P* values were adjusted by Benjamini–Hochberg procedure and the log-transformed *P* values averaged across all neocortical areas (NCX), prefrontal areas (PFC) and non-prefrontal areas (nonPFC) was plotted (size). Significance was labeled by color (True: red and False: gray). (**C**) Cell type-enrichment of selected genes showing human-distinct up- or down-regulation in adult brain regions or neocortical areas (*34*). Preferential Expression Measure (PEM) (size and color) was plotted to show the cell type specificity.

eNPC – embryonic neuroepithelial progenitor; eIPC – embryonic intermediate progenitor cell; eNasN – embryonic nascent neuron; ExN – excitatory neuron; InN – interneuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell; VSMC – vascular smooth muscle cell.

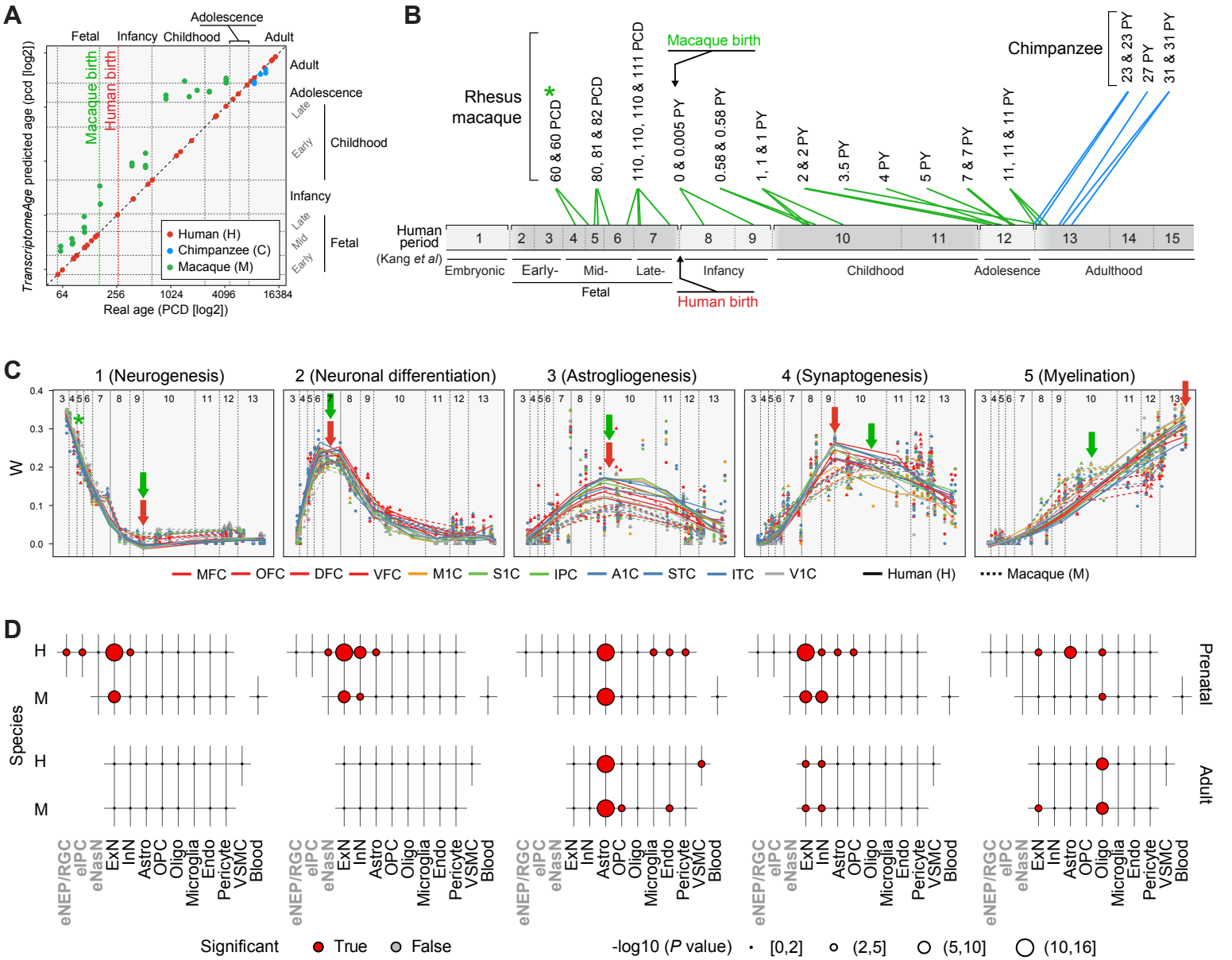
Fig. 5. Shared and divergent transcriptomic features of homologous cell types between human and macaque. (**A**) Dendrogram and heatmap showing diversity and correlation of prenatal cell types within and between the two species. The human single cells were from (*33*). (**B**) Dendrogram and heatmap showing diversity and correlation of adult cell types within and between the two species. (**C**) Cell type specificity of inter-species differentially expressed genes based on the single-cell/nucleus information. Blue, human down-regulated genes; and red, human up-regulated genes.

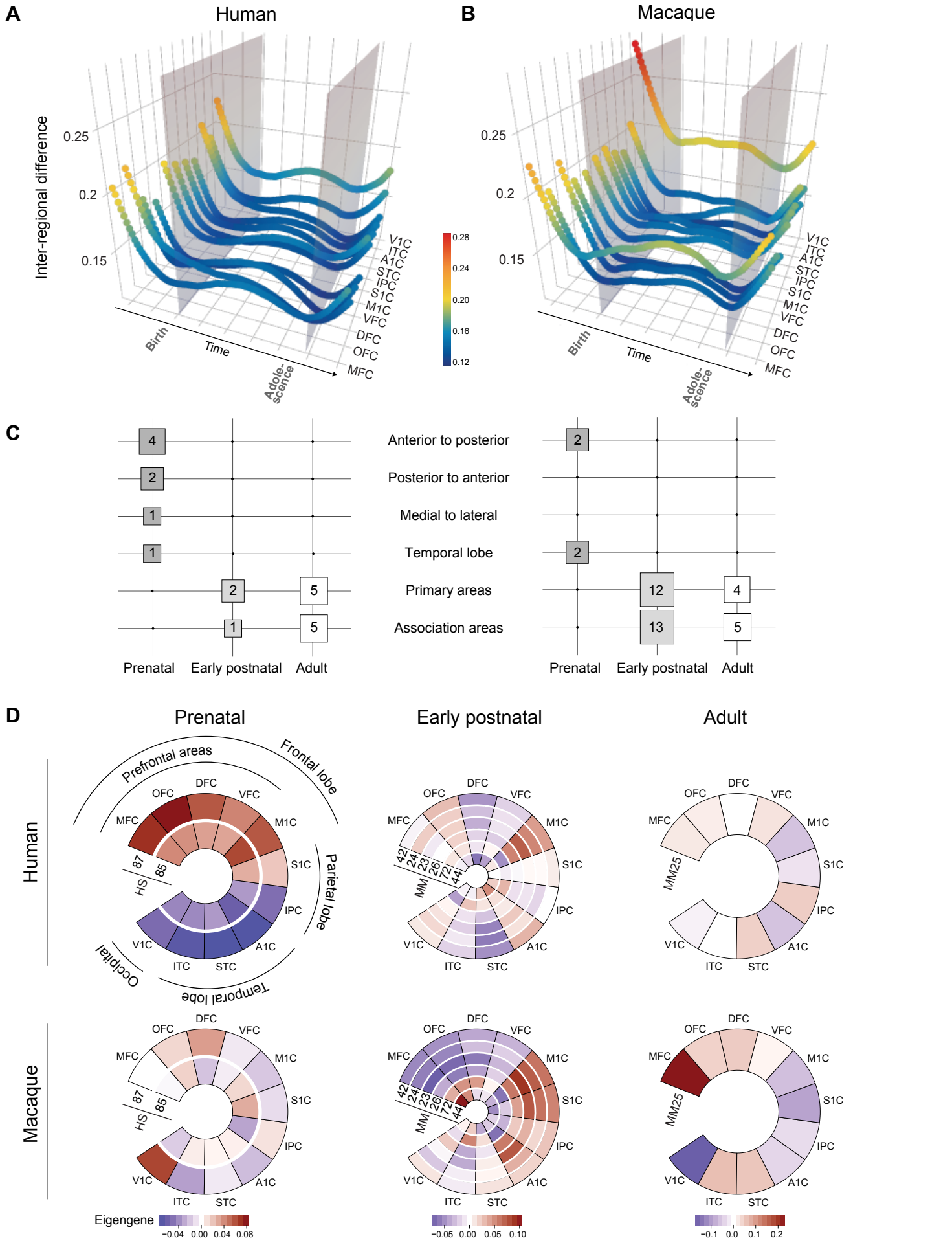
Fig. 6. Heterochronic expression of regional and inter-species gene clusters. (**A**) Clusters of genes exhibiting species-distinct regional heterochronic expression patterns in human and macaque brains at various prenatal periods and adulthood. The timing of expression of genes in the cluster is represented by blue (earlier expression) to red (later expression). Prenatal heterochronic regional clusters RC21 and RC34 show earlier expression (blue) in human prenatal fronto-parietal perisylvian neocortical areas (M1C, S1C, and IPC) and enrichment in neural progenitors. RC10 is composed of genes with earlier expression in the human prenatal prefrontal cortex and enrichment in astrocytes. These observed regional expression patterns are not present in the macaque prenatal

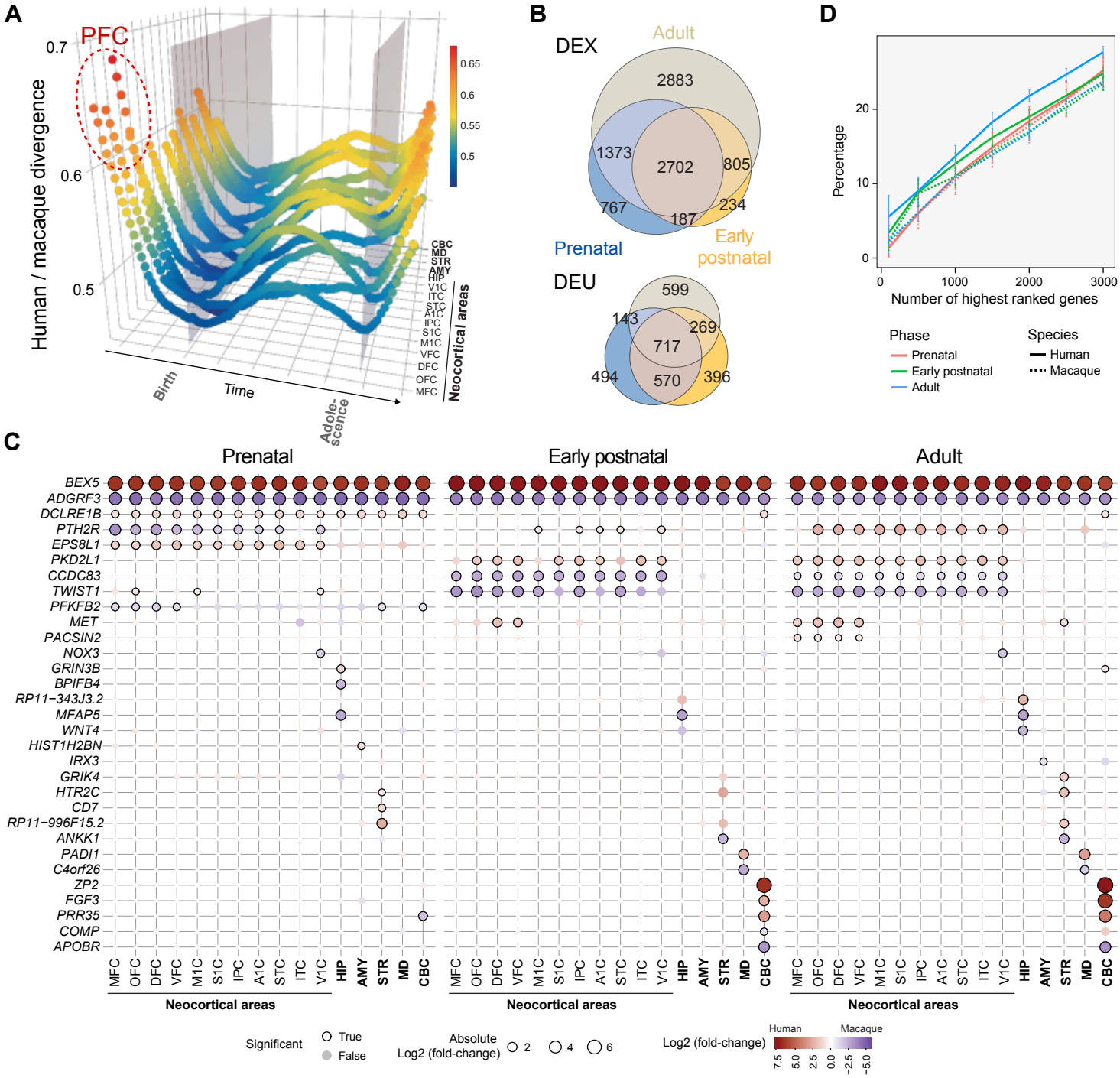
NCX. Adult heterochronic cluster RC25 shows earlier expression in primary areas of the macaque cortex and enrichment for genes associated with oligodendrocytes. (**B**) A network of 139 inter-species heterochronic genes (blue) is enriched for targets of putative upstream transcriptional regulators that include those encoded by eight genes of the same network (red), and TWIST1 (green), a transcription factor with inter-species heterotopic expression (fig. S34). Arrows indicate direction of regulation. (**C**) Top five canonical pathways enriched among inter-species heterochronic genes in at least one neocortical area. (**D**) Cluster EC14 shows inter-species heterochronic expression, exhibits a delayed expression specifically in the human prenatal prefrontal cortex, and is enriched for genes selectively expressed by intermediate progenitor cells (IPC).

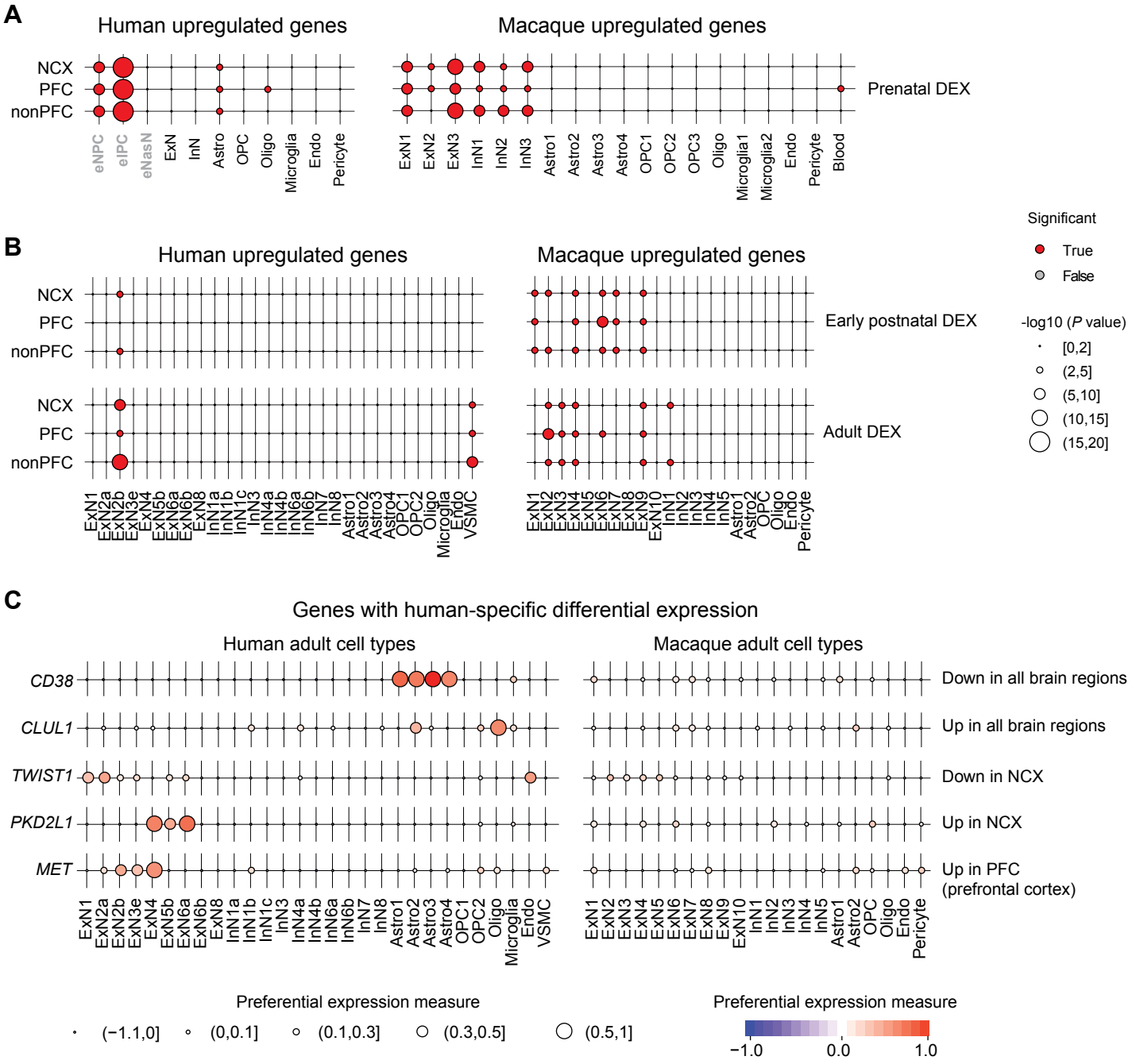
Fig. 7. Heterotopic and/or heterochronic expression of disease-associated genes between human and macaque. (**A**) Bar plot depicting the number of genes associated with autism spectrum disorder (ASD; high confidence [hc]), neurodevelopmental disorders (NDD), attention deficit hyperactivity disorder (ADHD), schizophrenia (SCZ), bipolar disorder (BD), major depressive disorder (MDD), Alzheimer's disease (AD), and Parkinson's disease (PD) that display heterochronic divergence between human and macaque. (**B**) Bubble matrix showing the heterochronic expression of ASD and SCZ-associated genes. Blue represents earlier expression in human; red represents earlier expression in macaque. (**C**) Bar plot depicting the number of genes associated with neuropsychiatric disorders that exhibit heterotopic divergence between human and macaque. The 14 SCZ-associated genes that displayed heterotopy were grouped into 5 clusters on the basis of their spatio-temporal expression profiles (fig. S41). **(D)** Donut

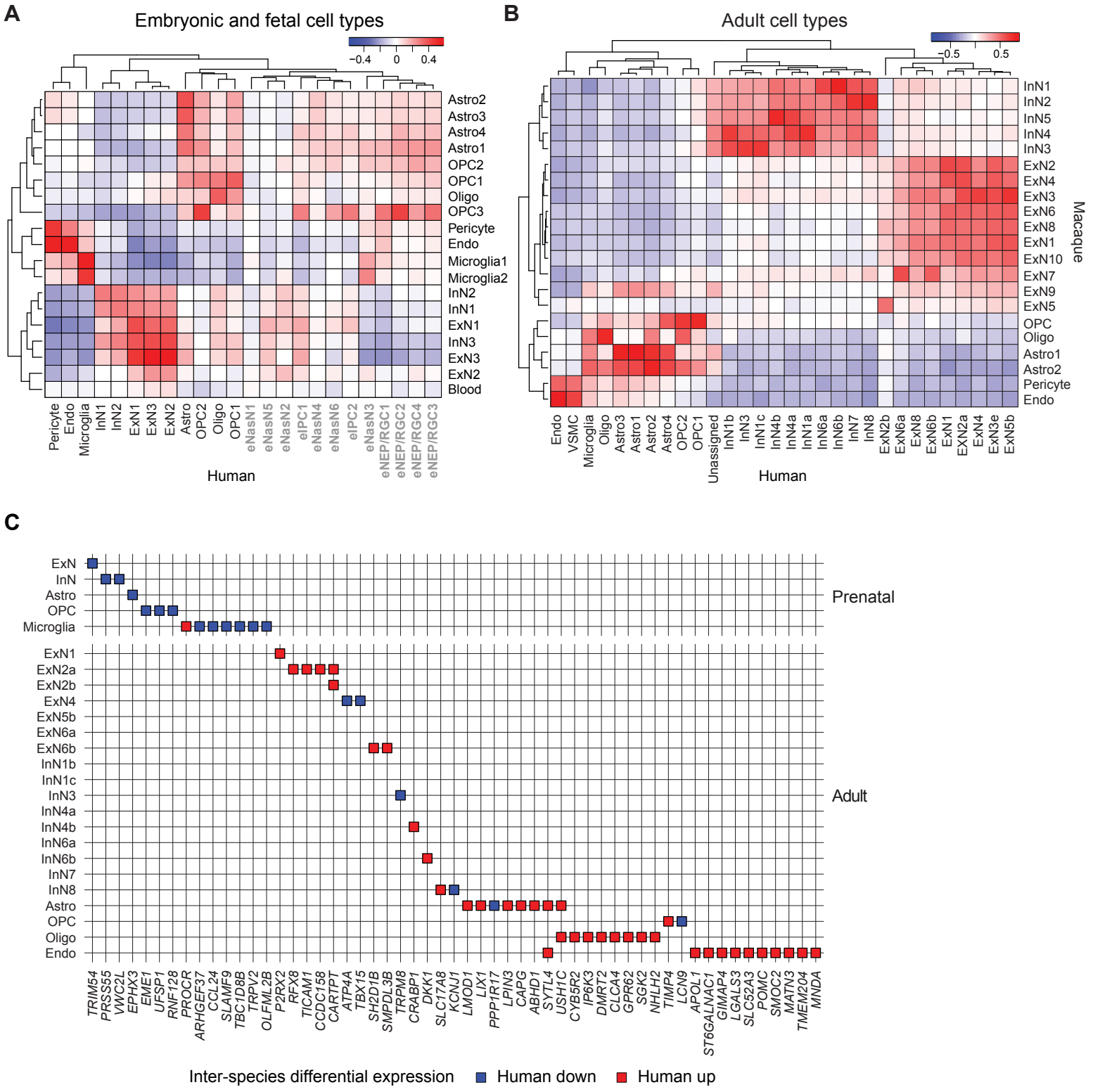
plots exhibiting the centered expression of the 5 SCZ-associated heterotopic clusters in prenatal, early postnatal development, and adulthood. Cluster numbers are labeled with the same color as in panel C. Clusters that are not significantly divergent between species in each period are grey and do not have a black border. Red indicates high expression; blue indicates low expression.

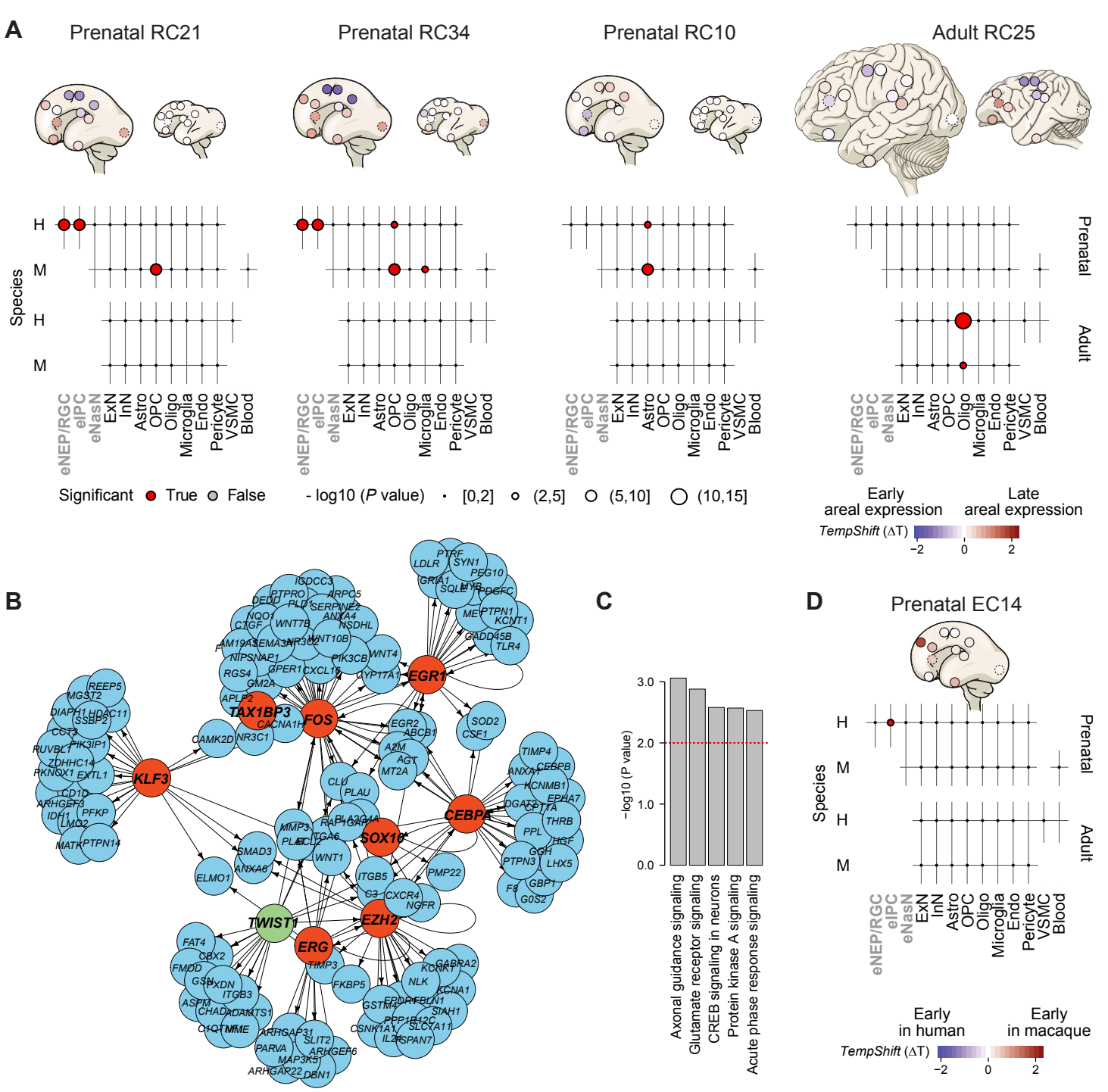

A
Adolescence
Fetal
Infancy
Childhood
Adult
Macaque birth
Human birth
TranscriptomeAge predicted age (pcd [log2])
Real age (PCD [log2])
64
256
1024
4096
16384
Human (H)
Chimpanzee (C)
Macaque (M)
Adult
Adolescence
Late
Early
Childhood
Infancy
Late
Mid
Early
Fetal
B
Rhesus macaque
Chimpanzee
Macaque birth
Human birth
60 & 60 PCD
80, 81 & 82 PCD
110, 110, 110 & 111 PCD
0 & 0.005 PY
0.58 & 0.58 PY
1, 1 & 1 PY
2 & 2 PY
3.5 PY
4 PY
5 PY
7 & 7 PY
11, 11 & 11 PY
23 & 23 PY
27 PY
31 & 31 PY
Human period (Kang et al)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Embryonic
Early-
Mid-
Late-
Fetal
Infancy
Childhood
Adolesence
Adulthood
C
1 (Neurogenesis)
2 (Neuronal differentiation)
3 (Astrogliogenesis)
4 (Synaptogenesis)
5 (Myelination)
W
MFC OFC DFC VFC M1C S1C IPC A1C STC ITC V1C
Human (H)
Macaque (M)
D
Species
H
M
Prenatal
Adult
eNEP/RGC
eIPC
eNasN
ExN
InN
Astro
OPC
Oligo
Microglia
Endo
Pericyte
VSMC
Blood
Significant
True
False
-log10 (P value)
[0,2]
(2,5]
(5,10]
(10,16]


A
Human
B
Macaque
Inter-regional difference
0.25
0.2
0.15
Birth
Time
Adole-scence
V1C
ITC
A1C
STC
IPC
S1C
M1C
VFC
DFC
OFC
MFC
0.28
0.24
0.20
0.16
0.12
C
Anterior to posterior
Posterior to anterior
Medial to lateral
Temporal lobe
Primary areas
Association areas
Prenatal
Early postnatal
Adult
D
Prenatal
Early postnatal
Adult
Human
Macaque
Prefrontal areas
Frontal lobe
Parietal lobe
Occipital
Temporal lobe
HS
MM
MM25
Eigengene
−0.04 0.00 0.04 0.08
−0.05 0.00 0.05 0.10
−0.1 0.0 0.1 0.2


A
PFC
Human / macaque divergence
0.7
0.6
0.5
Birth
Time
Adole-scence
Neocortical areas
CBC
MD
STR
AMY
HIP
V1C
ITC
STC
A1C
IPC
S1C
M1C
VFC
DFC
OFC
MFC
B
DEX
Adult
2883
1373
2702
805
767
187
234
Prenatal
Early postnatal
DEU
599
143
269
717
494
570
396
D
Percentage
Number of highest ranked genes
Phase
Prenatal
Early postnatal
Adult
Species
Human
Macaque
C
Prenatal
Early postnatal
Adult
BEX5
ADGRF3
DCLRE1B
PTH2R
EPS8L1
PKD2L1
CCDC83
TWIST1
PFKFB2
MET
PACSIN2
NOX3
GRIN3B
BPIFB4
RP11−343J3.2
MFAP5
WNT4
HIST1H2BN
IRX3
GRIK4
HTR2C
CD7
RP11−996F15.2
ANKK1
PADI1
C4orf26
ZP2
FGF3
PRR35
COMP
APOBR
MFC OFC DFC VFC M1C S1C IPC A1C STC ITC V1C HIP AMY STR MD CBC
Neocortical areas
Significant
True
False
Absolute Log2 (fold-change)
Log2 (fold-change)
Human
Macaque


A
Human upregulated genes
Macaque upregulated genes
NCX
PFC
nonPFC
Prenatal DEX
eNPC
eIPC
eNasN
ExN
InN
Astro
OPC
Oligo
Microglia
Endo
Pericyte
ExN1
ExN2
ExN3
InN1
InN2
InN3
Astro1
Astro2
Astro3
Astro4
OPC1
OPC2
OPC3
Oligo
Microglia1
Microglia2
Endo
Pericyte
Blood
B
Human upregulated genes
Macaque upregulated genes
Early postnatal DEX
Adult DEX
Significant
True
False
-log10 (P value)
[0,2]
(2,5]
(5,10]
(10,15]
(15,20]
C
Genes with human-specific differential expression
Human adult cell types
Macaque adult cell types
CD38
CLUL1
TWIST1
PKD2L1
MET
Down in all brain regions
Up in all brain regions
Down in NCX
Up in NCX
Up in PFC
(prefrontal cortex)
Preferential expression measure
(−1.1,0]
(0,0.1]
(0.1,0.3]
(0.3,0.5]
(0.5,1]
Preferential expression measure
−1.0
0.0
1.0

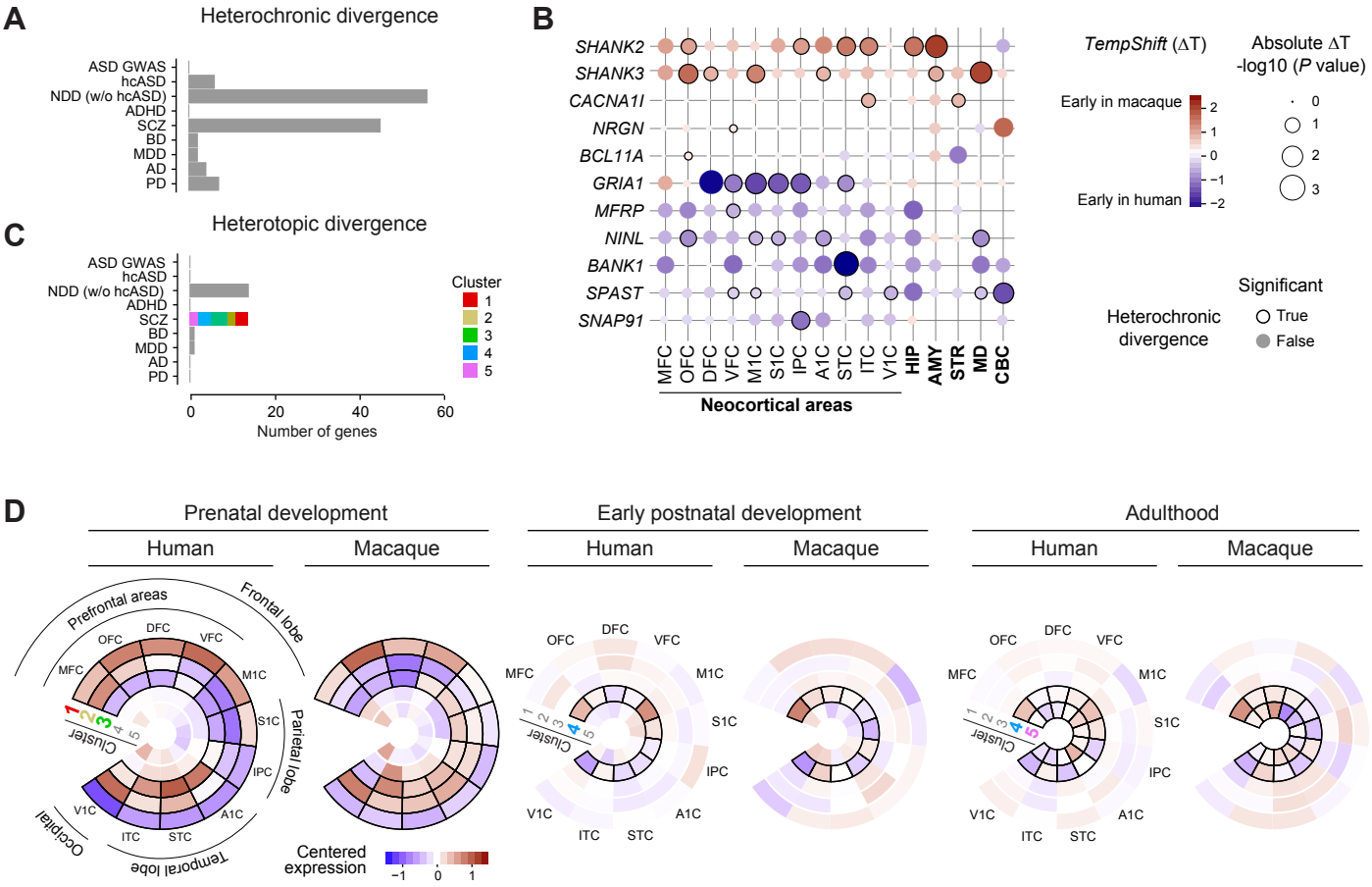

A
Embryonic and fetal cell types
−0.4
0
0.4
Astro2
Astro3
Astro4
Astro1
OPC2
OPC1
Oligo
OPC3
Pericyte
Endo
Microglia1
Microglia2
InN2
InN1
ExN1
InN3
ExN3
ExN2
Blood
Pericyte
Endo
Microglia
InN1
InN2
ExN1
ExN3
ExN2
Astro
OPC2
Oligo
OPC1
eNasN1
eNasN5
eNasN2
eIPC1
eNasN4
eNasN6
eIPC2
eNasN3
eNEP/RGC1
eNEP/RGC2
eNEP/RGC4
eNEP/RGC3
Human
B
Adult cell types
−0.5
0.5
InN1
InN2
InN5
InN4
InN3
ExN2
ExN4
ExN3
ExN6
ExN8
ExN1
ExN10
ExN7
ExN9
ExN5
OPC
Oligo
Astro1
Astro2
Pericyte
Endo
Macaque
Endo
VSMC
Microglia
Oligo
Astro3
Astro1
Astro2
Astro4
OPC2
OPC1
Unassigned
InN1b
InN3
InN1c
InN4b
InN4a
InN1a
InN6a
InN6b
InN7
InN8
ExN2b
ExN6a
ExN8
ExN6b
ExN1
ExN2a
ExN4
ExN3e
ExN5b
Human
C
ExN
InN
Astro
OPC
Microglia
Prenatal
ExN1
ExN2a
ExN2b
ExN4
ExN5b
ExN6a
ExN6b
InN1b
InN1c
InN3
InN4a
InN4b
InN6a
InN6b
InN7
InN8
Astro
OPC
Oligo
Endo
Adult
TRIM54
PRSS55
VWC2L
EPHX3
EME1
UFSP1
RNF128
PROCR
ARHGEF37
CCL24
SLAMF9
TBC1D8B
TRPV2
OLFML2B
P2RX2
RFX8
TICAM1
CCDC158
CARTPT
ATP4A
TBX15
SH2D1B
SMPDL3B
TRPM8
CRABP1
DKK1
SLC17A8
KCNJ1
LMOD1
LIX1
PPP1R17
LPIN3
CAPG
ABHD1
SYTL4
USH1C
CYB5R2
IP6K3
DMRT2
CLCA4
GPR62
SGK2
NHLH2
TIMP4
LCN9
APOL1
ST6GALNAC1
GIMAP4
LGALS3
SLC52A3
POMC
SMOC2
MATN3
TMEM204
MNDA
Inter-species differential expression
Human down
Human up


A
Prenatal RC21
Prenatal RC34
Prenatal RC10
Adult RC25
Species
H
M
Prenatal
Adult
eNEP/RGC
eIPC
eNasN
ExN
InN
Astro
OPC
Oligo
Microglia
Endo
Pericyte
VSMC
Blood
Significant
True
False
- log10 (P value)
[0,2]
(2,5]
(5,10]
(10,15]
Early areal expression
Late areal expression
TempShift (ΔT)
−2
0
2

B
KLF3
TAX1BP3
FOS
EGR1
CEBPB
SOX16
TWIST1
EZH2
ERG

C
−log10 (P value)
3.0
2.0
1.0
0.0
Axonal guidance signaling
Glutamate receptor signaling
CREB signaling in neurons
Protein kinase A signaling
Acute phase response signaling

D
Prenatal EC14
Species
H
M
Prenatal
Adult
eNEP/RGC
eIPC
eNasN
ExN
InN
Astro
OPC
Oligo
Microglia
Endo
Pericyte
VSMC
Blood
Early in human
Early in macaque
TempShift (ΔT)
−2
0
2


A
Heterochronic divergence
ASD GWAS
hcASD
NDD (w/o hcASD)
ADHD
SCZ
BD
MDD
AD
PD
C
Heterotopic divergence
Cluster
1
2
3
4
5
0
20
40
60
Number of genes
B
SHANK2
SHANK3
CACNA1I
NRGN
BCL11A
GRIA1
MFRP
NINL
BANK1
SPAST
SNAP91
MFC
OFC
DFC
VFC
M1C
S1C
IPC
A1C
STC
ITC
V1C
HIP
AMY
STR
MD
CBC
Neocortical areas
TempShift (ΔT)
Early in macaque
Early in human
Absolute ΔT
-log10 (P value)
Significant
Heterochronic divergence
True
False

D
Prenatal development
Early postnatal development
Adulthood
Human
Macaque
Prefrontal areas
Frontal lobe
Parietal lobe
Occipital
Temporal lobe
Cluster
Centered expression

Supplementary Materials for

Spatiotemporal transcriptomic divergence across human and macaque brain development

Ying Zhu[1,2,†], André M. M. Sousa[1,†], Tianliuyun Gao[1,†], Mario Skarica[1,†], Mingfeng Li[1,†], Gabriel Santpere[1], Paula Esteller-Cucala[3], David Juan[3], Luis Ferrández-Peral[3], Forrest O. Gulden[1], Mo Yang[1], Daniel J. Miller[1], Tomas Marques-Bonet[3,4,5,6], Yuka Imamura Kawasawa[7], Hongyu Zhao[2], Nenad Sestan[1,8,*]

Correspondence to: Nenad Sestan (nenad.sestan@yale.edu)

This PDF file includes:

Materials and Methods

Figs. S1 to S47

Captions for Tables S1 to S27

Other Supplementary Materials for this manuscript includes the following:

Table S1 to S27 (Excel)

Materials and Methods

Postmortem human and non-human primate specimens

All human tissue was procured and processed as described in Li et al. 2018 (*33*).

Tissue procurement: Rhesus macaque (*Macaca mulatta*) brain samples were collected postmortem from 26 specimens ranging from 60 postconceptional days (PCD) to adulthood. All experiments using non-human primates were carried out in accordance with a protocol approved by Yale University's Committee on Animal Research and NIH guidelines.

Neuropathological evaluation of macaque brains: All clinical histories, tissue specimens, and histological sections were evaluated to assess for signs of disease, injury, and gross anatomical and histological alterations. Tissue samples from the brain specimens analyzed in the study were fixed in 4% paraformaldehyde and processed for histological examination using Nissl stain and immunohistochemistry. No obvious signs of neuropathological alterations were observed in any of the macaque specimens analyzed in this study.

Selection criteria for brain specimens. To better ensure consistency between samples and decrease potential variation due to ante- and postmortem conditions, specific selection criteria were established. We collected tissue specimens from macaques that were not used for brain invasive procedures and had no signs of malformations or lesions. Macaques were tested negative for herpes B virus and tuberculosis.

Tissue dissection for tissue-level sequencing: Sixteen homologous brain regions to the human Capstone Developmental study (*33*) were sampled from fresh macaque brains using the same protocol and anatomical landmarks described in "A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates" (*74*). Briefly, brains were chilled on ice for 15–30 minutes prior to sectioning. Brains were placed ventral side up onto a chilled aluminum plate (1 cm-thick) on ice. The brainstem and cerebellum were removed from the cerebrum by making a transverse cut at the junction between the diencephalon and midbrain. Next, they were sectioned to obtain 1 cm thick serial, coronal sections. The tissue slabs were snap frozen in isopentane (J.T. Baker) on dry ice (-78.5ºC) and stored at -80ºC.

Histological verification of tissue sampling: To verify that the brain region or neocortical area of interest is properly and consistently sampled, we occasionally also collected small tissue blocks, adjacent to the tissue sample dissected for the RNA extraction. These tissue blocks were then fixed in 4% paraformaldehyde for 48 h, sectioned on a vibratome or cryostat, and Nissl-stained to verify the identities of dissected adjacent tissue.

RNA isolation, library preparation, sequencing, and quality assessment

Tissue processing: Dissected frozen tissue samples were pulverized in liquid nitrogen using a ceramic mortar and pestle (Coorstek #60322, #60323). Pulverized samples were transferred to chilled wide-mouth cryogenic vials (Nalgene, Thermo Fisher Scientific #5005-0015) and stored at -80ºC until used for RNA extraction.

RNA extraction: Total RNA was extracted using RNeasy Plus Mini Kit (Qiagen #74136) or mirVana kit (Thermo Fisher Scientific #AM1560) (table S1) with some modifications. A bead mill homogenizer (Bullet Blender, Next Advance #BBX24B) was used to lyse the pulverized tissue. Each pulverized tissue sample was transferred to a chilled safe-lock microcentrifuge tube (Eppendorf #022600028). A mass of chilled stainless steel beads (Next Advance #SSB14B) equal to the mass of the tissue was added to the tube. Two volumes of lysis buffer were added to the tissue and beads. Samples were mixed in a Bullet Blender for 1 min at a speed of 6. Samples were visually inspected to confirm desired homogenization and then incubated at 37°C for 5 min. The lysis buffer was added up to 0.6 ml, and samples were mixed in the Bullet Blender for 1 min. Each RNA sample was subjected to a DNase treatment (TURBO DNase, Thermo Fisher Scientific #AM2239) as per manufacturer's instructions. Optical density values of extracted RNA were measured using fluorospectrometer NanoDrop 3300 (Thermo Fisher Scientific #ND-3300) to confirm an A260:A280 ratio above 1.9. RIN was determined for each sample by using automated electrophoresis kit (Agilent RNA 6000 Nano Kit, Agilent Technologies #5067-1511) on 2100 Bioanalyzer (Agilent Technologies #G2939BA)

Library preparation for mRNA-sequencing: cDNA libraries were prepared using the TruSeq RNA Sample Prep Kit (Illumina: #FC-122-1001, #FC-122-1002) as per the manufacturer's instructions with some modifications. Briefly, polyA RNA was purified from 1 to 5 μg of total RNA using (dT) beads. Quaint-IT RiboGreen RNA Assay Kit

(Thermo Fisher Scientific #R11490) was used to quantitate purified mRNA with the NanoDrop 3300. Following mRNA quantitation, 2.5 µl spike-in master mixes, containing five different types of RNA molecules at varying amount (2.5×10^{-7} to 2.5×10^{-14} mol), were added per 100 ng of mRNA (*75*). The spike-in RNAs were synthesized by External RNA Control Consortium (ERCC) by *in vitro* transcription of *de novo* DNA sequences or of DNA derived from the *B. subtilis* or the deep-sea vent microbe *M. jannaschii* genomes (*34*) and were a generous gift of Dr. Mark Salit at The National Institute of Standards and Technology (NIST). These were used both to track the brain regions and to normalize expression levels across experiments. Each sample was tagged by adding a pair of spike-in RNAs unique to the region from which the sample was taken. Also, an additional three common spike-ins were added for controlling sequencing error rates, which is not influenced by SNP existence (*34*). The mixture of mRNA and spike-in RNAs were subjected to fragmentation, reverse transcription, end repair, 3'-end adenylation, and adapter ligation to generate libraries of short cDNA molecules. The libraries were size selected at 200-250 bp by gel excision, followed by PCR amplification and column purification. The TruSeq Kit allows addition of 12 different types of index, thus the multiplexing of samples when they are sequenced. The final product was assessed for its size distribution and concentration using Agilent DNA 1000 Kit (Agilent Technologies #5067-1504) on 2100 Bioanalyzer (Agilent Technologies #G2939BA).

Sequencing: We used Genome Analyzer IIx (Illumina GAIIx) for mRNA-sequencing by loading one sample per lane, or HiSeq 2000 (Illumina) by loading 4 samples per lane (table S1). For GAIIx sequencing, the library was diluted to 10 nM in EB buffer and then

denatured using the Illumina protocol. The denatured libraries were diluted to 12 pM, followed by cluster generation on a single-end Genome Analyzer IIx (GAIIx) flow cell (v4) using an automated clonal amplification system (Illumina cBOT), according to the manufacturer's instructions. The Illumina GAIIx flow cell was run for 75 cycles using a single-read recipe (v4 sequencing kits) according to the manufacturer's instructions. For HiSeq 2000 sequencing, the library was diluted to 10 nM in EB buffer and then denatured using the Illumina protocol. The denatured libraries were diluted to 15 pM, followed by cluster generation on a single-end HiSeq flow cell (v1.5) using an Illumina cBOT, according to the manufacturer's instructions. The HiSeq flow cell was run for 75 cycles using a single-read recipe (v2 sequencing kit) according to the manufacturer's instructions.

mRNA sequencing alignment, annotation, and expression quantification

Reference genomes and read alignment: The export files generated from Illumina's CASAVA pipeline were converted to fastq files. Reads were then mapped to the human (*Homo sapiens*, Dec. 2013 hg38) and rhesus macaque genomes (*Macaca mulatta*, Nov. 2015 rheMac8), respectively, using STAR (*76*) without providing junction annotation.

Generation of inter-species mRNA annotation: To minimize biases in comparative data analyses due to the disparate quality of gene annotation for the two species, we used the *XSAnno* pipeline, which was developed and independently tested by our group (*37*) to generate annotation of human-macaque orthologs, based on human Gencode v25 (*77*). The *XSAnno* pipeline incorporated whole genome alignment and local alignment with

multiple filters. The liftOver (*78*) parameter "-minMatch" was set to 0.913, based on bootstrapping. The inter-species PID and PL were both set to 0.9.

Expression quantification: The number of reads uniquely mapped to each gene were counted using featureCounts (*79*). The gene expression levels were quantified by read count and RPKM (Reads Per Kilobase per Million mapped reads) (*80*).

Quality control assessment and detection of outliers: Several quality control measures were implemented throughout sample preparation and transcriptome data generation steps as described in Li et al (*33*). The number and the percentage of uniquely mapped reads were also calculated for each sample. A sample passes quality control, only if it meets the following three criteria: 1) the number of uniquely mapped reads is greater than 5 million; 2) the percentage of uniquely mapped reads is greater than 40%; and 3) the percentage of uniquely mapped reads that align to exons is greater than 40%. After filtering, we additionally assessed the sample quality by principal component analysis (PCA) and hierarchical clustering. No obvious outliers and batch effects were detected (figs. S3 and S4).

Single-cell/nucleus RNA sequencing, quality assessment, and cell type assignment

Dissection and single-cell dissociation: We dissected dorsal frontal cortex (DFC), hippocampus (HIP), amygdala (AMY), striatum (STR), mediodorsal nucleus of thalamus (MD) and cerebellar cortex (CBC) from two E110 Rhesus macaques on ice. The dissected tissues were transferred immediately to warm (37 ℃) papain-protease-dispase

(PPD) solution in 1.5ml conical tubes separately and incubated for 30 min. Tissue was gently pipetted every 5-10 minutes during the incubation until tissue pieces were no longer visible. Then the cell suspension was centrifuged at 300 x *g* for 3 minutes and re-suspended in 1x PBS (Gibco #14190) with 0.04% BSA (Sigma #SRE0036). After re-suspension, the cell suspension was filtered into 5 ml polystyrene tube with 40 um filter. Cell concentrations were measured using both automated cell counter (Countess, Thermo Fisher Scientific #C10227) and hemocytometer and adjusted accordingly. The cell suspension concentrations were firstly adjusted to 500~1000 cells/µl and placed on ice.

Brain Cell Nuclei Isolation: Dorsolateral prefrontal cortex was dissected from three frozen adult macaque brains (table S2). In order to avoid experimental bias for cell nuclei isolation, whole tissue was finely pulverized to powder in liquid nitrogen with mortar and pestle (Coorstek #60316, #60317). All buffers were ice-cold and all reagents used for consequent nuclear isolation were molecular biology grade unless stated otherwise. 50 - 100 mg of pulverized tissue was added into 5 ml of ice-cold lysis buffer: 320 mM sucrose (Sigma #S0389), 5 mM $CaCl_2$ (Sigma #21115), 3 mM $Mg(Ace)_2$ (Sigma #63052), 10mM Tris-HCl (pH 8) (AmericanBio #AB14043), protease inhibitors w/o EDTA (Roche #11836170001), 0.1 mM EDTA (AmericanBio #AB00502), RNAse inhibitor (80U/ml) (Roche #03335402001), 1mM DTT (Sigma #43186), 0.1% TX-100 (v/v) (Sigma #T8787). Reagents: DTT, RNAse Protector, protease inhibitors, TX-100 were added immediately before use. The suspension was transferred to Dounce tissue grinder (15ml volume, Wheaton #357544; autoclaved, RNAse free, ice-cold) and homogenized with loose and tight pestles, 30 cycles each, with constant pressure and without introduction of

air. The homogenate was strained through 40 um tube top cell strainer (Corning #352340) which was pre-wetted with 1ml isolation buffer: 1800 mM sucrose (Sigma #S0389), 3 mM $Mg(Ace)_2$ (Sigma #63052), 10mM Tris-HCl (pH 8) (AmericanBio #AB14043), protease inhibitors w/o EDTA (Roche #11836170001), RNAse inhibitor (80U/ml) (Roche #03335402001), 1mM DTT (Sigma #43186). Additional 9 ml of isolation buffer was added to wash the strainer. Final 15 ml of solution was mixed by inverting the tube 10x and carefully pipetted into 2 ultracentrifuge tubes (Beckman Coulter #344059) onto the isolation buffer cushion (5 ml) without disrupting the phases. The tubes were centrifuged at 30000 x *g*, for 60 min at 4 ˚C on ultracentrifuge (Beckman L7-65) and rotor (Beckman SW41-Ti). Upon end of ultracentrifugation, the supernatant was carefully and completely removed and 100 ul of resuspension buffer (250 mM sucrose (Sigma #S0389), 25 mM KCl (Sigma #60142), 5mM MgCl2 (Sigma #M1028), 20mM Tris-HCl (pH 7.5) (AmericanBio #AB14043; Sigma #T2413), protease inhibitors w/o EDTA (Roche #11836170001), RNAse inhibitor (80U/ml) (Roche #03335402001), 1mM DTT (Sigma #43186)) was added dropwise on the pellet in each tube and incubated on ice for 15 minutes. Pellets were gently dissolved by pipetting 30x with 1ml pipette tip, pooled and filtered through 40 um tube top cell strainer (Corning #352340). Finally, nuclei were counted on hemocytometer and diluted to 1 million/ml with sample-run buffer: 0.1% BSA (Gemini Bio-Products #700-106P), RNAse inhibitor (80U/ml) (Roche #03335402001), 1mM DTT (Sigma #43186) in DPBS (Gibco #14190).

Single cell and/or nucleus microfluidic capture and cDNA synthesis: The cells and nuclei samples were placed on ice and taken to Yale Center for Genome Analysis core

facility and processed within 15 minutes for single cell and/or nucleus RNA sequencing with targeted cell and nuclei recovery of 10000 cells and/or nuclei, respectively, on microfluidic Chromium System (10x Genomics) by following precisely manufacturers detailed protocol (10x Genomics, CG00052_RevE, https://support.10xgenomics.com/permalink/3vzDu3zQjY0o2AqkkkI4CC) with Chromium Single Cell 3' Library & Gel Bead Kit v2, (10x Genomics #120237) and Chromium Single Cell A Chip Kit (10x Genomics #120236) on Chromium Controller (10x genomics). Specifically, due to limitations imposed by source RNA quantity, cell derived cDNA was amplified for 10 cycles while cDNA from nuclei was amplified for 14 cycles.

Single cell and/or nucleus RNA-seq library preparation: The post cDNA amplification cleanup and construction of sample-indexed libraries and their amplification subsequently precisely followed manufacturer's directions (10x Genomics, CG00052_RevE). Specifically, the amplification step directly depended on the quantity of input cDNA and it varied form 8 – 14 cycles.

Sequencing, reads alignment and gene expression quantification

Sequencing of libraries: In order to reach optimal sequencing depth (30000 - 50000 raw reads per cell and/or nucleus), single cell and/or nucleus libraries were run using paired-end sequencing with single indexing on the HiSeq 4000 platform (Illumina) by following manufacturer's instructions (Illumina, 10x Genomics: CG00052_RevE,

CG000089_RevB). To avoid lane bias, multiple uniquely indexed samples were mixed and distributed over several lanes.

Processing of single cell and single nuclei RNA sequencing: The commercial software, *cellranger*, was employed to implement the preliminary data processing for single cell and single nuclei RNA sequencing generated from 10X genomic platform. In brief, the cellranger mkfastq converted binary base call (BCL) files to FASTQ files and simultaneously decoded the multiplexed samples, and then cellranger count took FASTQ files performed sequencing alignment against reference genome, quality filtering of low quality reads, cell barcode counting and filtering, and unique molecular identifier (UMI) counting for each annotated genes. We used the default parameters, except for customizing the number of expected cell number to 10,000 and constructing the "pre-mRNA" reference for single nuclei RNA-seq. We utilized Mmul_8.0.1 reference genome and Ensembl 90 annotation for monkey, and utilized GRCh38/hg38 reference genome and GENCODE v21 annotation for human. We did not use annotation generated by XSAnno, as UMI requires complete gene model and is less dependent on the mappability.

Classification of cell types and subtypes: The sparse expression matrix generated from *cellranger* analysis pipeline served as the only input to the *Seurat* software (*81*) and other accompanied software for the classification of cell type and subtype. To optimize this cell classification analysis and reduce the unwanted technical variance at the very beginning, we firstly did the quality filtering, normalization, and data scaling according to the

recommendations from *Seurat's* guideline. For an example, we excluded the cell with the number of expressed genes less than 300 or more than 4,000, or with the number of UMI less than 300 or more than 15,000, or the percentage of mitochondria reads more than 15%. The normalization method was "LogNormalize" and the scale factor was 10,000. In addition, the linear regression was performed by choosing the percentage of mitochondria reads as variable. After purifying the expression matrix, the 100 principle components were computed and their significances were tested by *JackStraw* and visualized by *Elbow plot,* with the result of finally choosing the top 25 principle components for tSNE and clustering analyses. It is worth to note that we used two different resolution parameters, i.e., 1 and 3, to gain the different resolution levels of cell clusters, which had assisted the assignment of cell subtype.

Next, we used the R scripts *SpecScore.R* (*82*) to compute the specificity score for each gene in each cell cluster. Since the cell cluster mostly corresponds to the cell subtype, we computed the correlation matrix to study the relationship between cell clusters by using the gene specificity score. We manually checked the hierarchical clusters and the top ranked genes in each cell cluster to determine which cell clusters could be assigned to the same cell type. After we determined the cell type, we re-computed the specificity score for each gene to choose the top ranked gene as the cell type signature markers in other downstream analysis. Under each cell type, the cell clusters were assigned to different cell subtypes. The inconsistency between low and high resolution were assigned to unknown, because a very small number of inconsistent cells were observed. Lastly, we recruited the well-known gene markers to verify the assignment of cell types and subtypes and the inconsistent cells were assigned to the

unknown category. For the adult human single-nucleus RNA-seq, the subtypes of excitatory neurons and interneurons were defined according to Lake et al. (*59*), by correlating the expression of marker genes.

Visualization of cell type alignment between species: To perform cross-species comparison between human and monkey species, we divided this analysis in three steps. Firstly, we filtered the gene annotation to get a shared list of genes between human and macaque. Secondly, we aligned the monkey and human cells together to survey the consistency and divergence in single cell transcriptome. The union of top 3,000 variable genes in either species was used, and the canonical correlation analysis from *Seurat* was adopted to implement this integration. We used *MetageneBicorPlot* function to examine the dependence between canonical correlation number and correlation strength, and consequently chosen 10 as the most optimized canonical correlation number for downstream clustering and visualization analyses. We observed the correspondence of excitatory neurons, interneurons, astrocytes, oligodendrocytes, oligodendrocyte progenitor cells, and endothelia, except for microglia. Lastly, we employed the correlation analysis to detect the correspondence of cell subtype for excitatory neurons and interneurons by using expression matrix of these published cell types marker genes (*59*).

Integrative analysis of transcriptomic differences between human and macaque neurodevelopment

Matching developmental age between species: We developed a model named *TranscriptomeAge* to predict the equivalent human age for macaque brains based on transcriptome of neocortical samples. We divided the entire human neocortical samples into two overlapping subsets and fitted early and late models, respectively. In the early and late models, we log-transformed (fetal to adolescent) or square-root-transformed (birth to adult) the age (postconceptional days) of these samples, using different transformations due to the different time scales present during prenatal and postnatal development. For each subset, the transformed age was then regressed against the expression of all protein-coding genes using a penalized regression model (elastic net). The elastic net model was applied using the glmnet function in glmnet R package (*83*), with $\alpha = 0.5$ and λ selected by ten-fold cross validation using the *cv.glmnet* function. We next picked a boundary that determines which model to use for prediction. We refined the separation point to 12 years (human age) because this age minimized the prediction error in an independent human brain transcriptome dataset (*84*) (mean prediction error is 0.32 developmental period) and also served as the boundary separating childhood from adolescence in a previous study (*29*). The separation age for macaque was chosen to 2 years, by maximizing the number of consecutive brains with predicted human ages greater than 12 years old using the late model. The matched human age was calculated for all neocortical samples of each macaque, and the median value was used as the matched human age for this macaque brain.

To match chimpanzee age with human age (Fig. 1B), we first normalized the adult data set (*34*) to match it with our data set presented in this manuscript, using overlapping samples. For each gene, we fitted a linear model using overlapping human and macaque samples in the two data sets, and transformed the chimpanzee gene expression based on this model. The matched human age for chimpanzee brains were then predicted based on the late model as described above.

Identification of transcriptomic signatures: The human neocortical samples from Period 3 (10 postconceptional weeks) and older, when all 11 neocortical areas analyzed are defined, were used. We preselected 6160 genes based on variance explained by age ($R^2 > 0.5$) in a cubic model $y \sim 1 + d + d^2 + d^3$, where y is gene expression and d is age in log2 (pcd). The human expression matrix (**G**) is factorized into a factor signature matrix (**F**) and a factor weight matrix (**W**), i.e., $\mathbf{G} = \mathbf{FW}$ (Fig. S17), through non-negative matrix factorization (NMF), with the assumption that the gene expression in each sample is a weighted sum of gene expression signatures (factors). Therefore, a column of **F** represents the gene expression of a factor. A column of **W** represents the weights of factors in a sample, and a row of **W** represents the weight of a factor across all samples. NMF was performed using the *nmf* function in the NMF R package (*85*), with nrun=100. We tried ranks from 2 to 10 and selected rank=5, as increasing the rank does not produce more patterns of signature trajectories (rows of **W**).

To compare between species, we used the human signature matrix computed above, and calculated the weight on each factor by fitting a linear regression model

$\boldsymbol{y}_s \sim \sum_{i=1}^{5} w_{is}\boldsymbol{f}_i$, where $\mathbf{y}_\mathrm{s}$ is the gene expression vector of macaque sample s; $\boldsymbol{f}_i$ is the signature of factor i and w_{is} is weight of macaque sample s on factor i.

Each factor is associated with a neurodevelopmental event based on cell-type enrichment and developmental trajectories. The trajectory of each factor was compared with neurodevelopmental trajectories estimated based on gene markers, trajectories plotted based on external staining data, and changes of cellular composition (see cell-type deconvolution below). Marker genes of neurodevelopmental process and external immunohistochemistry data was obtained as described in (*29*). The first principal component of marker gene expression was calculated to represent the trajectories of the corresponding developmental process. The cell type enrichment of each factor was tested in four single-cell/nucleus RNA-seq data sets: the prenatal human neocortical single-cell and adult human DFC single-nucleus RNA-seq data from (*33*), and the fetal macaque DFC single-cell and adult macaque DFC single-nucleus RNA-seq data generated in this study. The Preferential Expression Measure (PEM) score (*86*) was calculated for each gene to quantify cell-type enrichment. Top 100 genes ranked by PEM were selected as cell type markers. Enrichment of the cell-type markers in each factor ($\boldsymbol{f}_i$) was tested by Kolmogorov–Smirnov test. Significance is defined by Benjamini-Hochberg adjusted $P < 0.01$ (*87*).

Cell-type deconvolution: We first created a cell-type gene expression signature matrix for NPC, ExN fetal, InN fetal, ExN adult, InN adult, astrocyte adult, OPC adult, oligodendrocyte adult, microglia adult, and endothelia adult, by combining a fetal human prefrontal cortex single cell RNA-seq data set (*35*) with the adult human single nuclei

RNA-seq data we generated. Top 24 genes ranked by PEM were selected per cell type to create the signature matrix, as its cell type mean expression matrix minimized the 2-norm condition number. The abundances of cell types in each sample were estimated by CIBERSORT (*88*). The nomenclatures of cell types used in this analysis were listed in the table S25. Analysis was conducted using online version of CIBERSORT with default parameters.

Analysis of inter-regional differences: As the distribution of sample age is difficult to be exactly matched between species, it is difficult to fairly compare the regional variation between species. To circumvent the problem, we fitted a Gaussian process model (*89*) for each gene across development and calculated the difference between brain regions at 100 selected time points. The Gaussian process model was fitted using the independence model in *TempShift* (see Section "Identification of genes with heterochronic expression"). In brief, for each gene, we assumed the age vector of brain region r is denoted by $\boldsymbol{x}_r$, and the corresponding gene expression vector is denoted by $\boldsymbol{y}_r$. The combined age vector across all 16 brain regions is $\boldsymbol{x} = (\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \dots, \boldsymbol{x}_{16}^T)^T$ and the combined expression vector is $\boldsymbol{y} = (\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \dots, \boldsymbol{y}_{16}^T)^T$. The age was measured in log2-postconceptual days and matched human age was used for macaque brains. In the independence model, we assumed $\boldsymbol{y}$ follows a multivariate Gaussian distribution, with the between-region covariance matrix set to a zero matrix $\mathbf{0}$ (eq. 1), where $K(\boldsymbol{x_i}, \boldsymbol{x_j})$ is a covariance function or kernel between samples from brain regions i and j, and $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \mathbf{0}$, if $i \neq$ j. In this study, we adopted Gaussian kernel $\mathrm{K}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \sigma_f^2 exp\ (-\frac{(x_i - x_j)^2}{2l^2})$, where l denotes the length scale and σ_f^2 denotes the amplitude, which determines the average distance of the

function away from its mean. The parameters were determined by maximum likelihood estimation, and optimized using *optim* function in R.

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} K(\boldsymbol{x_1}, \boldsymbol{x_1}) & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & K(\boldsymbol{x_{16}}, \boldsymbol{x_{16}}) \end{bmatrix}\right) + \sigma^2 \boldsymbol{I_n} \quad (1)$$

We evenly selected 100 time points ranging from 5.8 to 14 log2 (pcd), the range of sample age, and predicted the expression of all genes at these time points. For each gene at each time point, we calculated the pairwise distance between each pair of brain regions, as the absolute difference in gene expression. For each gene, the difference of a selected region from other brain regions was defined as the average of pairwise distance of this region to all other regions. The upper quartile of gene inter-regional differences was plotted in Fig. 2A and B, and fig. S21.

Analysis of inter-species divergence: We analyzed the inter-species divergence in each brain region separately, by fitting a Gaussian process model for each gene. The Gaussian process model was fitted using the independence model in *TempShift* (see Section "Identification of genes with heterochronic expression"). In the selected brain region, for each gene, assume the age vectors of human and macaque are denoted by $\boldsymbol{x}_h$ and $\boldsymbol{x}_m$, and the corresponding gene expression vectors are denoted by $\boldsymbol{y}_h$ and $\boldsymbol{y}_m$, respectively. The age vector combing both species is $\boldsymbol{x} = (\boldsymbol{x}_h^T, \boldsymbol{x}_m^T)^T$ and the combined expression vector is $\boldsymbol{y} = (\boldsymbol{y}_h^T, \boldsymbol{y}_m^T)^T$. The age was measured in log2-postconceptual days and matched human age was used for macaque brains. We assume $\boldsymbol{y}$ follows a multivariate Gaussian distribution, with the between-species covariance matrix set to a zero matrix $\boldsymbol{0}$ (eq. 2). $K(\boldsymbol{x_i}, \boldsymbol{x_j})$ is a covariance function or kernel between samples from species i and j.

$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \mathbf{0}$, when $i \neq$ j, and Gaussian kernel $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \sigma_f^2 exp\ (-\frac{(x_i - x_j)^2}{2l^2})$ is used, when $i =$ j. The parameters were determined by maximum likelihood estimation.

$$\boldsymbol{y} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(\boldsymbol{x_h}, \boldsymbol{x_h}) & \mathbf{0} \\ \mathbf{0} & K(\boldsymbol{x_m}, \boldsymbol{x_m}) \end{bmatrix}\right) + \sigma^2 \boldsymbol{I_n} \quad (2)$$

As in the inter-regional model, we evenly selected 100 time points ranging from 5.8 to 14 log2 (PCD), the range of sample age, and predicted the expression of all genes at these time points. At each time point, we calculated the inter-species divergence as the absolute difference between species. The upper quartile of gene inter-species divergence was plotted in Fig. 3A.

To explore what genes drive regional differences in inter-species divergence, we performed PCA using inter-species divergence of neocortical areas by phase. Top 100 genes correlated with either direction of PC1 were selected and tested for enrichment in GO terms from *gageData* R package, using Fisher's exact test (FDR < 0.05).

Identification of inter-species differentially expressed genes: As bias in sample age between species may result in false discovery of temporal dynamic genes as inter-species differentially expressed genes, we used the inter-species Gaussian process model fitted in the above section to estimate the mean (**μ**) and covariance matrix (**Σ**) of gene expression at 100 time points (**x**) evenly selected from 5.8 to 14 [log2(pcd)]. For gene g, at time points $\boldsymbol{x}^*$, the predicted expression of species s is $\boldsymbol{y}_{gs}^* | \boldsymbol{x}^*, \boldsymbol{x}_s, \boldsymbol{y}_{gs} \sim \mathcal{N}(\boldsymbol{\mu}_{gs} = K(\boldsymbol{x}^*, \boldsymbol{x}_s)[K(\boldsymbol{x}_s, \boldsymbol{x}_s) + \sigma_n^2 \boldsymbol{I}]^{-1} \boldsymbol{y_{gs}}$,

$K(\boldsymbol{x}^*, \boldsymbol{x}^*) - K(\boldsymbol{x}^*, \boldsymbol{x}_s)[K(\boldsymbol{x}_s, \boldsymbol{x}_s) + \sigma^2 \boldsymbol{I}]^{-1} K(\boldsymbol{x}_s, \boldsymbol{x}^*))$, where $\boldsymbol{y_{gs}}$ is the input expression vector of gene g at time points $\boldsymbol{x}_s$ in species s, and K(.,.) is the kernel function. As every

finite collection of the random variables in a Gaussian process has a multivariate normal distribution, to identify inter-species differentially expressed genes in each phase (p) of the hourglass, for gene g, we tested the null hypothesis $\boldsymbol{\mu}_{ghp} = \boldsymbol{\mu}_{gmp}$ against the alternative hypothesis $\boldsymbol{\mu}_{ghp} \neq \boldsymbol{\mu}_{gmp}$ by chi square test, based on the conditional distribution of $\boldsymbol{y}^*_{ghp}|\boldsymbol{x}^*, \boldsymbol{x}_h, \boldsymbol{y}_{gh}$ and $\boldsymbol{y}^*_{gmp}|\boldsymbol{x}^*, \boldsymbol{x}_m, \boldsymbol{y}_{gm}$. $\boldsymbol{\mu}_{ghp}$ and $\boldsymbol{\mu}_{gmp}$ are the mean expression vectors of gene g at time points within selected phase p in human (h) and macaque (m), respectively. According to the hourglass-like pattern, we divided the 100 time points into 3 phases, with boundaries at birth (266 pcd) and adulthood (20 years old). Given $\boldsymbol{y}^*_{gs}|\boldsymbol{x}^*, \boldsymbol{x}_s, \boldsymbol{y}_{gs} \sim \mathcal{N}(\boldsymbol{\mu}^*_{\boldsymbol{gs}}, \boldsymbol{\Sigma}_{\mathbf{gs}})$, where $\boldsymbol{\mu}^*_{gs} = \left(\boldsymbol{\mu}_{gs1}, \boldsymbol{\mu}_{\mathrm{gs2}}, \boldsymbol{\mu}_{\mathrm{gs3}}\right)^{\mathrm{T}}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{gs11}} & \boldsymbol{\Sigma}_{gs12} & \boldsymbol{\Sigma}_{gs13} \\ \boldsymbol{\Sigma}_{gs21} & \boldsymbol{\Sigma}_{gs22} & \boldsymbol{\Sigma}_{gs23} \\ \boldsymbol{\Sigma}_{gs31} & \boldsymbol{\Sigma}_{gs32} & \boldsymbol{\Sigma}_{gs33} \end{bmatrix}$, at phase p, $\boldsymbol{y}^*_{gsp}|\boldsymbol{x}^*, \boldsymbol{x}_s, \boldsymbol{y}_{gs} \sim \mathcal{N}(\boldsymbol{\mu}^*_{\boldsymbol{gsp}}, \boldsymbol{\Sigma}_{\mathbf{gspp}})$. To test if $\boldsymbol{\mu}_{gph} = \boldsymbol{\mu}_{gpm}$, we tested if $\boldsymbol{y}^*_{ghp} - \boldsymbol{y}^*_{gmp} \sim \mathcal{N}(0, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\mathrm{ghp}} + \boldsymbol{\Sigma}_{\mathrm{gmp}})$. Under null hypothesis, $\left(\boldsymbol{y}^*_{ghp} - \boldsymbol{y}^*_{gmp}\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y}^*_{ghp} - \boldsymbol{y}^*_{gmp}\right)$ follows chi-square distribution with d degrees of freedom, where d is the dimension of $\boldsymbol{x}^*$. A gene is considered as differentially expressed, if FDR (*53*) < 0.01 and fold-change > 1.5.

Analysis of evolutionary conservation in sequence of inter-species differentially expressed genes: Human and macaque gene annotations were retrieved from Ensembl (*88*) using BioMart (*90*). For coding evolutionary analyses of gene sequences we retrieved information for genes annotated as protein-coding by Ensembl in both species. In order to avoid paralogy-related ambiguities we focused on one to one orthologs with high orthology confidence according to Ensembl. For these human-macaque orthologs

we retrieved their gene structure, genomic coordinates and the corresponding nonsynonymous (dN) and synonymous (dS) substitution rates. In order to ensure a subset of fully comparable gene annotations among Ensembl- and XSAnno-based annotations, we focused on the 10,609 genes detected by both pipelines in the same genomic regions (≥50% of overlapping exonic sequences in both annotations). In addition, we removed one gene for which we could not discard saturation (dS higher than the mean plus 2 standard deviations of the whole inter-species dS values) as well as a subset of 48 genes whose classification as protein coding was dubious (*91*). By implementing this pipeline we retrieved a dataset of 10,011 high quality one to one human-macaque ortholog protein coding genes (with 4,429 genes belonging to the set of genes changing their expression from macaque to human).

To compare the coding sequence evolution of the different gene sets of interest, we performed a Wilcoxon-Mann-Whitney test between the genes in the corresponding set and the remaining genes in our dataset.

Weighted gene co-expression correlation network analysis (WGCNA): Signed co-expression networks in neocortex were built using the R package *WGCNA* (*92*) within species, roughly according to the three phases in the regional developmental hourglass: prenatal development, postnatal development and adult. The age of samples included in each network is listed in table S26.

For prenatal development, only human samples from Period 3-7 were used, as neocortical areas are not well defined in Period 1 and 2. In each phase, genes with RPKM > 1 in at least 3 human or macaque samples were selected and subjected to network

analysis. For all genes included, a pairwise correlation matrix was computed, and an adjacency matrix was calculated by raising the correlation matrix to a power. The power was set to 20 for all networks according to a scale-free topology criterion (*92*). We then implemented the *blockwiseModules* function in the WGCNA package, with the following setting: power=20, networkType="signed", TOMType="signed", deepSplit=2, detectCutHeight=0.995, minModuleSize=10, minCoreKME=0.7, minCoreKMESize=3, minKMEtoStay=0.7, minCoreKMESize=3, and mergeCutHeight=0.15. Each module was summarized by an eigengene, which is the first principal component of the scaled module expression. Thus, the module eigengene explained the maximum amount of variation of the module expression levels.

The regional expression patterns of modules in each species were presented by averaging module eigengenes by region in each phase. The lists of genes in human modules were also used to calculate regional expression pattern of the modules in macaques and vice versa. A module is considered as species-distinct if the correlation of regional expression pattern between species is less than 0.5.

To explore whether the regional expression of a module changes with time for each phase and for each species, module eigengenes were calculated using the gene expression predicted from inter-species independence model fitted in the above section (see Section "Analysis of evolutionary conservation in sequence of inter-species differentially expressed genes").

Identification of genes with heterochronic expression: To identify genes with temporal shift between neocortical areas or between species, we implemented *TempShift*, which

fits an independence model (eq. 3), a no-shift model (eq. 4) and a shift model (eq. 5) for each gene. Assume we have samples from m groups, the combined time vector across all the samples is $\boldsymbol{x} = \left(\boldsymbol{x_1^T}, \boldsymbol{x_2^T}, \ldots, \boldsymbol{x_m^T}\right)^T$ and the combined expression vector is $\boldsymbol{y} = \left(\boldsymbol{y_1^T}, \boldsymbol{y_2^T}, \ldots, \boldsymbol{y_m^T}\right)^T$. In the three models, $\boldsymbol{y}$ follows a multivariate Gaussian distribution with different covariance matrices. $\boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j})$ is a covariance function or kernel between groups i and j, with n_i rows and n_j columns. In this study, we adopted Gaussian kernel $\mathrm{K}\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) = \sigma_f^2 \exp\left(-\frac{\left(\boldsymbol{x_i}-\boldsymbol{x_j}\right)^2}{2l^2}\right)$, where l denotes the length scale and σ_f^2 denotes the amplitude, which determines the average distance of the function away from its mean. We evaluated whether a gene shows temporal shifts across groups in two steps: 1) we compared the shift model with the independence model to identify genes following the same trajectory across groups based on log-likelihood ratio $\mathrm{LLR_{shape}}$ (eq. 6); we then compared the shift model with the no-shift model to identify genes with temporal shift by calculating log-likelihood ratio $\mathrm{LLR_{shift}}$ (eq. 7).

Independence model:
$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{K}(\boldsymbol{x_1}, \boldsymbol{x_1}) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{K}(\boldsymbol{x_m}, \boldsymbol{x_m}) \end{bmatrix}\right) + \sigma^2 \boldsymbol{I_n} \tag{3}$$

No-shift model:
$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{K}(\boldsymbol{x_1}, \boldsymbol{x_1}) & \cdots & \boldsymbol{K}(\boldsymbol{x_1}, \boldsymbol{x_m}) \\ \vdots & \ddots & \vdots \\ \boldsymbol{K}(\boldsymbol{x_m}, \boldsymbol{x_1}) & \cdots & \boldsymbol{K}(\boldsymbol{x_m}, \boldsymbol{x_m}) \end{bmatrix}\right) + \sigma^2 \boldsymbol{I_n} \tag{4}$$

Shift model:
$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{K}(\boldsymbol{x_1}, \boldsymbol{x_1}) & \ldots & \boldsymbol{K}\left(\boldsymbol{x_1}, \boldsymbol{x_j} - \Delta t_j \vec{\mathbf{1}}_{n_j}\right) & \ldots & \boldsymbol{K}\left(\boldsymbol{x_1}, \boldsymbol{x_m} - \Delta t_m \vec{\mathbf{1}}_{n_m}\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{K}\left(\boldsymbol{x_i} - \Delta t_i \vec{\mathbf{1}}_{n_i}, \boldsymbol{x_1}\right) & \ldots & \boldsymbol{K}\left(\boldsymbol{x_i} - \Delta t_i \vec{\mathbf{1}}_{n_i}, \boldsymbol{x_j} - \Delta t_j \vec{\mathbf{1}}_{n_j}\right) & \ldots & \boldsymbol{K}\left(\boldsymbol{x_i} - \Delta t_i \vec{\mathbf{1}}_{n_i}, \boldsymbol{x_m} - \Delta t_m \vec{\mathbf{1}}_{n_m}\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{K}\left(\boldsymbol{x_m} - \Delta t_m \vec{\mathbf{1}}_{n_m}, \boldsymbol{x_1}\right) & \ldots & \boldsymbol{K}\left(\boldsymbol{x_m} - \Delta t_m \vec{\mathbf{1}}_{n_m}, \boldsymbol{x_j} - \Delta t_j \vec{\mathbf{1}}_{n_j}\right) & \ldots & \boldsymbol{K}\left(\boldsymbol{x_m} - \Delta t_m \vec{\mathbf{1}}_{n_m}, \boldsymbol{x_m} - \Delta t_m \vec{\mathbf{1}}_{n_m}\right) \end{bmatrix}\right) + \sigma^2 \boldsymbol{I_n} \tag{5}$$

$$LLR_{shape} = \log\left(\frac{max\{L(\boldsymbol{\Delta t}, \boldsymbol{\theta}_{shift} | \boldsymbol{x}, \boldsymbol{y})\}}{max\{L(\boldsymbol{\theta}_{independence} | \boldsymbol{x}, \boldsymbol{y})\}}\right) \tag{6}$$

$$LLR_{shift} = \log\left(\frac{max\{L(\boldsymbol{\Delta t}, \boldsymbol{\theta_{shift}} | \boldsymbol{x}, \boldsymbol{y})\}}{L(\boldsymbol{\Delta t}=\mathbf{0}, \boldsymbol{\theta_{shift}} | \boldsymbol{x}, \boldsymbol{y})}\right) \tag{7}$$

To identify temporal shift genes across neocortical areas, we applied the above approach to samples from each species separately. As the input to the model, $\boldsymbol{x}_i$ is the age vector in log2-postconceptional days of samples from neocortical area i and $\boldsymbol{y}_i$ is the corresponding expression vector. To identify temporal shift genes between species, we fitted the models using samples from each brain region of both species. In this case, $\boldsymbol{x}_i$ is the age vector in log2-postconceptional days of samples from species i and $\boldsymbol{y}_i$ is the corresponding expression vector. For macaque samples, matched human age is used in both analyses.

A gene g is considered as heterochronic, if $LLR^{g}_{random} > median(LLR_{non-random}) - MAD(LLR_{non-random})$, $LLR^{g}_{shape} > median(LLR_{shape}) - MAD(LLR_{shape})$, and $LLR^{g}_{shift} > median(LLR_{shift}) + 2 * MAD(LLR_{shift})$. $\mathrm{LLR}_{\mathrm{non-random}}$ is the log-likelihood ratio between independence model and a random model assuming gene expression of each sample follows i.i.d. Gaussian distribution.

Genes that fulfilled the first two criteria in all neocortical areas and were heterochronic in any of the areas, were used for clustering. The inter-areal and inter-species heterochronic genes were clustered using *dynamicTreeCut* (*93*) with default parameters, by neocortical areal $\boldsymbol{\Delta t}$ in both species, and inter-species $\boldsymbol{\Delta t}$ in all regions, respectively. All selected inter-species heterochronic genes were used to conduct enrichment analysis in canonical pathways and upstream transcriptional regulators by Ingenuity Pathway Analysis (IPA, QIAGEN). Significance was defined as $P < 0.05$. The regulatory network was visualized by *igraph (94)* R package.

Inter-species difference in regional expression of disease-associated genes: The list of genes associated with psychiatric disorders and cognitive traits were obtained from *(33)*. The regional expression of disease-associated genes was presented by average expression in each region and phase. Genes were then centered across neocortical areas by areal mean and the distance between species was calculated for each gene. Genes with centered distance > 3 were considered as heterotopic genes between species. The heterotopic schizophrenia genes were first clustered by the phase when they show inter-species difference and then by hierarchical clustering.

Identification of cell type-specific inter-species differentially expressed genes: To identify genes differentially expressed between species in each cell type, we compared age-matched human and macaque single-cell RNA/nucleus-seq data from fetal and adult neocortex. To detect species-specific expression in fetal, we compared cells from prefrontal cortex of GW23 and GW26 human neocortex (*35*) with single cells from DFC of age-matched E110 monkey brain. Given the limited cell number in Zhong et. al data (*35*), we considered cell types instead of subtypes in the comparisons. To test inter-species differential expression in adult cells, we used single-nucleus RNA-seq from adult human and macaque DFC. To match subtypes between species, we first reclassified the macaque excitatory neurons and interneurons using the random forest built on human data. The correlation between human and macaque matched subtypes were plotted in fig. S47 and the overlap between macaque subtypes and macaque subtypes matched to human is listed in table S24. A gene is considered as robustly expressed in only one species for a certain cell type, if it is expressed (UMI count > 0) in more than 30 cells in one species,

but fewer than 10 in the other. To reduce false positives and to select genes with cell-type specific inter-species differential expression, the genes were filtered using the following criteria: first, the genes should change in the same direction between species in all cell types showing inter-species differential expression; second, the genes should change in the same direction in neocortex tissue and in single cells, and finally, the genes are enriched in the cell type with inter-species difference (PEM in cell types with species difference are greater than 0.3 and greater than PEM in other cell types).

Cell type-enrichment analysis: Cell type enrichment was tested in four single-cell/nucleus RNA-seq data sets: the human prenatal neocortical single-cell and adult single-nucleus RNA-seq data from (*33*), and the fetal macaque DFC single-cell and adult macaque DFC single-nucleus RNA-seq data generated in this study. Gene lists subjected to cell type-enrichment analysis includes: co-expressed gene modules, inter-species differentially expressed genes and heterochronic genes. The enrichment analysis is performed as follows: 1) the Preferential Expression Measure (PEM) score (*86*) was calculated for each gene to quantify cell-type enrichment.; 2) the PEM distribution of selected genes was compared with that of background genes using Kolmogorov–Smirnov test. A gene list is considered to be significantly enriched in a cell type if Benjamini–Hochberg adjusted $P < 0.01$.

Functional enrichment analysis: Enrichment in KEGG pathways and GO terms was tested based on *gageData* R package with Fisher's exact test. The false discovery rate (FDR) was controlled by applying Benjamini–Hochberg procedure (*87*). Enrichment in

canonical pathways and upstream transcriptional regulators was analyzed by Ingenuity Pathway Analysis (IPA, QIAGEN).

Transcription factor enrichment: We retrieved regulatory elements which had been previously defined for the human and macaque genomes in different regions of the adult brain (*58*). These regulatory elements were mainly composed of enhancer and promoter regions differentially enriched in the two species and further classified as either higher or lower enriched in human relative to macaque. We examined the overrepresentation of specific transcription factor binding sites (TFBS) for each brain region and enrichment direction (high or low) using findMotifsGenome.pl (HOMER v4.10) (*95*) and retrieved those transcription factors specific to the enriched TFBS which were inter-species differentially expressed (DEX) genes themselves. Then we retrieved all the inter-species differentially expressed genes showing binding motifs for the enriched DEX-TF using findMotifs.pl (HOMER v4.10) (*95*).

For each brain region, we also tested if the number of inter-species differentially expressed genes was enriched in TF-DEX binding sites. For that we used findMotifs.pl (HOMER v4.10) (*95*) which analyses the promoters of DEX genes relative to other species-specific promoters. We thus performed this enrichment for all anatomical brain regions and developmental stages considered in this study.

Differential exon usage: An inter-species differential exon usage analysis in 16 brain regions across developmental stages (prenatal, early postnatal and adulthood) was conducted using DEXSeq R package (*96, 97*). For this analysis we focused on human-

macaque orthologous exons retrieved by XSAnno framework (*37*) and samples from the same species from a given brain region and developmental stage were treated as biological replicates. Read counting for orthologous exonic elements was carried out using HTSeq library (*96*) considering only uniquely mapping reads. To account for differences in sequencing depth, size factors were estimated within DEXSeq and used to normalize the expression estimates in different RNA-seq libraries. The dispersion for each counting bin was estimated by DEXSeq using the method described in DESeq2 (*98*). Then, relative exon usage fold changes between species were calculated for each exonic element. Due to the high number of samples in some comparatives, differential exon usage analyses were parallelized through BiocParallel R package in order to reduce computational times. In order to minimize the effect of noise we focused on highly expressed exonic elements (>10 mean normalized counts across samples) showing very clear differential usage (adjusted p-value <0.01 and absolute log2 fold change >2). Moreover, we kept only those differentially used exonic elements pointing to a given transcript with a consistent pattern through all its exonic elements. To do this, exonic elements were considered as differentially expressed (normalized count mean across samples > 10, adjusted p-value < 0.01 and absolute fold change > 2) or not differentially expressed (normalized count across samples > 10 and adjusted p-value > 0.01), leaving the rest as unclear. Thus, we focused on differentially used exonic elements compatible with the differential expression of whole transcripts (none of their exonic elements are not differentially expressed).

Validation of gene expression and histology

Droplet digital PCR for gene expression quantitation: We employed droplet digital PCR (ddPCR) to reliably quantitate gene expression. An aliquot of the total RNA that was previously extracted from ventrolateral prefrontal cortex (VFC) was used for secondary validation through ddPCR analysis. The samples used for validation are described in table S27.

One µg of total RNA was used for cDNA synthesis using SuperScript III First-strand synthesis Supermix (Invitrogen) and subsequently diluted with nuclease-free water. Custom gene-specific primers and probe for each gene of interest were designed using NCBI/Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and PrimerQuest tool (IDT). ddPCR was carried out using the Bio-Rad QX100 system. After each PCR reaction mixture consisting of ddPCR master mix and custom primers/probe set was partitioned into 15,000–20,000 droplets, parallel PCR amplification was carried out. Endpoint PCR signals were quantified and Poisson statistics was applied to yield target copy number quantification of the sample. Two-color PCR reaction was utilized for the normalization of gene expression by the housekeeping gene *TBP*. Sequences of primers used in validation are as follows:

PADI2: Forward: CTTCGGCCAACGCTATATCC

Probe (5'FAM/ZEN/3'IBFQ): ACGGGTGGCTCCGCGGAGCT

Reverse: AGATGGGAGTCAGGGGAATG

UAP1L1: Forward: CTGCTCACCCAGCACTACCG

Probe (5'FAM/ZEN/3'IBFQ): GCCGGGGCCCGCTTCCTGGA

Reverse: CGGCCTTGCAGGTACACTTCT

LINC00299: Forward: TGGCGGTTGTATGATCCTGG

Probe (5'FAM/ZEN/3'IBFQ): TGGGGCCTTTTCCTGCCGCTCCA

Reverse: GCAGTTGGGGTTAAGCTCCT

DGCR8: Forward: GCGAAGAATAAAGCTGCCCG

Probe (5'FAM/ZEN/3'IBFQ): CCACATCAGCATCGAGGACTCGCGG

Reverse: TTGCTGGTCAGCTCGTAGAC

PCDHGA11: Forward: TAGGCAAATGTGACCCGACA

Probe (5'FAM/ZEN/3'IBFQ): CCCGCCCAACACGGACTGGCG

Reverse: GGTGTCATCGCCATTTTGGG

TBP: Forward: CCACTCACAGACTCTCACAAC

Probe (5'HEX/ZEN/3'IBFQ): CCATCACTCCTGCCACGCCA

Reverse: GTACAATCCCAGAACTCTCCG

For each region, the gene expression was compared between species using *t*-test. The level of significance was set at 0.05

Immunohistochemistry: For validation studies, adult human and macaque brain tissue samples were fixed in 4% paraformaldehyde for 2 days at 4°C. Tissue sections were mounted and dried overnight. Antigen retrieval was performed using R-Buffer A pH 6.0 (Electron Microscopy Sciences #62706-10) and Electron Microscopy Sciences Retriever according to manufacturer's instructions. Sections were incubated in 1% hydrogen peroxide/PBS to quench endogenous peroxidase activity. Sections were washed in PBS (3 x 15 min) and incubated in blocking solution containing 5% (v/v) normal donkey serum (Jackson ImmunoResearch Laboratories), 1% (w/v) bovine serum albumin, and 0.4% (v/v) Triton X-100 in PBS for 1 h at room temperature. Primary antibody (PTH2R;

Sigma #SAB2900515) was diluted 1:100 in blocking solution and incubated with tissues sections overnight at 4°C. Sections were washed with PBS (3 x 15 min) prior to being incubated with the appropriate biotinylated or fluorescent secondary antibodies (Jackson ImmunoResearch Labs) for 1.5 h at room temperature. All secondary antibodies were raised in donkey and diluted at 1:250 in blocking solution. Sections were subsequently washed in PBS and incubated with avidin-biotin-peroxidase complex (Vectastain ABC Elite kit; Vector Laboratories) for 1 h at room temperature. Finally, sections were washed in PBS (3 x 15 min) and signals were developed using a DAB peroxidase substrate kit according to the manufacturer's protocol (Vector Laboratories). Following washes in PBS, sections were dried, dehydrated, and cover-slipped with Permount (Fisher Scientific). For immunofluorescence, sections were washed in PBS with 0.3% triton and treated with the Autofluorescence Eliminator Reagent (Millipore #2160) according to manufacture instructions, and cover-slipped with Vectashield (Vector laboratories, #H-1000). Sections were digitized using Aperio CS2 scanner or Zeiss LSM 510 Meta Confocal Microscope and images were assembled in Zeiss Zen, ImageJ, Adobe Photoshop, and Adobe Illustrator.

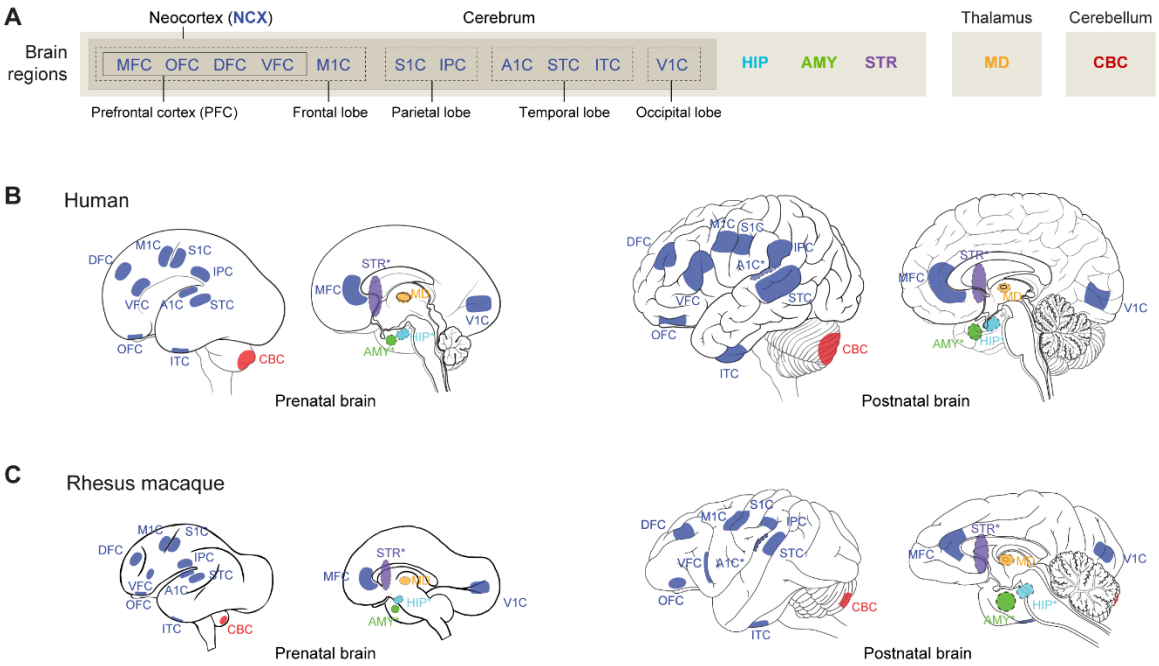


Fig. S1. Brain regions and neocortical areas of prenatal and postnatal human and rhesus macaque used for transcriptome profiling. (A) Schematic represents the sampled regions from each brain structure, including brain lobes and non-cortical structures. **(B-C)** Schematic shows anatomical positions of dissections of prenatal and postnatal brain samples in (**B**) human and (**C**) macaque. MFC – medial prefrontal cortex; OFC – orbital prefrontal cortex; DFC – dorsolateral prefrontal cortex; VFC – ventrolateral prefrontal cortex; M1C – primary motor cortex; S1C – primary somatosensory cortex; IPC – inferior posterior parietal cortex; A1C – primary auditory cortex; STC – superior temporal cortex; ITC – inferior temporal cortex; V1C – primary visual cortex; HIP – hippocampus; AMY – amygdala; STR – striatum; MD – mediodorsal nucleus of the thalamus; CBC – cerebellar cortex. Asterisks denote structures that are internal and not visible from the view shown.

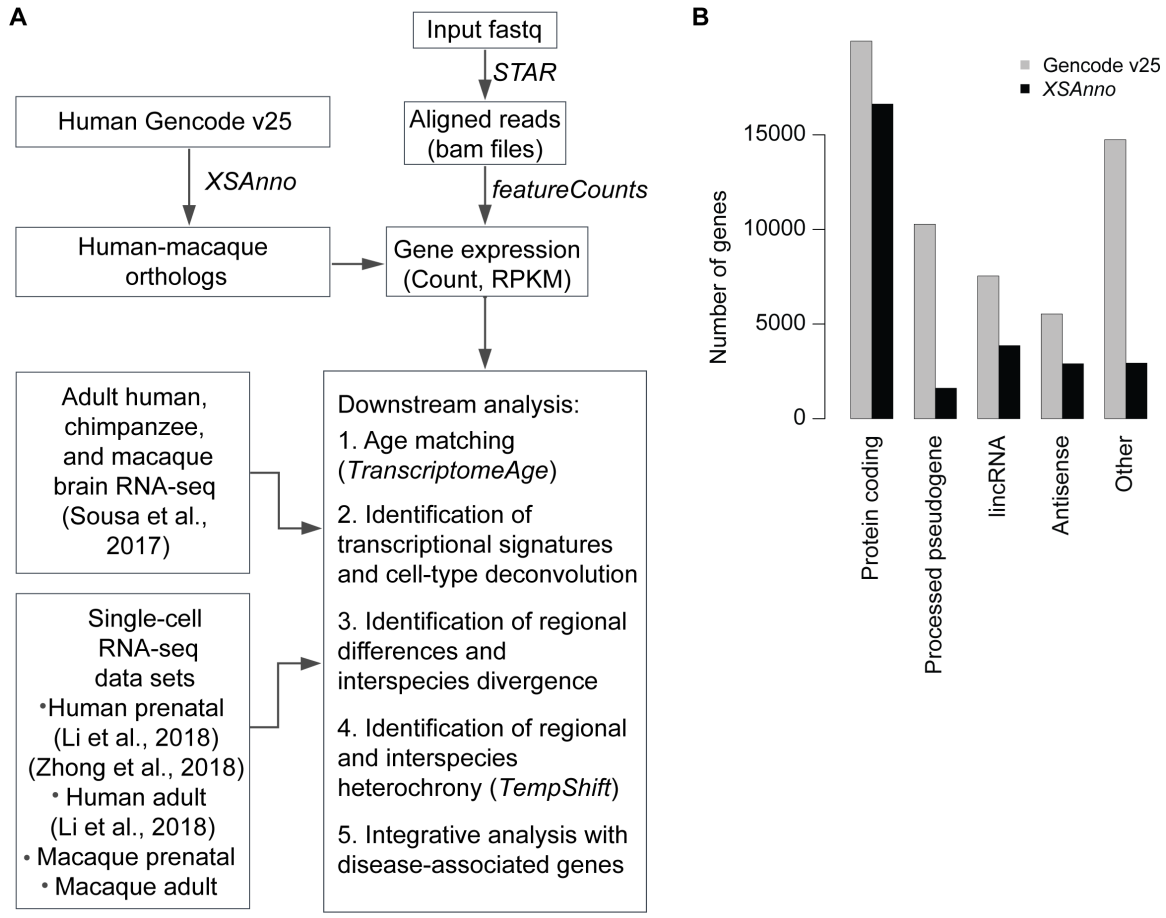


Fig. S2. Bulk tissue RNA-Seq analysis pipeline. (A) Pipeline describing all analytical steps and integration with other published data sets. **(B)** Number of genes and human-macaque orthologs generated by *XSAnno* (*37*). All analyses were conducted after filtering gene orthologs with *XSAnno*. Genes were divided into 5 main categories: protein coding, processed pseudogene, lincRNA, or antisense, with all other subcategories included in other.

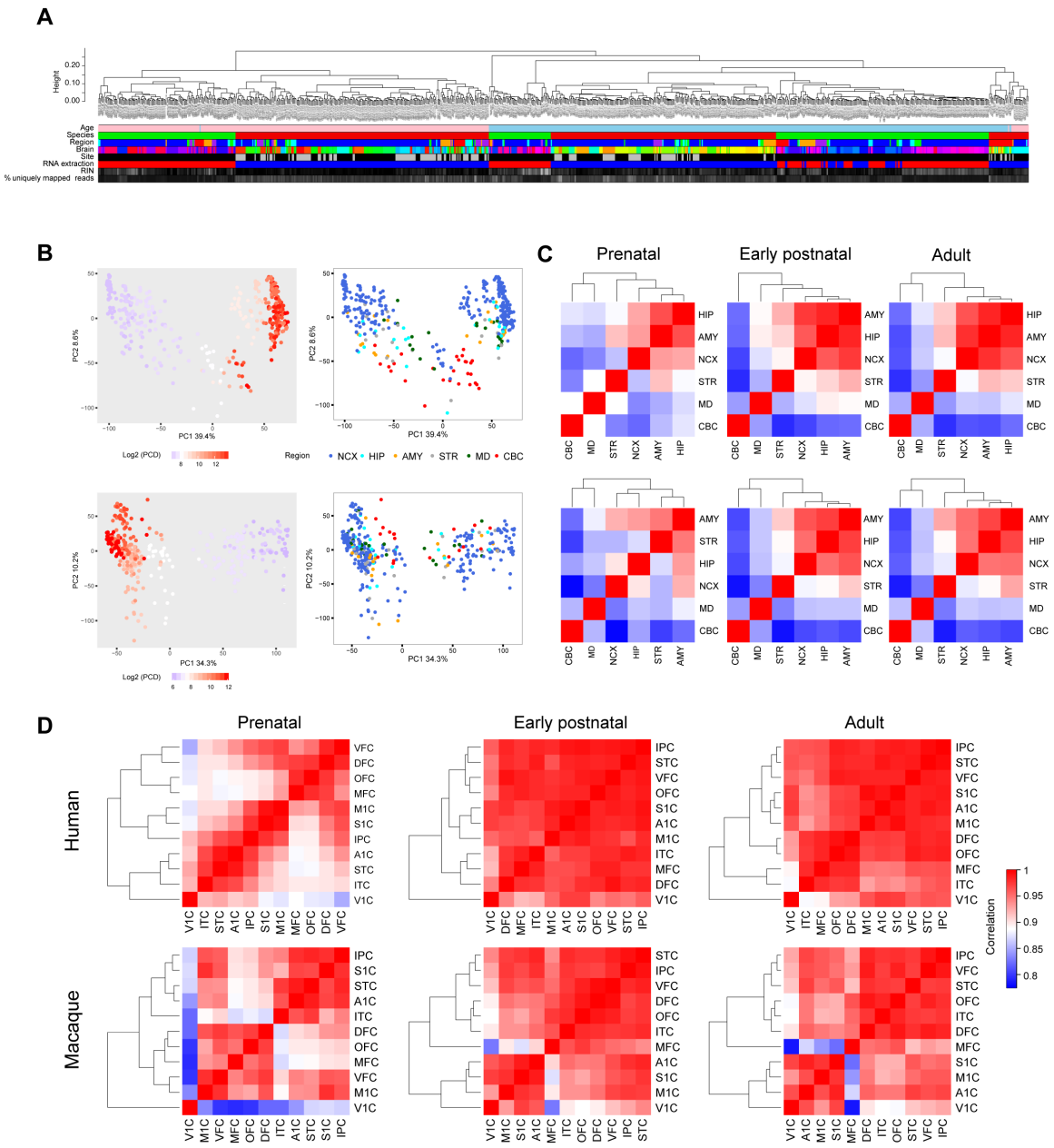


Fig. S3. Transcriptomic architecture of the developing and adult human and macaque brains. (A) Hierarchical clustering of all analyzed tissue-level samples. Unsupervised hierarchical clustering of all analyzed samples reveals that samples first cluster based on whether they correspond to prenatal or postnatal periods, as observed in the Principal Component Analysis, except samples from the 37 PCW human brain, which clustered with postnatal samples. Samples from prenatal or postnatal neurodevelopmental periods cluster mainly by species, except those from cerebellum. The color bar in the bottom represents variables including age: prenatal

(pink) and postnatal (blue); species: human (red) and macaque (green); region: NCX (blue), HIP (cyan), AMY (green), STR (purple), MD (orange) and CBC (red); sequencing site Yale (black) and USC (gray); RNA extraction kit: miRvana (red) and QIAGEN (blue); RNA integrity number (RIN): black (10.0) to white (0.00) gradient; percentage of uniquely mapped reads: black (100%) to white (0). **(B)** Principal component analysis of human (top) and macaque (bottom) samples. The samples were colored by postconceptional days (PCD; left) and regions (right), respectively. In both species, the white color represents birth, blue labels prenatal samples, and red labels postnatal samples. Samples were clustered based on age in PC1, but were not obviously clustered by brain region. **(C)** Heatmap and hierarchical clustering of prenatal, early postnatal, and adult human (top) and macaque (bottom) brain regions. Correlation of top 1,000 genes with highest regional differences was plotted. In both species, the correlation between NCX and HIP increased with development, while MD and CBC became more distant from other regions. **(D)** Heatmap and hierarchical clustering of prenatal, early postnatal and adult human (top) and macaque (bottom) neocortical areas. Correlation of top 1,000 genes with highest inter-areal differences was plotted. In the prenatal neocortex, samples clustered by topographical proximity, while in the early postnatal and adult neocortices, primary areas clustered together, except V1C in human. Furthermore, MFC shows larger distance from other regions in the early postnatal and adult macaque neocortices than in human.

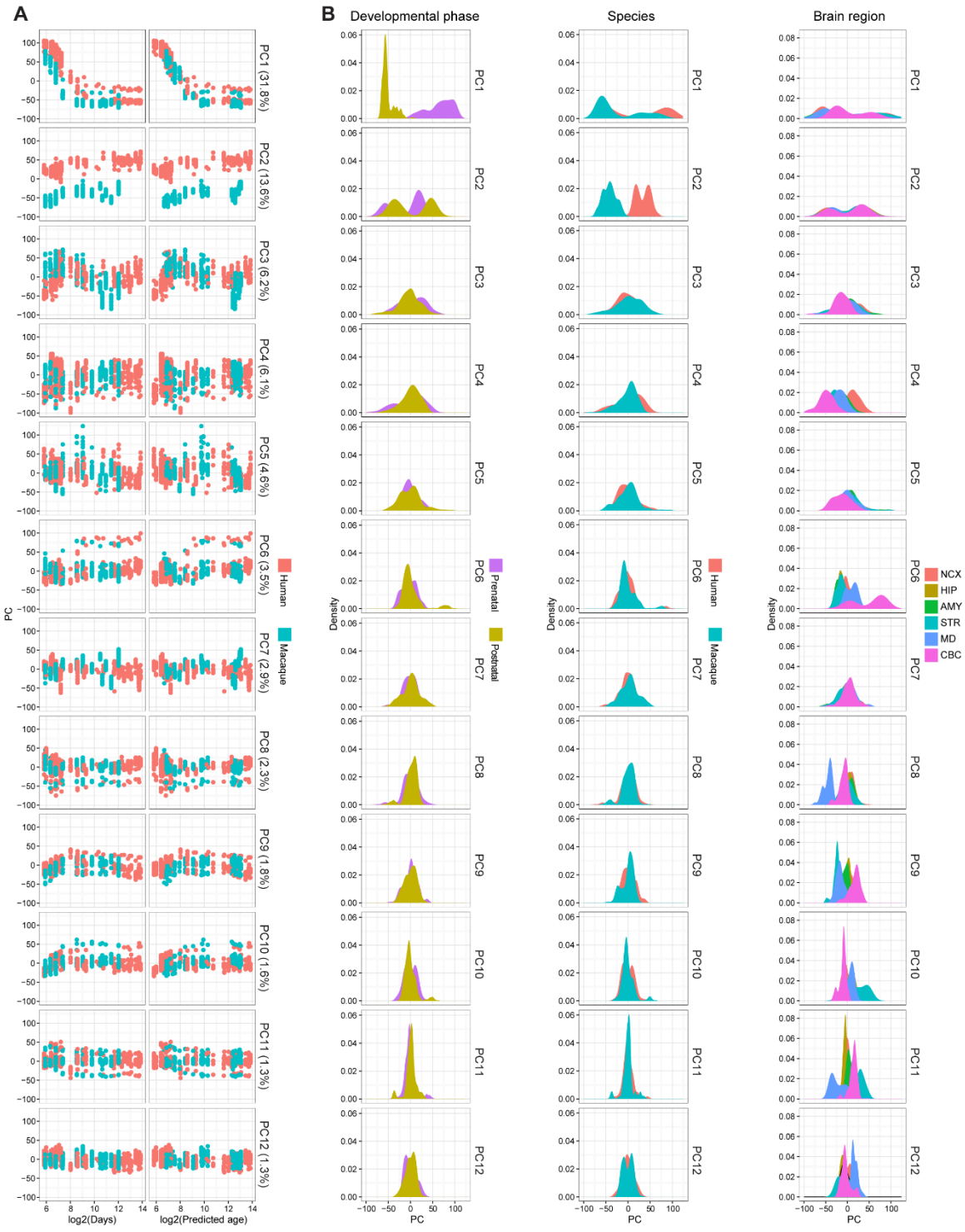


Fig. S4. Principal component analysis. (A) Graphical representation of the 12 first Principal Components (PC; variance explained > 1%) based on gene expression of brain samples corresponding to the real (left panel) or predicted (right panel) ages of human and macaque

neurodevelopment. Each point represents one sample. The percentage of variation that is explained by each PC is shown. **(B)** The first PC clearly separates prenatal and postnatal development in both human and macaque (left panel); the second PC separates human from macaque samples (middle panel); PC3 to PC12 separate different brain regions in both species (right panel).

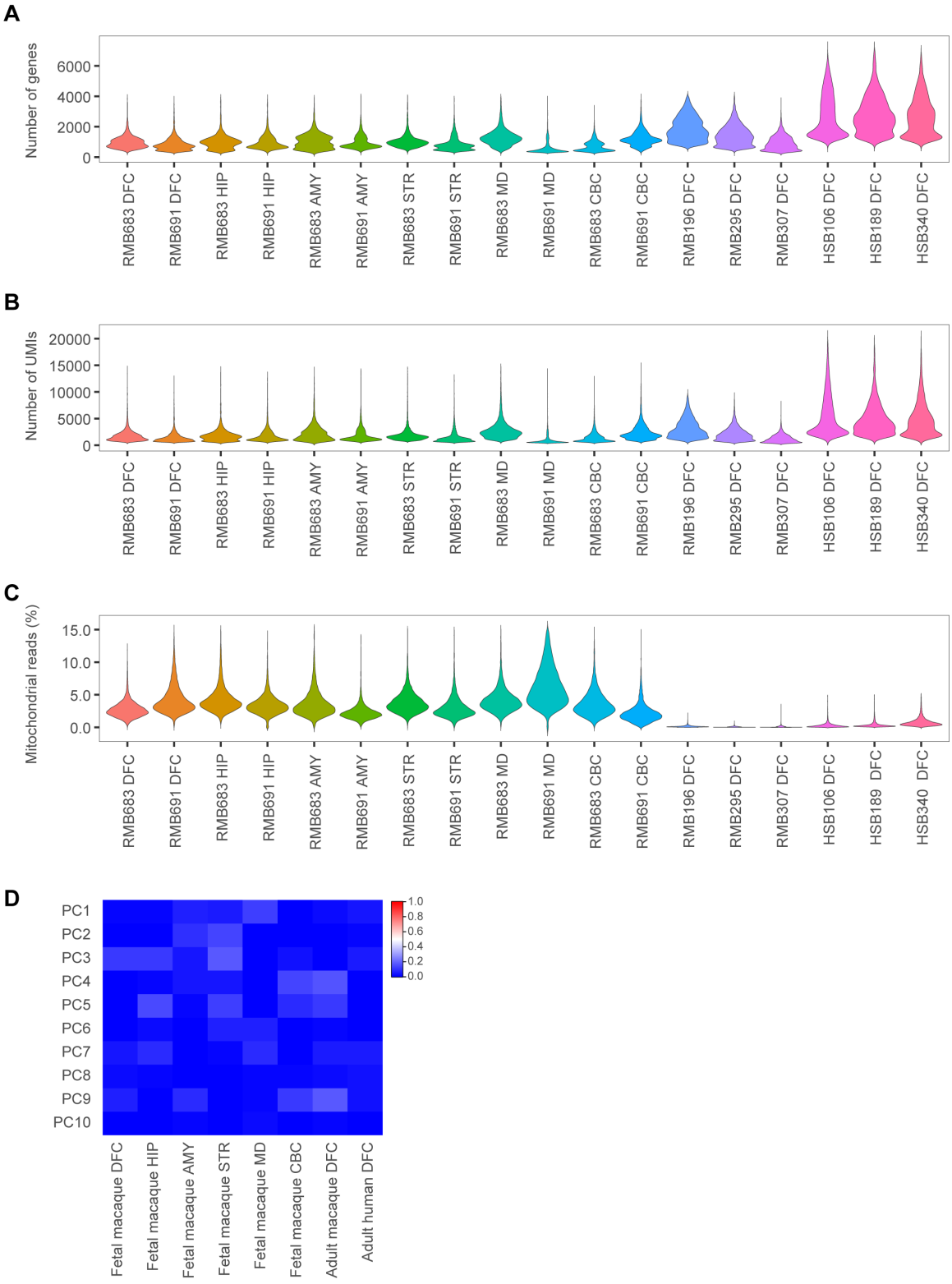


Fig. S5. Single-cell and single-nucleus RNA-Seq quality controls. (A) Violin plots showing the distribution of the number of genes detected in each human and macaque sample. **(B)** Violin plots

showing the distribution of the number of unique molecular identifiers (UMIs) detected in each human and macaque sample. **(C)** Violin plots showing the distribution of the percentage of mitochondrial reads in human and macaque samples. **(D)** Heatmap showing the Pearson correlation between the top 10 principal components (PCs) and the sequencing or library preparation batches. The color bar scales the correlation coefficient value.

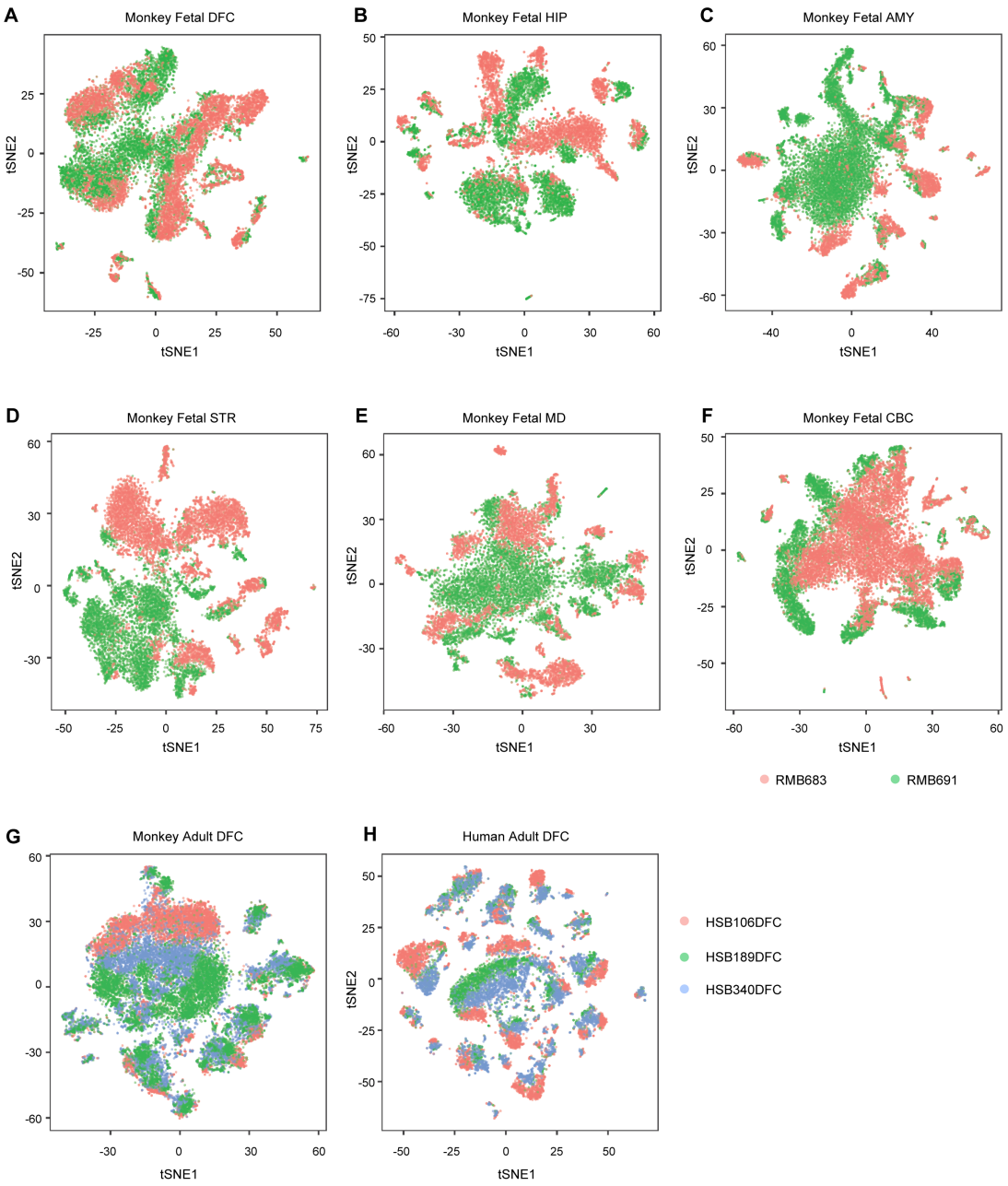


Fig. S6. Specimen representation in macaque fetal single-cell and human and macaque adult single-nucleus clusters.

(A-F) t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for macaque fetal (**A**) dorsolateral prefrontal neocortex, (**B**) hippocampus, (**C**) amygdala, (**D**) striatum, (**E**) mediodorsal nucleus of the thalamus, and (**F**) cerebellar cortex cells clustering. Cells are colored by specimen (RMB683 – salmon; RMB691 – green).

(G-H) t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for adult dorsolateral prefrontal neocortex of **(G)** macaque and **(H)** human. Nuclei are colored by specimen (HSB106 – salmon; HSB189 – green; HSB340 – blue).

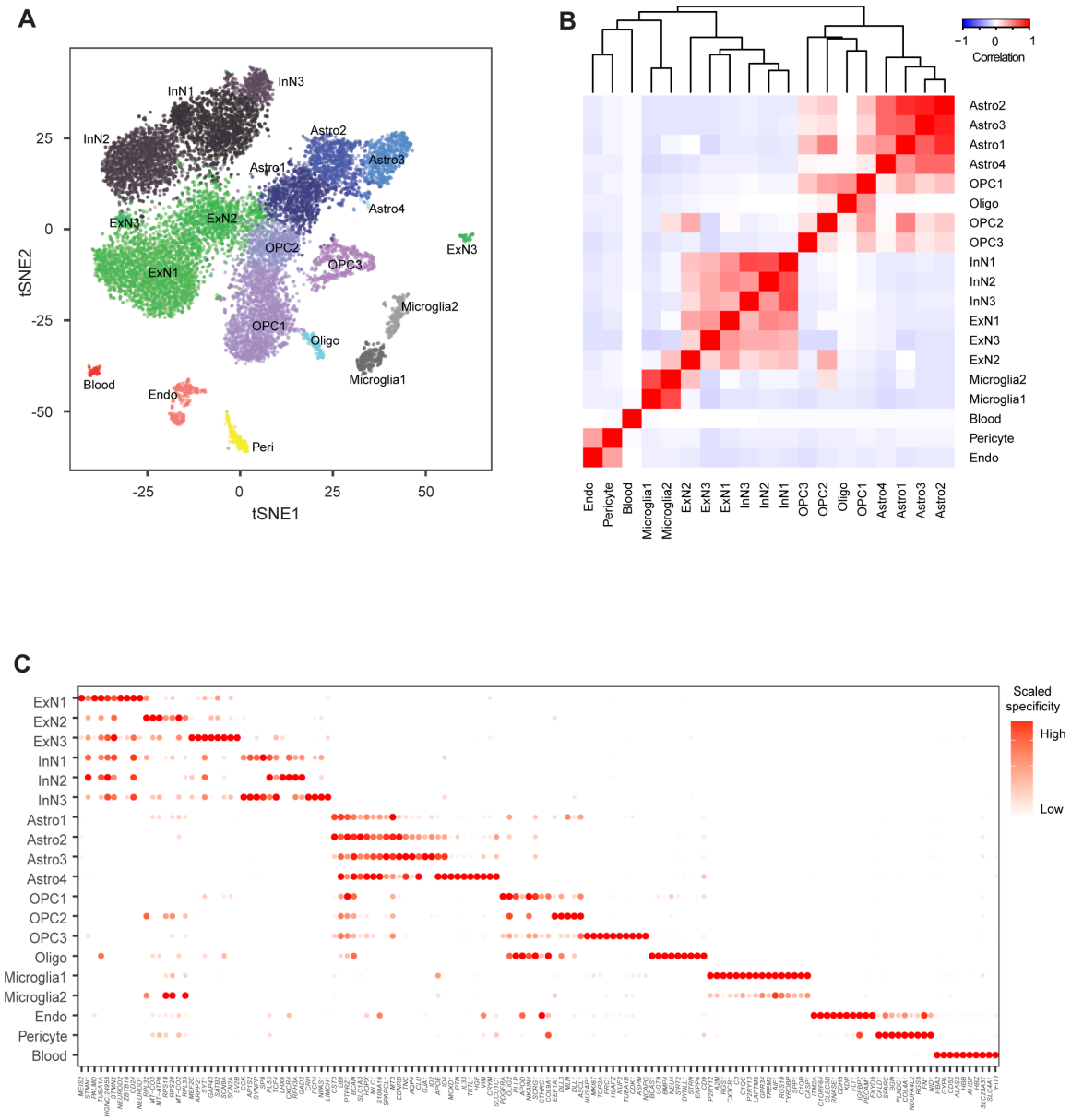


Fig. S7. Single cell RNA-Seq analysis of fetal macaque dorsolateral prefrontal cortex (DFC). **(A)** t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for macaque fetal dorsolateral prefrontal neocortex cells clustering (n = 2; total 15,427 cells). Cells are colored by cluster identity into 19 clusters. **(B)** The heatmap matrix presents the correlation between any pair of identified cell types. The dendrogram shows the unsupervised hierarchical clustering of the pairwise expression correlations. The color bar corresponds to the correlation coefficient value. **(C)** Dot plot illustrating the expression patterns of genes with highest

specificity to any single cell type. The color bar corresponds to the extent of scaled specificity score. ExN – excitatory neuron; InN – inhibitory neuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell.

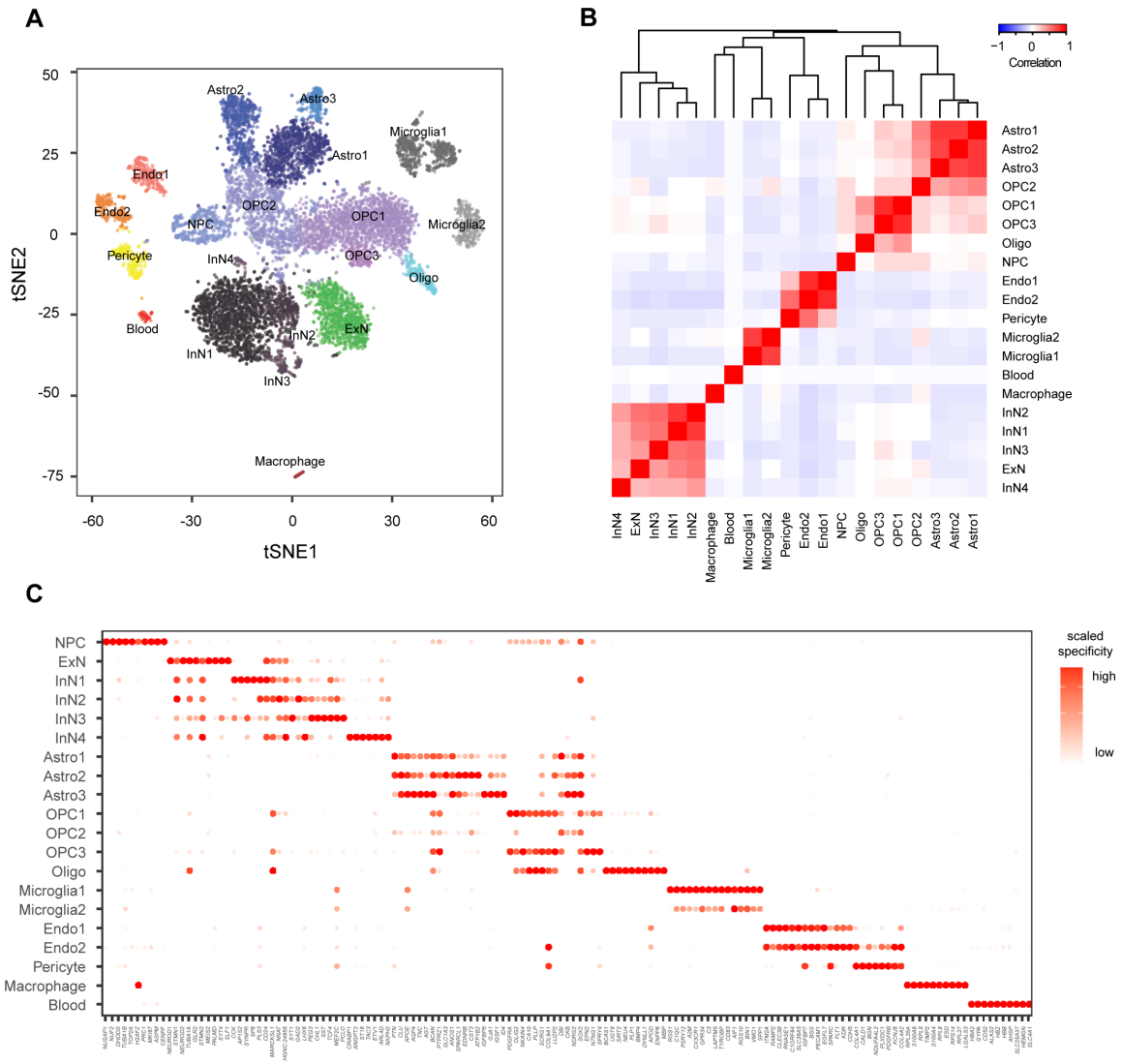


Fig. S8. Single cell RNA-Seq analysis of fetal macaque hippocampus (HIP). (A) t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for macaque fetal hippocampus cells clustering (n = 2; total 11,461 cells). Cells are colored by cluster identity into 20 clusters. **(B)** The heatmap matrix presents the correlation between any pair of identified cell types. The dendrogram shows the unsupervised hierarchical clustering of the pairwise expression correlations. The color bar corresponds to the correlation coefficient value. **(C)** Dot plot illustrating the expression patterns of genes with highest specificity to any single cell type. The color bar corresponds to the extent of scaled specificity score. NPC – neuronal progenitor cell;

ExN – excitatory neuron; InN – inhibitory neuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell.

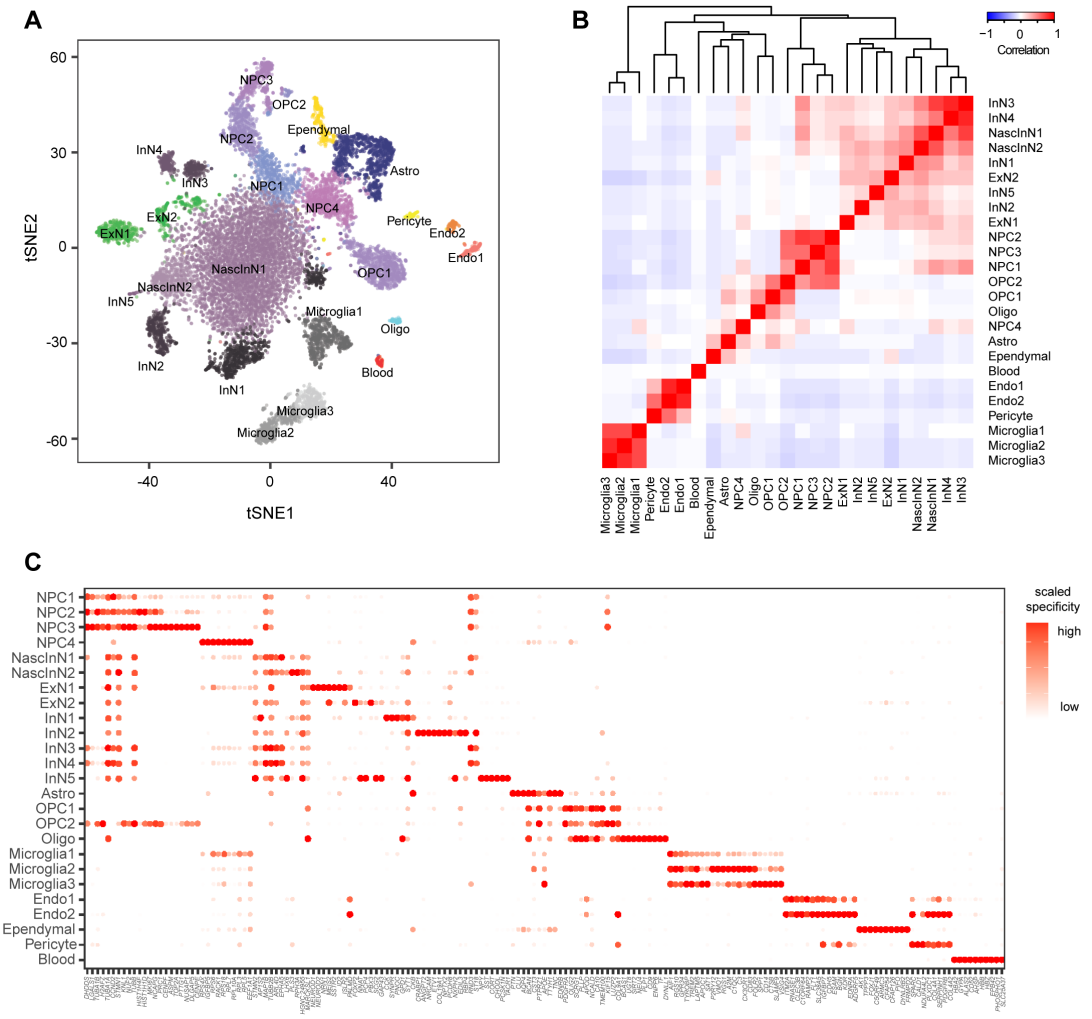


Fig. S9. Single cell RNA-Seq analysis of fetal macaque amygdala (AMY). (A) t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for macaque fetal amygdala cells clustering (n = 2; total 14,440 cells). Cells are colored by cluster identity into 25 clusters. **(B)** The heatmap matrix presents the correlation between any pair of identified cell types. The dendrogram shows the unsupervised hierarchical clustering of the pairwise expression correlations. The color bar corresponds to the correlation coefficient value. **(C)** Dot plot illustrating the expression patterns of genes with highest specificity to any single cell type. The color bar corresponds to the extent of scaled specificity score. NPC – neuronal progenitor cell; NascInN – nascent inhibitory neuron; ExN – excitatory neuron; InN – inhibitory neuron; Astro –

astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell.

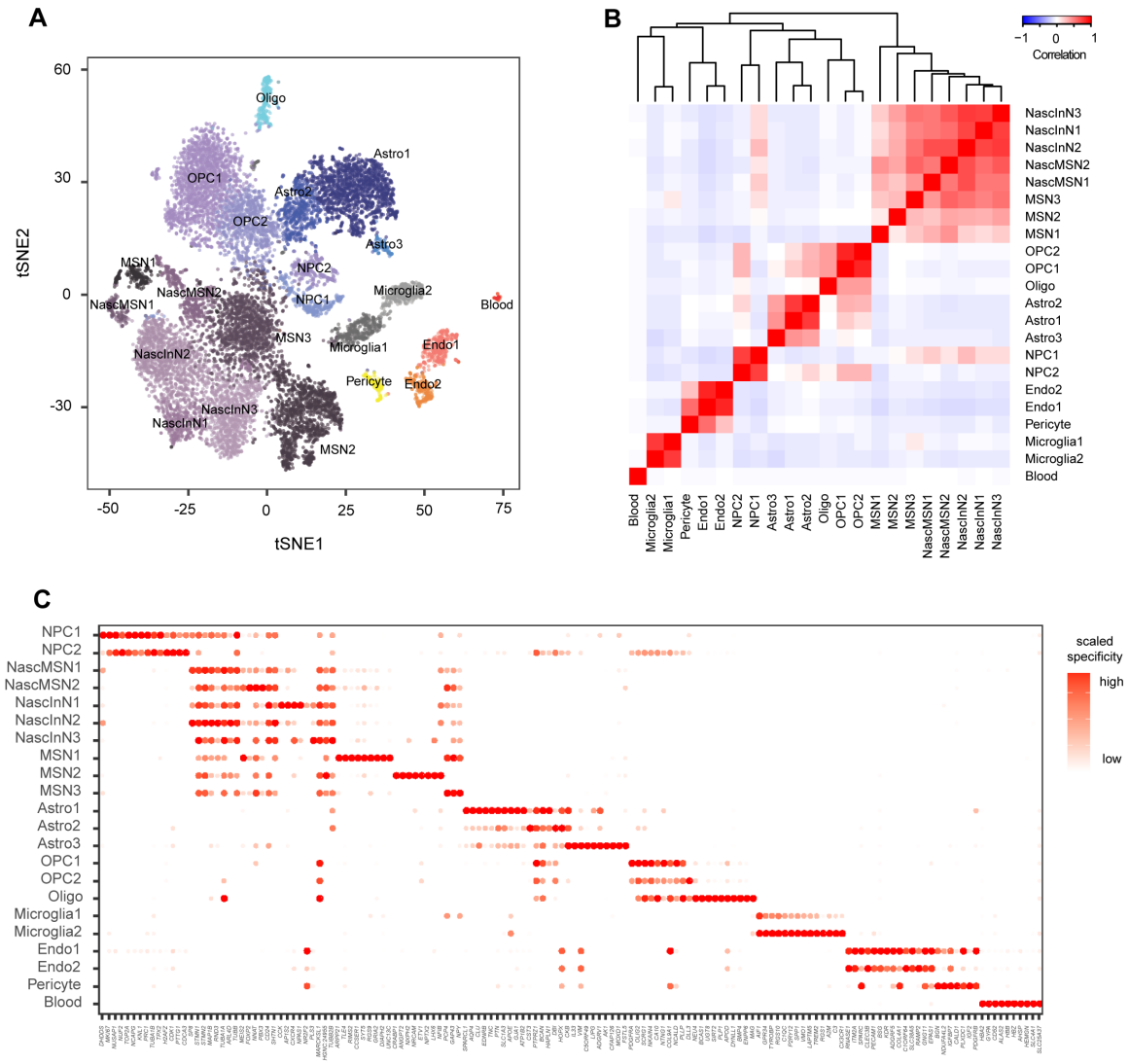


Fig. S10. Single cell RNA-Seq analysis of fetal macaque striatum (STR). (A) t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for macaque fetal striatum cells clustering (n = 2; total 14,255 cells). Cells are colored by cluster identity into 22 clusters. **(B)** The heatmap matrix presents the correlation between any pair of identified cell types. The dendrogram shows the unsupervised hierarchical clustering of the pairwise expression correlations. The color bar corresponds to the correlation coefficient value. **(C)** Dot plot illustrating the expression patterns of genes with highest specificity to any single cell type. The color bar corresponds to the extent of scaled specificity score. NPC – neuronal progenitor cell; NascMSN – nascent medium spiny neuron; NascInN – nascent inhibitory neuron; MSN –

medium spiny neuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell.

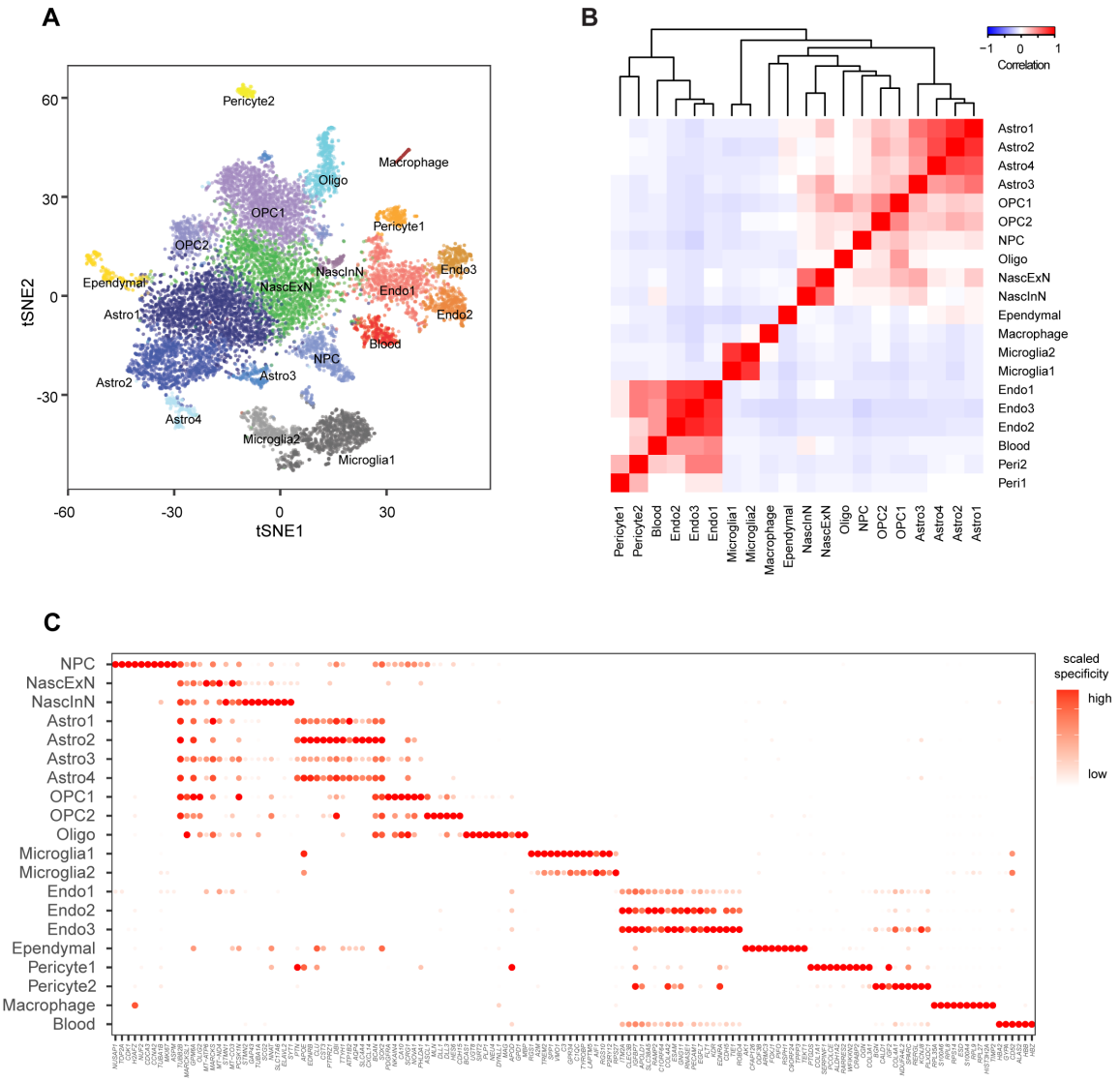


Fig. S11. Single cell RNA-Seq analysis of fetal macaque mediodorsal thalamic nucleus (MD). (A) t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for macaque fetal mediodorsal thalamic nucleus cells clustering (n = 2; total 12,696 cells). Cells are colored by cluster identity into 20 clusters. **(B)** The heatmap matrix presents the correlation between any pair of identified cell types. The dendrogram shows the unsupervised hierarchical clustering of the pairwise expression correlations. The color bar corresponds to the correlation coefficient value. **(C)** Dot plot illustrating the expression patterns of genes with highest specificity to any single cell type. The color bar corresponds to the extent of scaled specificity score. NPC – neuronal progenitor cell; NascExN – nascent excitatory neuron;

NascInN – nascent inhibitory neuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell**.**

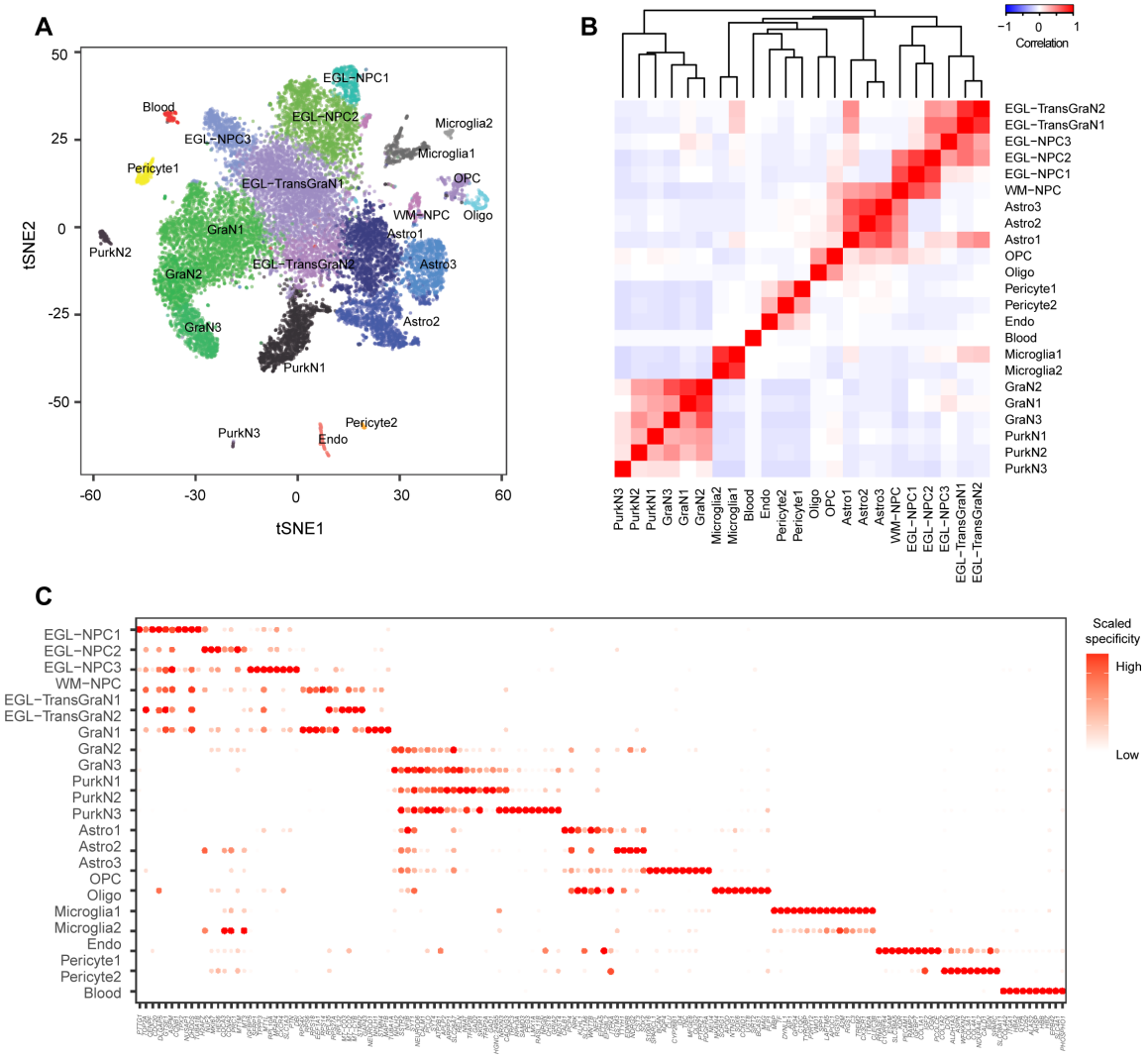


Fig. S12. Single cell RNA-Seq analysis of fetal macaque cerebellum (CBC). (A) t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for macaque fetal cerebellum cells clustering (n = 2; total 18,062 cells). Cells are colored by cluster identity into 23 clusters. **(B)** The heatmap matrix presents the correlation between any pair of identified cell types. The dendrogram shows the unsupervised hierarchical clustering of the pairwise expression correlations. The color bar corresponds to the correlation coefficient value. **(C)** Dot plot illustrating the expression patterns of genes with highest specificity to any single cell type. The color bar corresponds to the extent of scaled specificity score. EGL-NPC – external granular layer neuronal progenitor cell; WM-NPC – white matter neuronal progenitor cell; EGL-TransGraN –

external granular layer transition to granule neuron; GraN – granule neuron; PurkN – Purkinje neuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell.

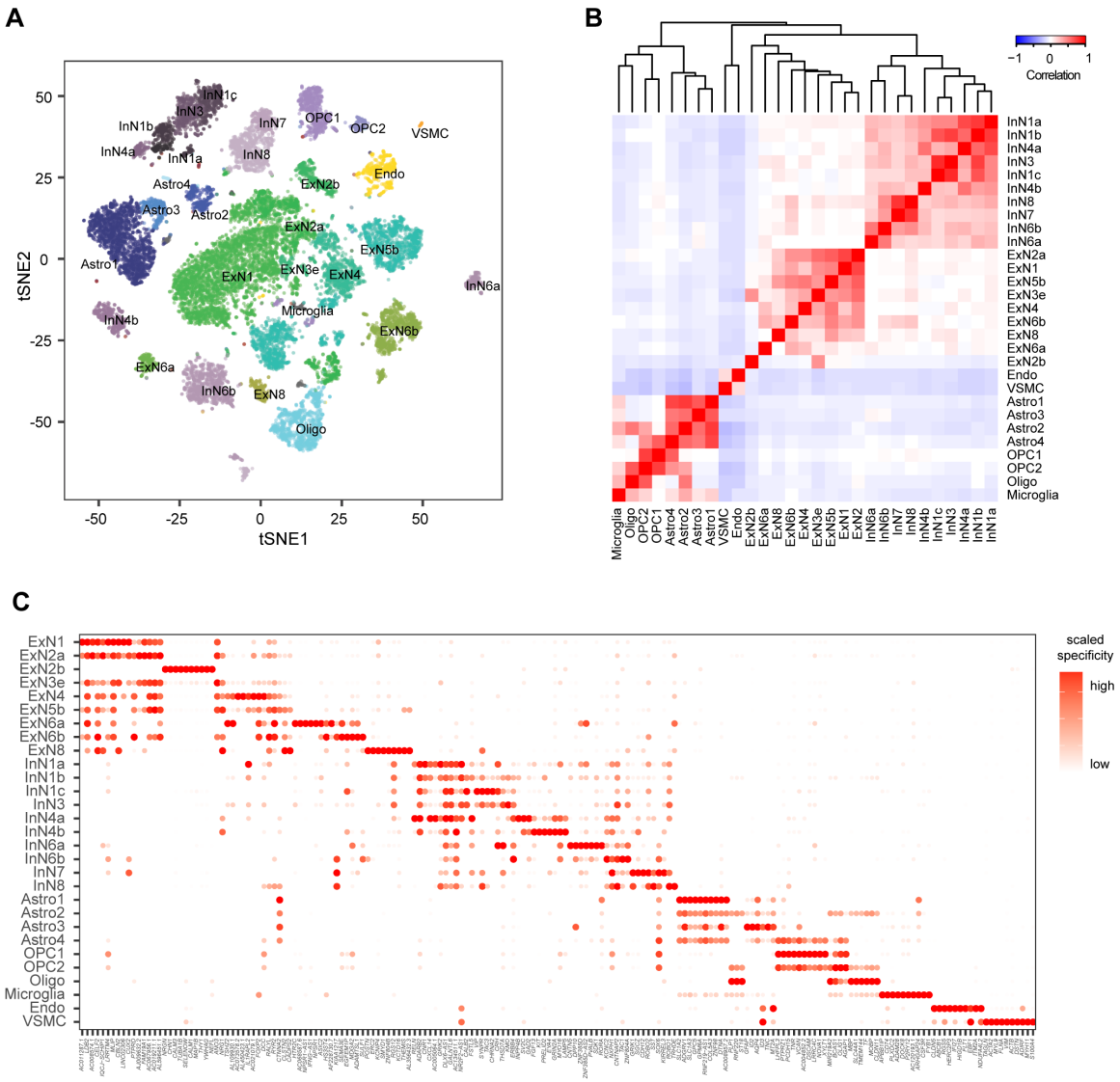


Fig. S13. Single nucleus RNA-Seq analysis of adult human dorsolateral prefrontal cortex (DFC). (A) t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for human adult dorsolateral prefrontal cortex cell nuclei clustering (n = 3; total 17,093 nuclei). Nuclei are colored by cluster identity into 29 clusters. **(B)** The heatmap matrix presents the correlation between any pair of identified cell types. The dendrogram shows the unsupervised hierarchical clustering of the pairwise expression correlations. The color bar corresponds to the correlation coefficient value. **(C)** Dot plot illustrating the expression patterns of genes with highest specificity to any single cell type. The color bar corresponds to the extent of scaled specificity score. ExN – excitatory neuron; InN – inhibitory neuron; Astro – astrocyte;

OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell; VSMC – vascular smooth muscle cell.

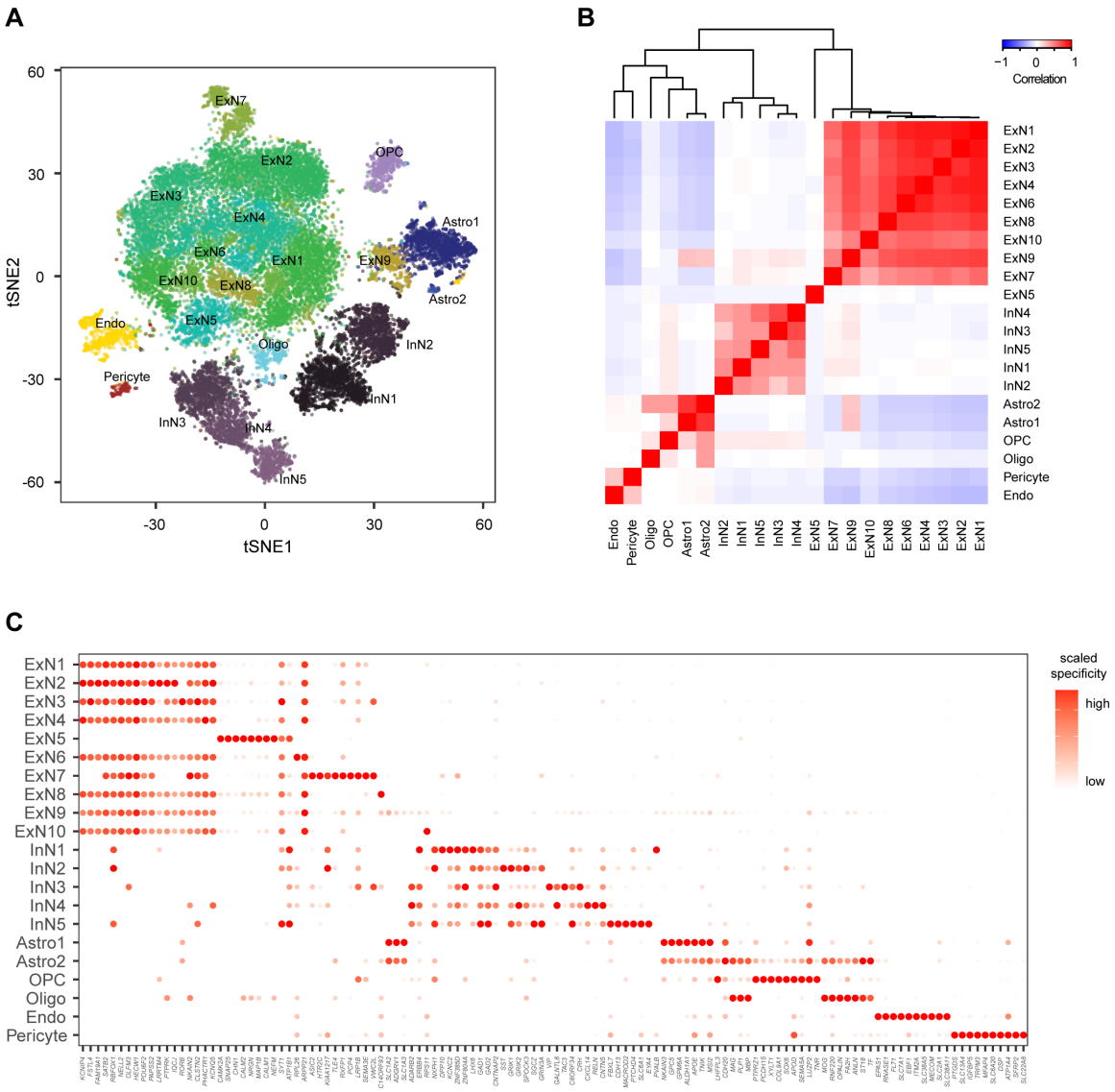


Fig. S14. Single nucleus RNA-Seq analysis of adult macaque dorsolateral prefrontal cortex (DFC). (A) t-distributed stochastic neighbor embedding (t-SNE) plot of gene expression relationships for macaque adult dorsolateral prefrontal cortex cell nuclei clustering (n = 3; total 26,933 nuclei). Nuclei are colored by cluster identity into 21 clusters. **(B)** The heatmap matrix presents the correlation between any pair of identified cell types. The dendrogram shows the unsupervised hierarchical clustering of the pairwise expression correlations. The color bar corresponds to the correlation coefficient value. **(C)** Dot plot illustrating the expression patterns of genes with highest specificity to any single cell type. The color bar corresponds to the extent of

scaled specificity score. ExN – excitatory neuron; InN – inhibitory neuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell.

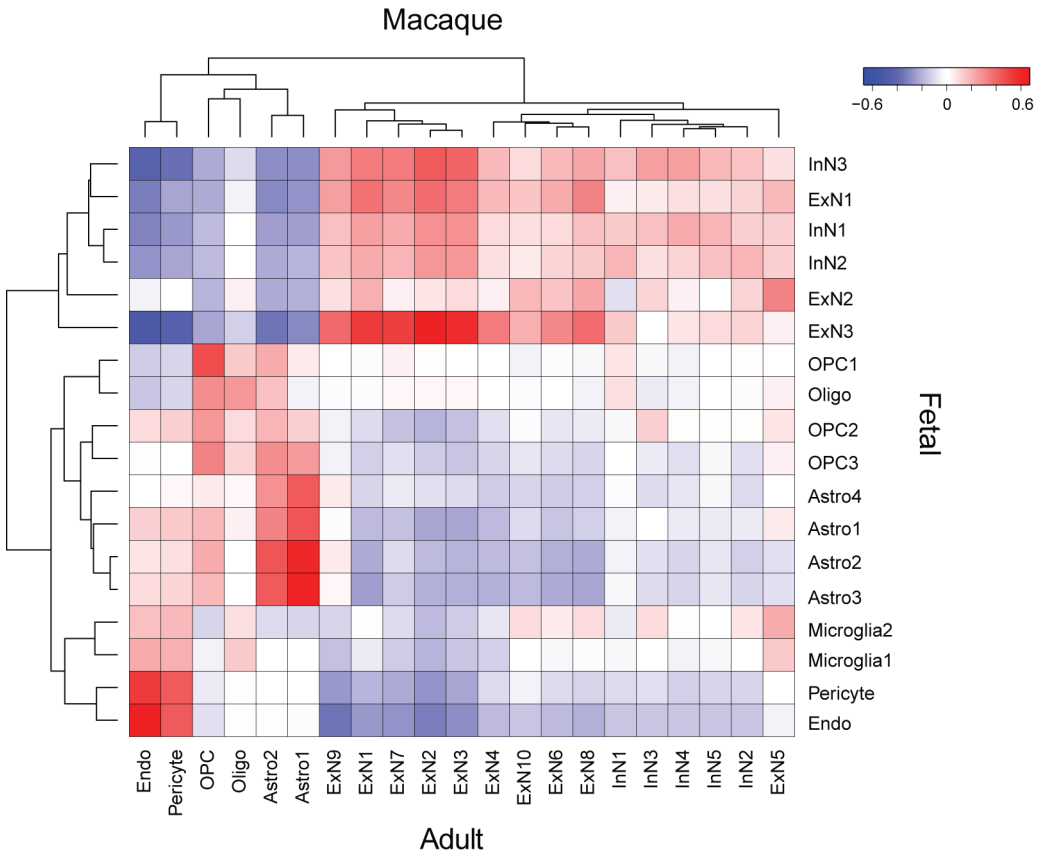


Fig. S15. Comparison of cell types between fetal and adult in monkey dorsal prefrontal cortex (DFC). The heatmap matrix displays the pairwise correlation of cell types identified from fetal and adult monkey DFC single cell/nucleus datasets. The dendrogram, using unsupervised hierarchical cluster analysis, represents the relationship between cell types between fetal cell types with adult cell types.

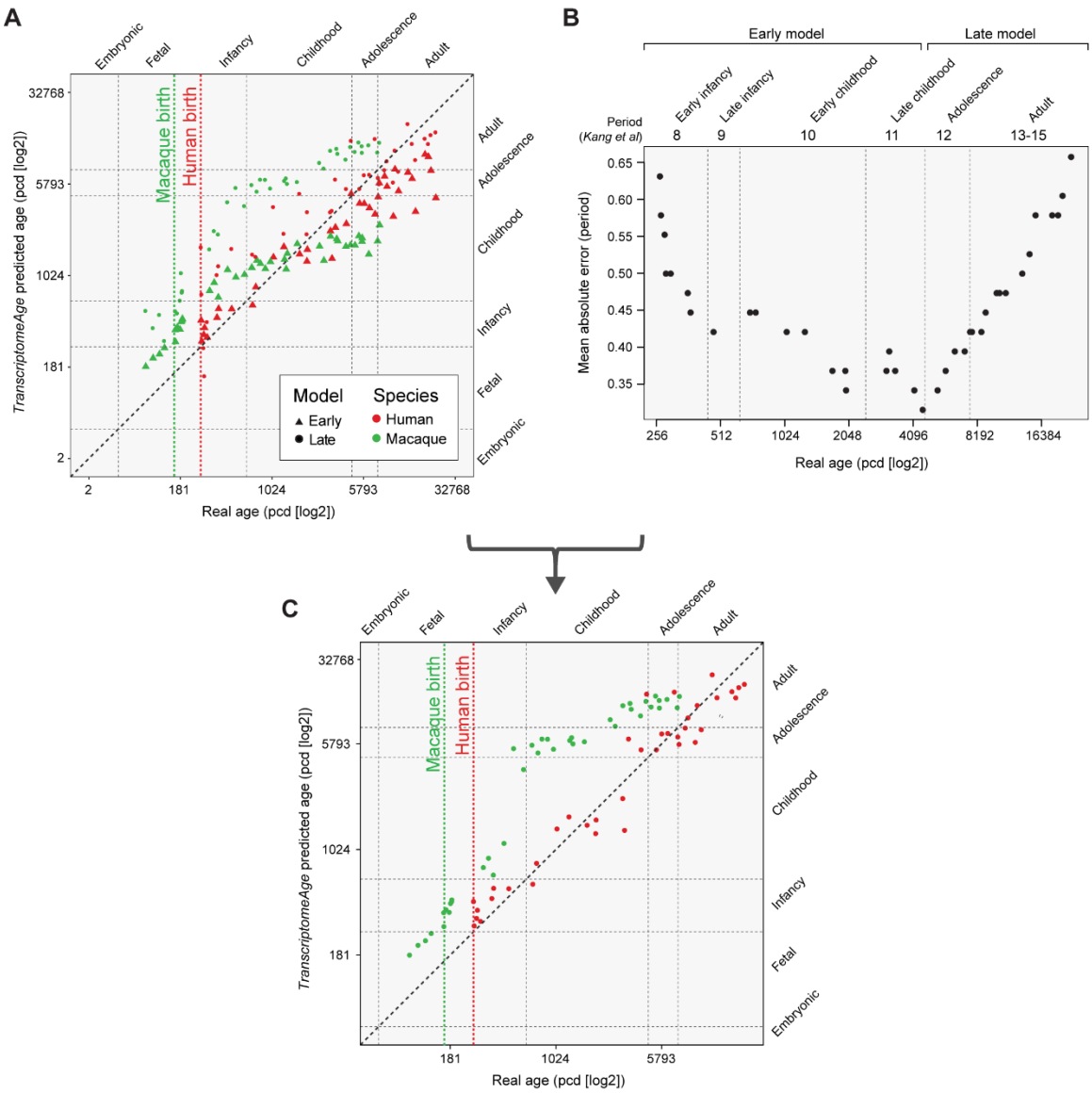


Fig. S16. Determination of model boundary for age-matching algorithm. (A) Predicted age of human (red) and macaque (green) dorsolateral prefrontal cortex samples from (*84*) using early (triangle) and late (circle) models (*31*). **(B)** The boundary determining which model to use for prediction was set to 12 years (real human age) because this age minimized the prediction errors; the mean prediction error is 0.32 developmental periods for human samples. **(C)** The predicted age of human and macaque samples in (*84*).

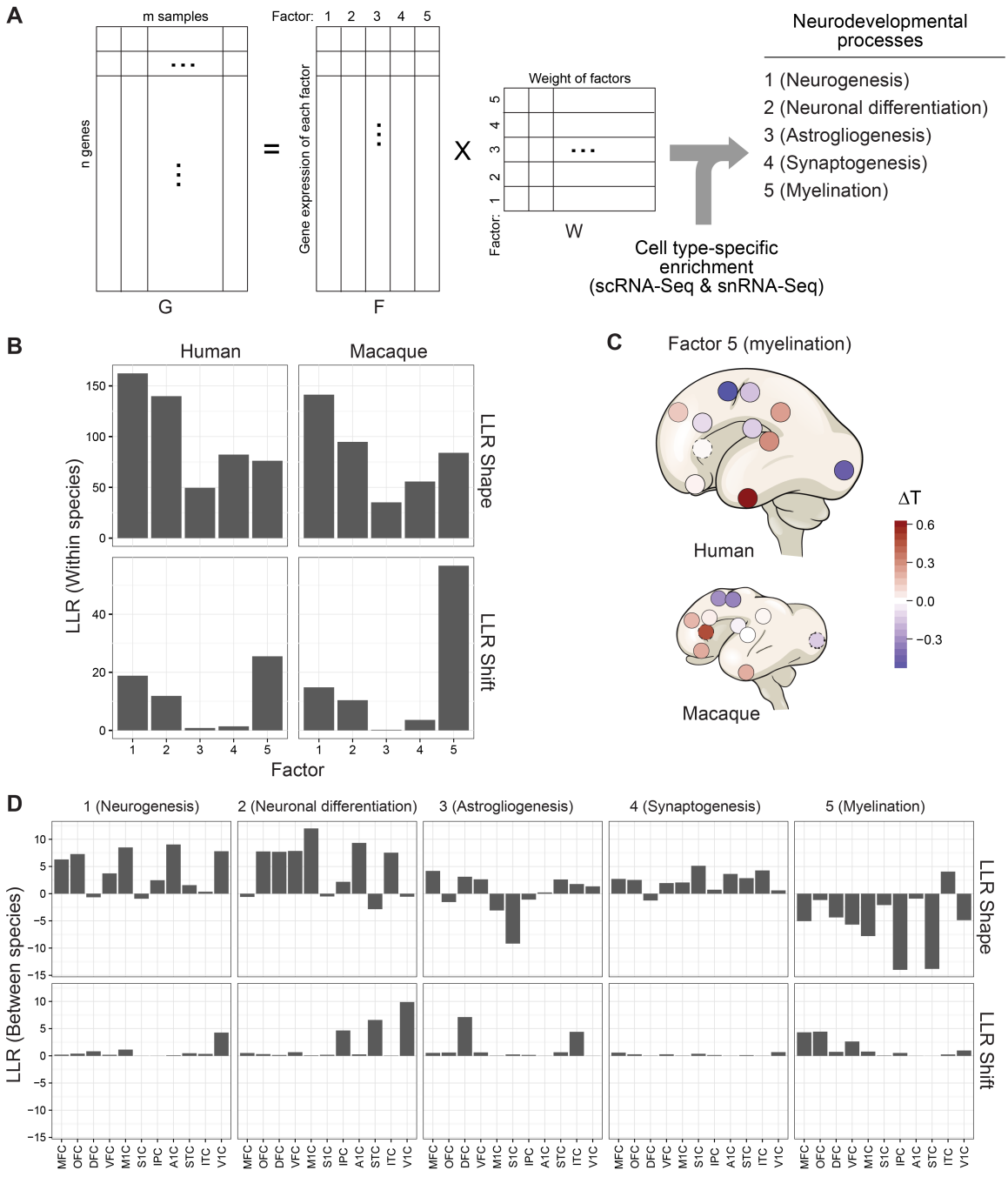


Fig. S17. Identification of five main transcriptomic signatures. (A) Schematic representing the non-negative matrix factorization (NMF) algorithm. The gene expression matrix of human neocortical samples (G) was factorized into a factor signature matrix (F) and a factor weight matrix (W; (*32*). Cell type-specific enrichment allowed us to ascribe each signature to specific neurodevelopmental processes. **(B)** Comparison of trajectories of the five main transcriptomic

signatures between neocortical areas within a given species. LLR Shape (*32*) measures the likelihood that the trajectories are conserved across either neocortical areas in a given species (B) or between species (D) and LLR Shift measures the likelihood that the trajectories are temporally shifted between areas (B) or species (D). The trajectory of factor 5 (ascribed to myelination) shows broad conservation across neocortical areas and a high likelihood of temporal shift between neocortical areas. **(C)** The heterochrony between neocortical areas for Factor 5 measured in log2 (PCD) indicates heterochronic development within the neocortex. Blue indicates earlier expression; red indicates later expression. ΔT: temporal difference in gene expression among regions. **(D)** Comparison of trajectories of the five transcriptomic signatures between human and macaque. The low LLR Shape values of factor 5 suggest the progression of oligodendrocyte maturation and myelination exhibits species-distinct trajectories.

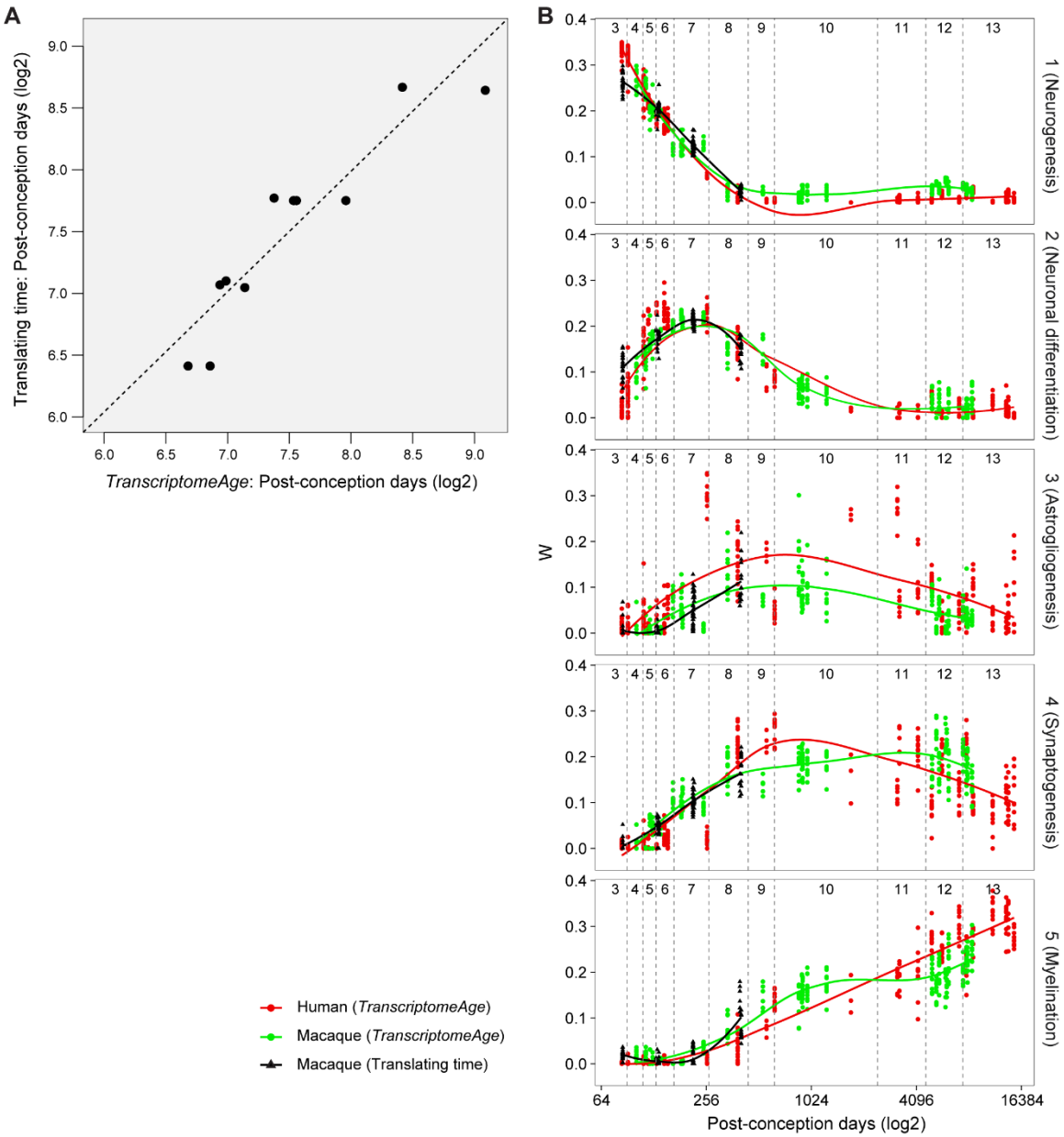


Fig. S18. Age-matching comparison with Translating Time (*29*). (A) Scatter plot comparing the predicted age of macaque brains based on the transcriptome using *TranscriptomeAge* and the matching age based on Translating Time (www.translatingtime.org). Translating Time, which predicts the developmental age of macaques corresponding to up to 347 postconceptional days in humans, predicts younger developmental ages for early prenatal macaque brains than our model. **(B)** The trajectories of the five main transcriptomic signatures based on the age predicted by Translating Time (black) or *TranscriptomeAge* (green).

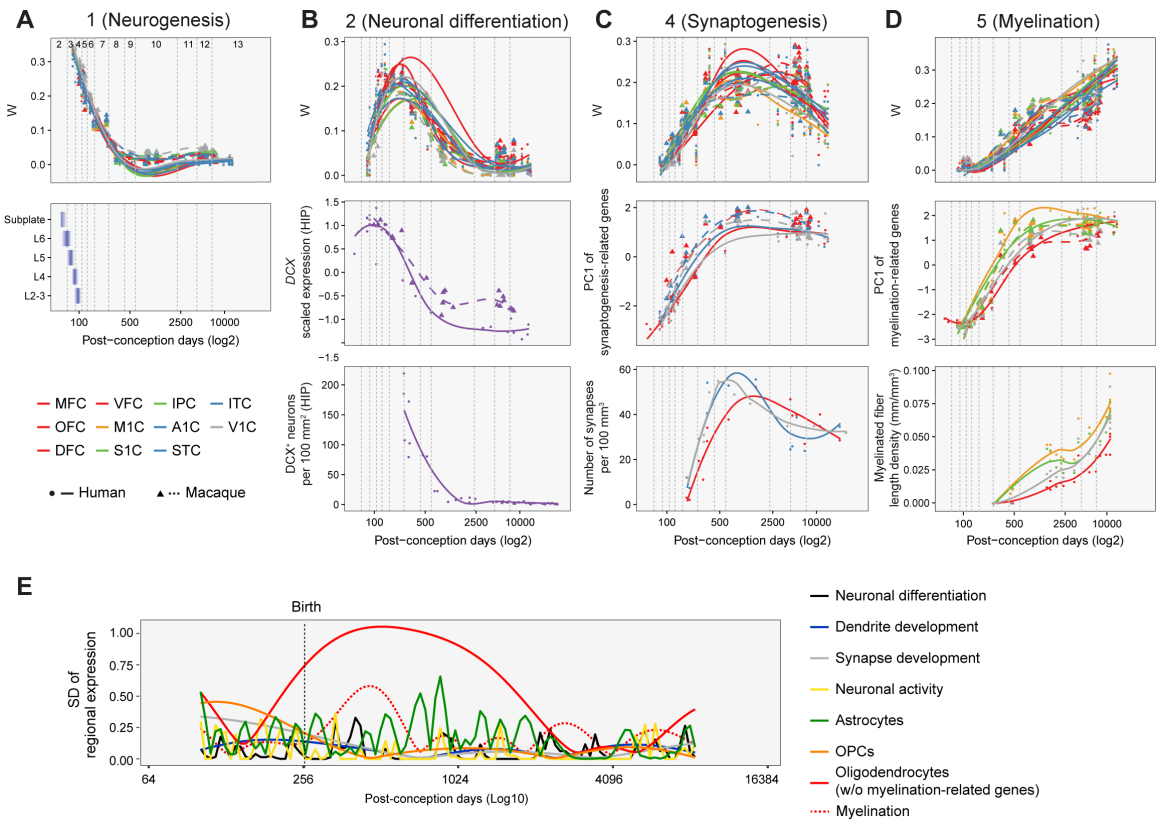


Fig. S19. Transcriptomic signatures correlate with biological processes. (A) Plots depicting the weight (W) of the first transcriptomic signature (top), which we associated with neurogenesis, and the start, peak (dark blue), and end of neurogenesis in human subplate and layers (L) 2-6 (*40*) (bottom). **(B)** Plot showing the weight (W) of the second transcriptomic signature (top), which we associated with neuron differentiation. The expression of *DCX* in human and macaque hippocampus was used as a proxy for neural differentiation (middle; note that macaque hippocampus retains a higher expression of DCX during postnatal development, extending into adulthood). The number of DCX-immunopositive neurons in the human hippocampus (*41*) (bottom) shows a similar pattern to the expression profile of *DCX* in human. **(C)** Plot depicting the weight (W) of the fourth transcriptomic *signature* (top), associated to synaptogenesis. The pattern of the first principal component (PC1) of synaptogenesis-related genes, as curated in (*29*) (middle), is developmentally similar to the transcriptomic signature 4, and to the developmental profile of the number of synapses per area in different cortical areas in human (*42*) (bottom). **(D)**

Plot showing the weight (W) of the fifth transcriptomic signature (top), which we associated with myelination (top). The first principal component (PC1) of myelination-related genes, as curated in (*29*), matches the developmental pattern of the transcriptomic signature 5 (middle). This correlation is further corroborated by the developmental profile of myelinated fiber length density in several human cortical areas (*43*). (**E**) Variation in oligodendrocyte (not including myelination-related genes) and myelination derived genes signatures and myelination-associated genes peak in early postnatal development, as evidenced by the standard deviation of the fitted regional mean, driving inter-regional variation during this and neighboring periods. Other biological processes and cell-type signatures indicated do not show a major inter-areal peak at any time point.

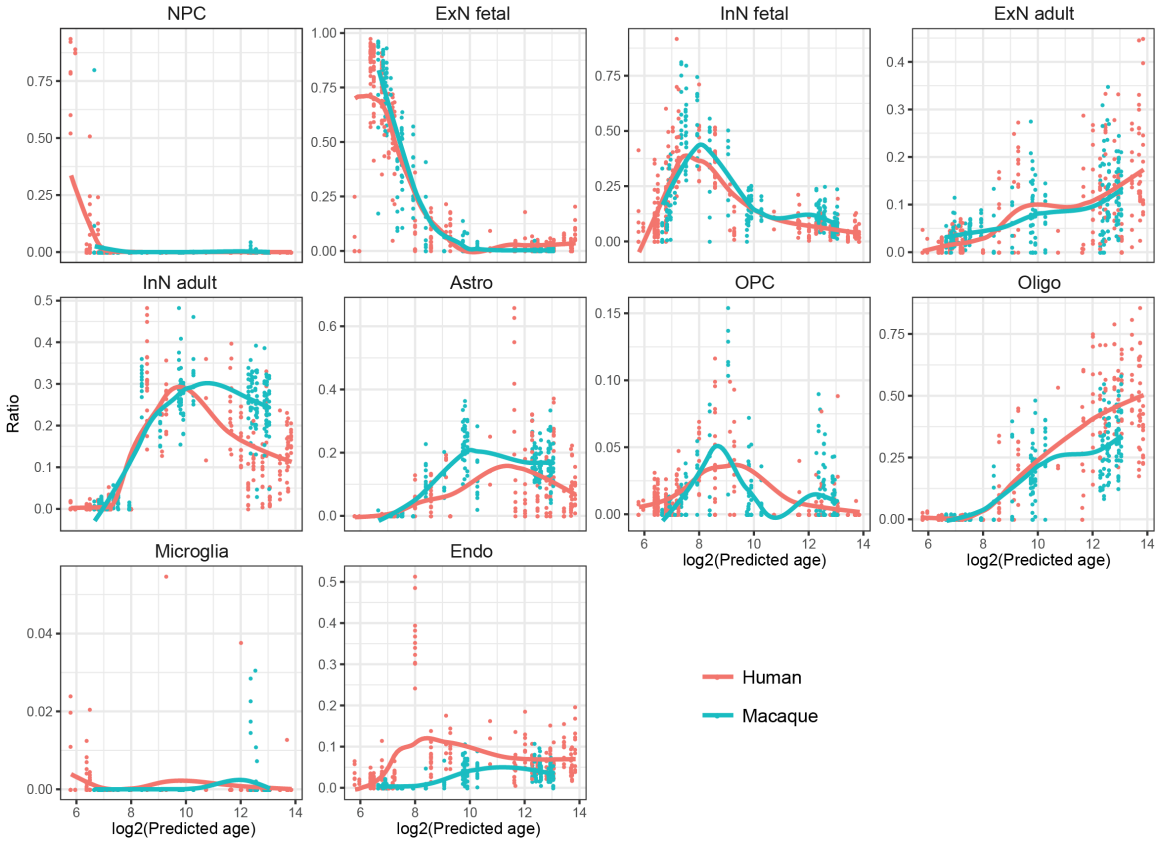


Fig. S20. Bulk tissue deconvolution with single cell/nucleus RNA-seq information. The proportion of the tissue-level transcriptome contributed by cell type-specific signatures was estimated based on single-cell transcriptome data (*33*). The trajectories of progenitors (NPC), neurons (fetal and adult), astrocytes, and oligodendrocytes match those of the five transcriptomic signatures described in Figure 1. NPC – neural progenitor cell; ExN – excitatory neuron; InN – interneuron; Astro – astrocytes; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell.

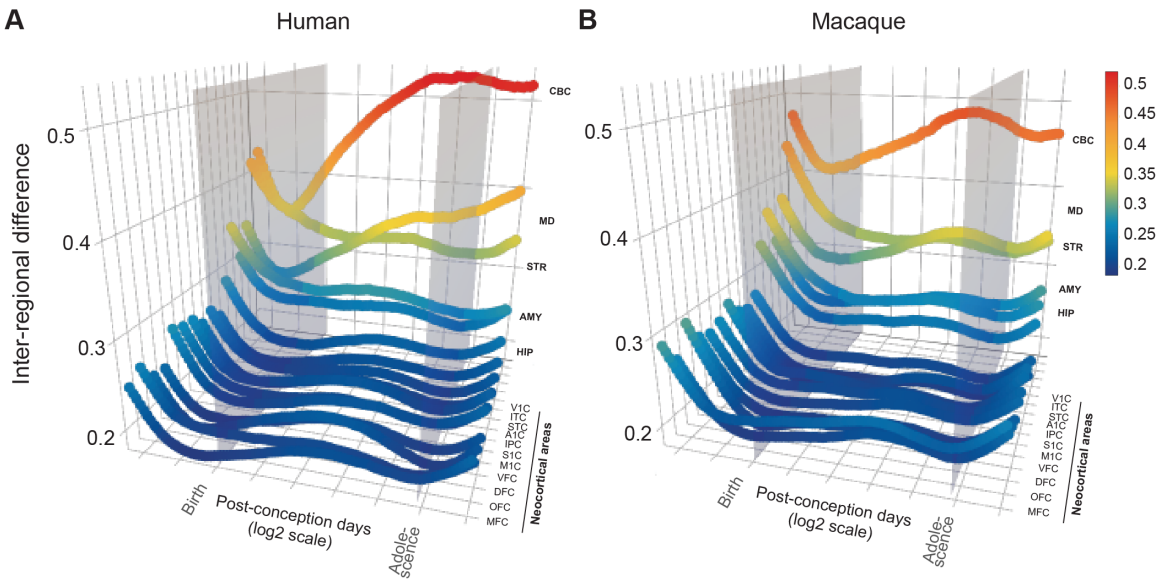


Fig. S21. Inter-regional transcriptomic differences in human and macaque brain development exhibit a cup-shaped pattern. Inter-regional differences were measured as the average distance of each region to all other regions in the **(A)** human and **(B)** macaque brain. The upper quartile difference among all genes was plotted. The gray planes represent the transition from prenatal to postnatal development (birth) and from adolescence to adulthood.

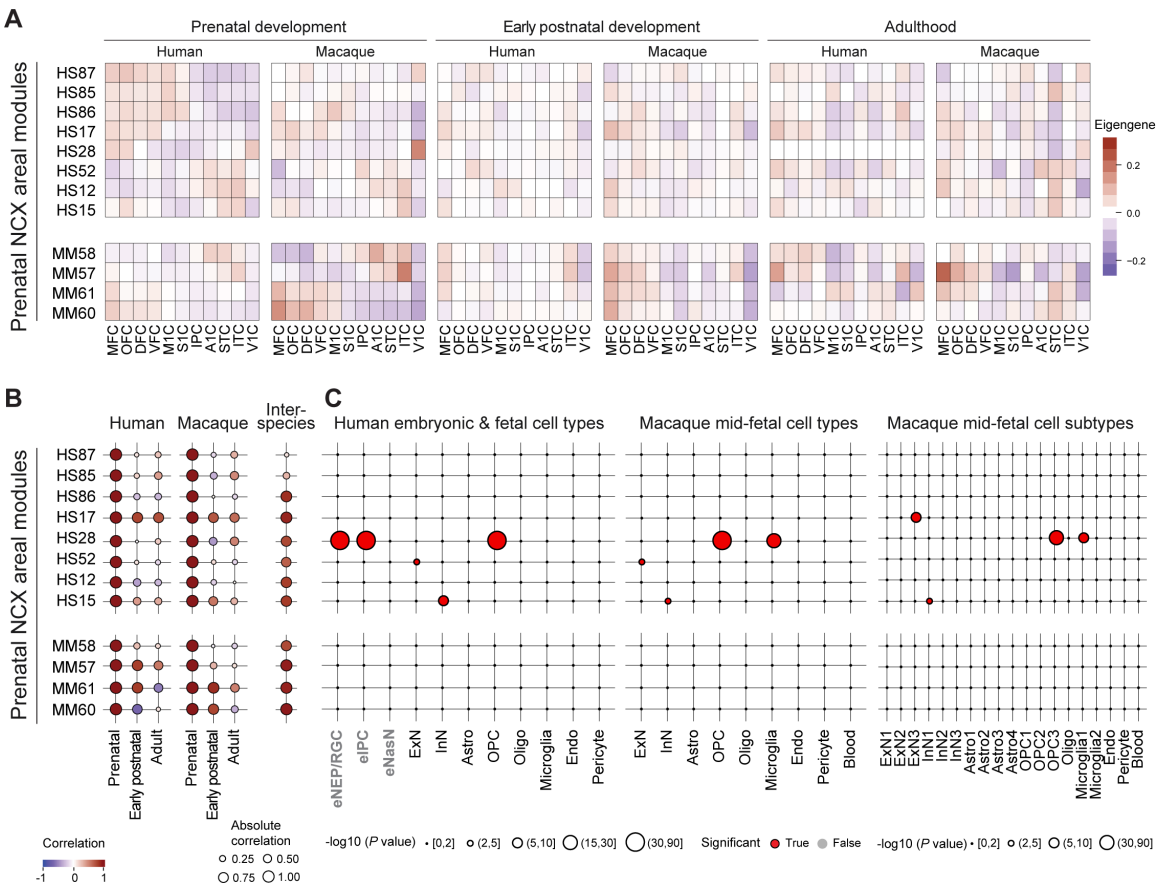


Fig. S22. Species-distinct expression and cell-type enrichment in prenatal modules. (A) Heatmap displaying the average of module eigengenes by neocortical areas in both species. For human (HS) modules, values shown are those observed from human brain and we calculated for macaque neocortex; for macaque (MM) modules, we report observed eigengenes in macaque and estimated eigengenes in the human neocortex. **(B)** Correlation of regional expression patterns between phases within species (left) and between species for the phase in which the module is identified (right). **(C)** Cell type enrichment of the modules. Significance is based on the Kolmogorov–Smirnov test ($P < 0.01$). eNEP/RGC – embryonic neural epithelial progenitor/radial glial cell; eIPC – embryonic intermediate progenitor cell; eNasN – embryonic nascent neuron; ExN – excitatory neuron; InN – interneuron; Astro – astroglial lineage; OPC – oligodendrocyte progenitor cell; Oligo – oligodendroglial lineage; Endo – endothelial cell.

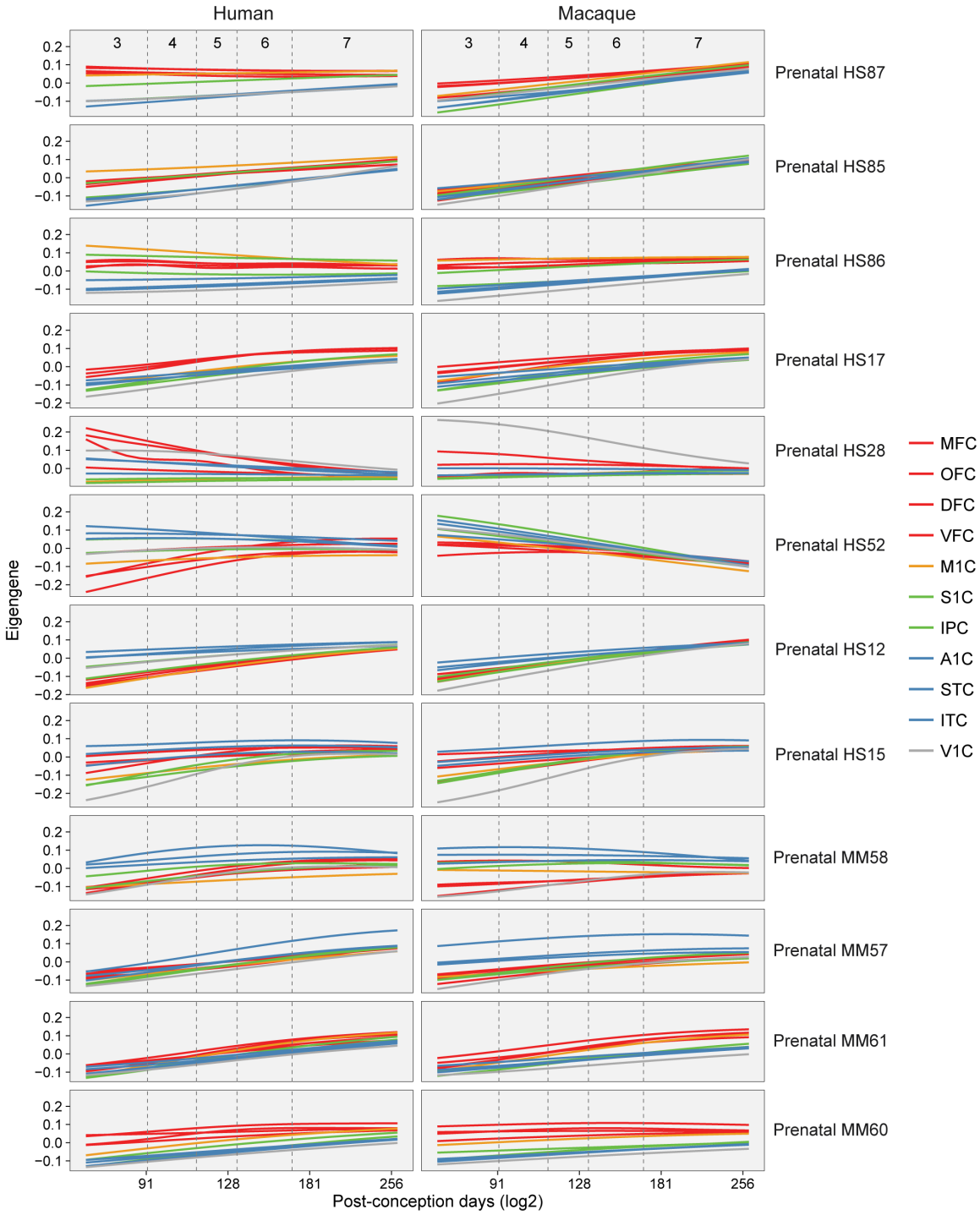


Fig. S23. Transient gradient-like expression patterns in prenatal modules. The expression pattern of each module is shown as the eigengenes in the neocortical areas of human (left) and macaque (right). Eigengenes were calculated based on predicted expression at evenly spaced time

points from Gaussian process models (*32*). For human modules, we calculated the eigengenes in the macaque neocortex and for macaque modules we calculated the eigengenes in the human neocortex.

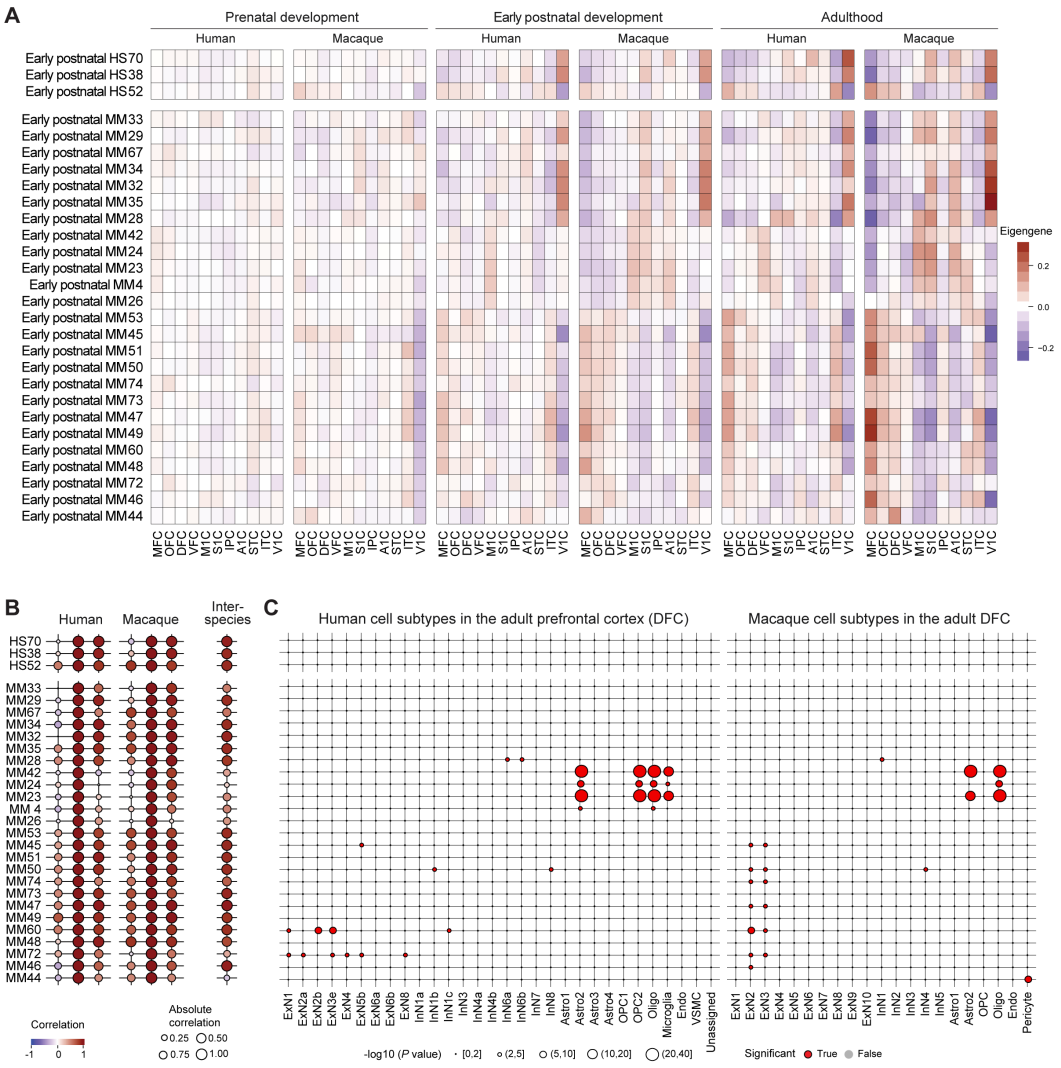


Fig. S24. Species-distinct expression and cell-type enrichment in inter-areal differentially expressed early postnatal modules. (A) Heatmap displaying the average of module eigengenes by neocortical areas in both species. For human modules, we calculated the eigengenes in the macaque neocortex and for macaque modules we calculated the eigengenes in the human neocortex. **(B)** Correlation of regional expression pattern between phases within species (left) and between species for the phase where the module is identified (right). **(C)** Cell type enrichment of the modules. Significance is based on the Kolmogorov–Smirnov test ($P < 0.01$). ExN – excitatory

neuron; InN – interneuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell; VSMC – vascular smooth muscle cell.

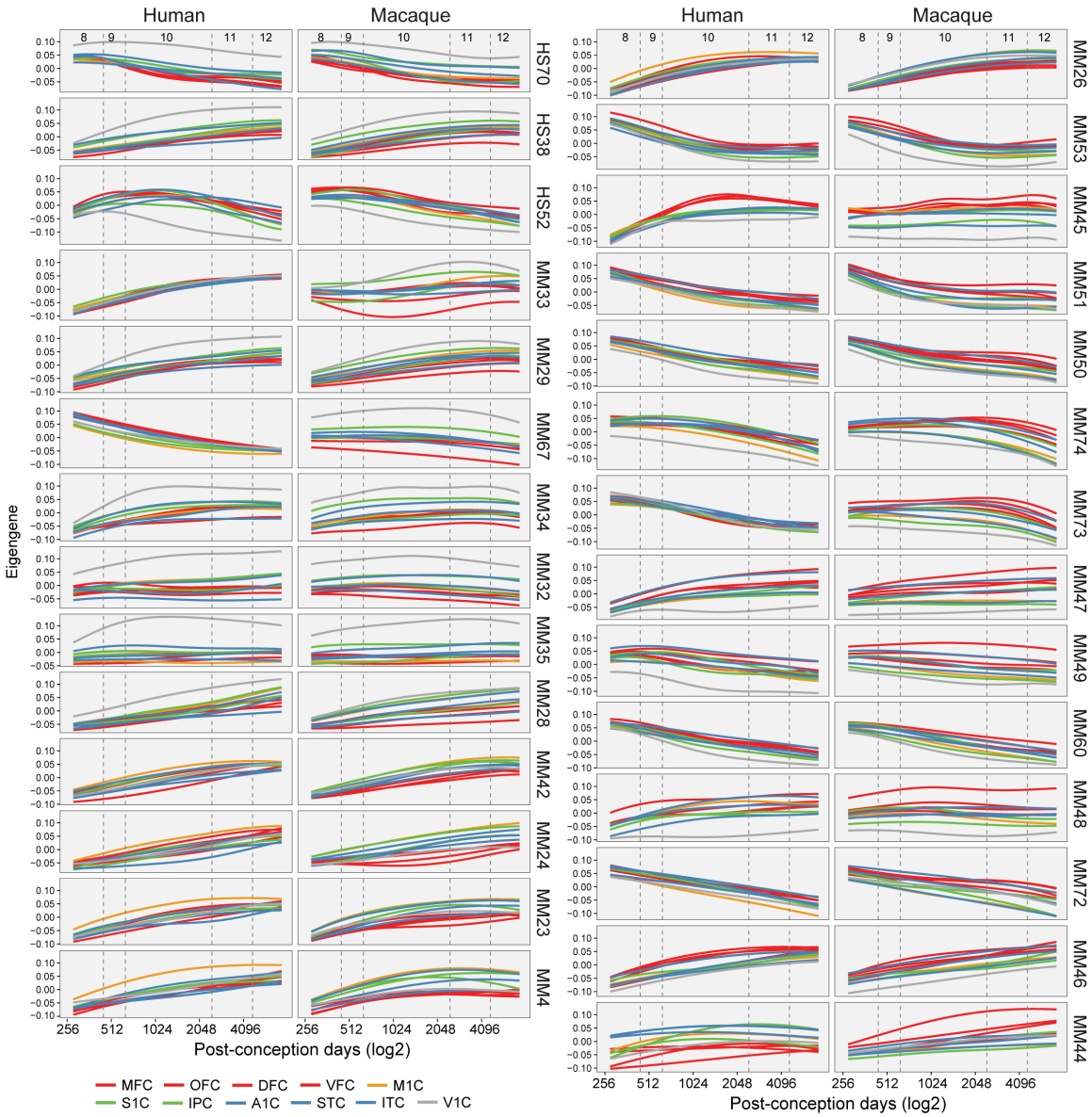


Fig. S25. Stable inter-areal differential expression among early postnatal modules. The expression pattern of each module is shown as the eigengenes in the neocortical areas of human (left) and macaque (right). Eigengenes were calculated based on predicted expression at evenly spaced time points from Gaussian process models (*32*). For human modules, we calculated the eigengenes in the macaque neocortex and vice versa.

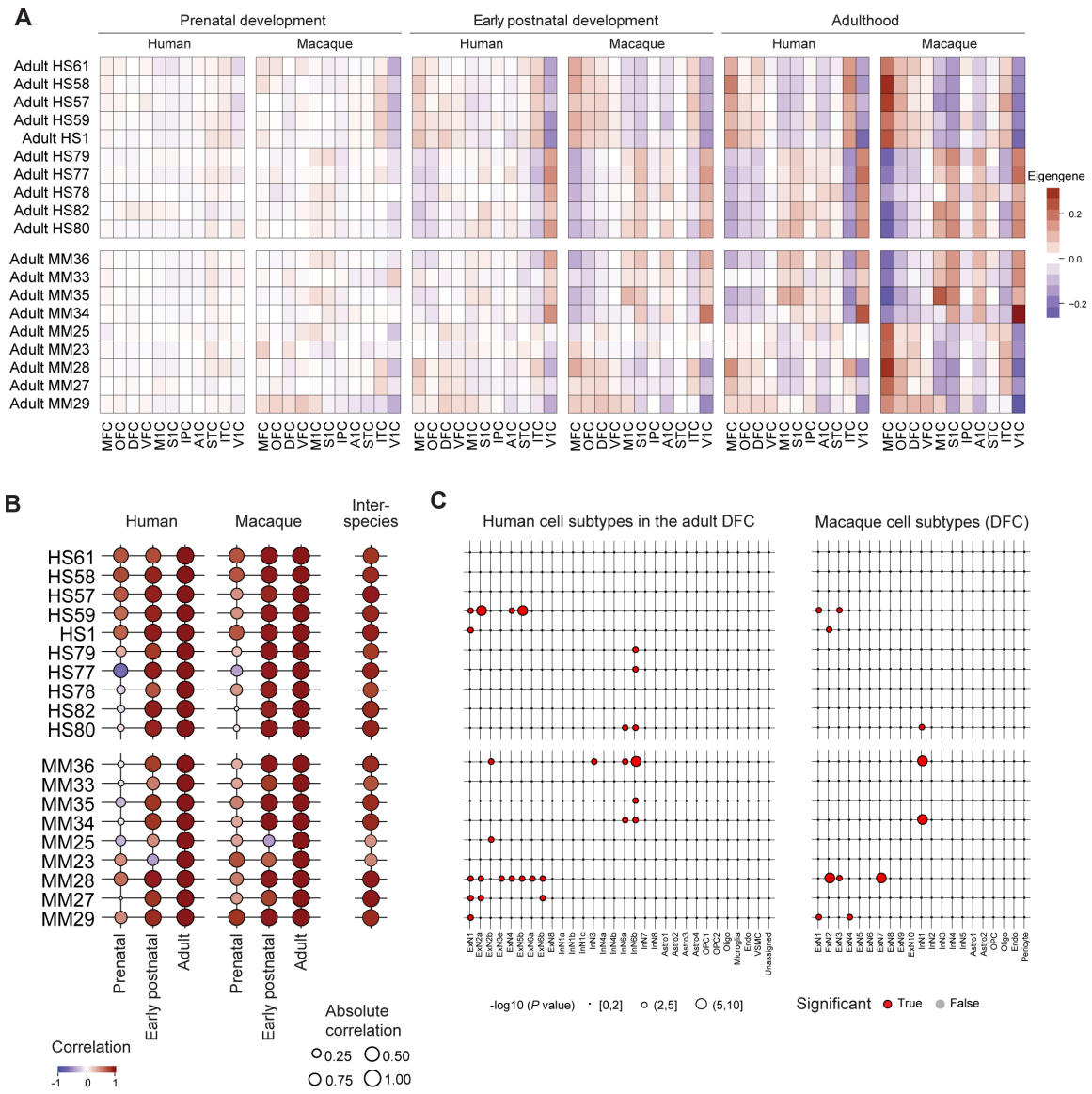


Fig. S26. Species-distinct expression and cell-type enrichment in inter-areal differentially expressed adult modules. (A) Heatmap displaying the average of module eigengenes by neocortical areas in both species. For human modules, we calculated the eigengenes in the macaque neocortex and for macaque modules we calculated the eigengenes in the human neocortex. **(B)** Correlation of regional expression patterns between phases within species (left) and between species for the phase where the module is identified (right). **(C)** Cell type enrichment of the modules. Significance is based on the Kolmogorov–Smirnov test ($P < 0.01$). ExN – excitatory neuron; InN – interneuron; Astro – astrocyte; OPC – oligodendrocyte

progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell; VSMC – vascular smooth muscle cell.

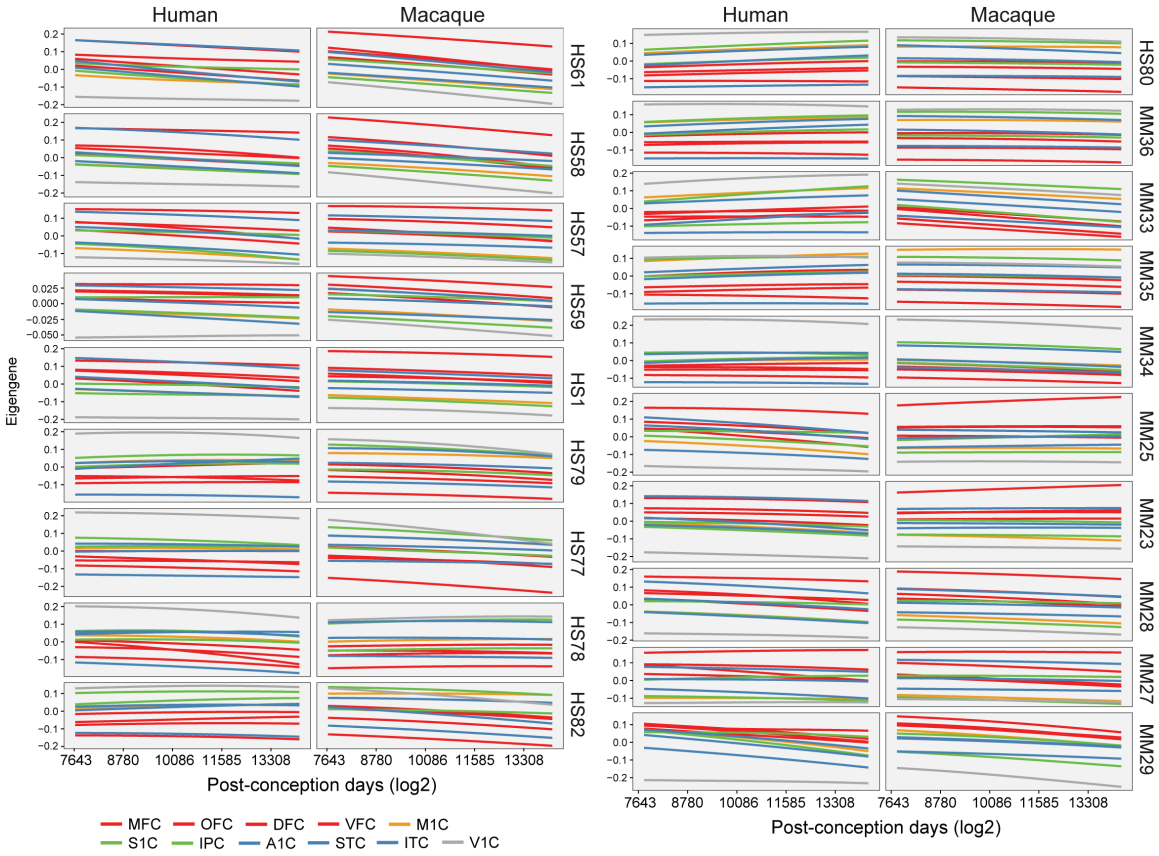


Fig. S27. Stable inter-areal differential expression among adult modules. The expression pattern of each module is shown as the eigengenes by neocortical areas in human (left) and macaque (right). Eigengenes were calculated based on predicted expression at evenly spaced time points from Gaussian process models (*32*). For human modules, we calculated the eigengenes in the macaque neocortex and for macaque modules we calculated the eigengenes in the human neocortex.

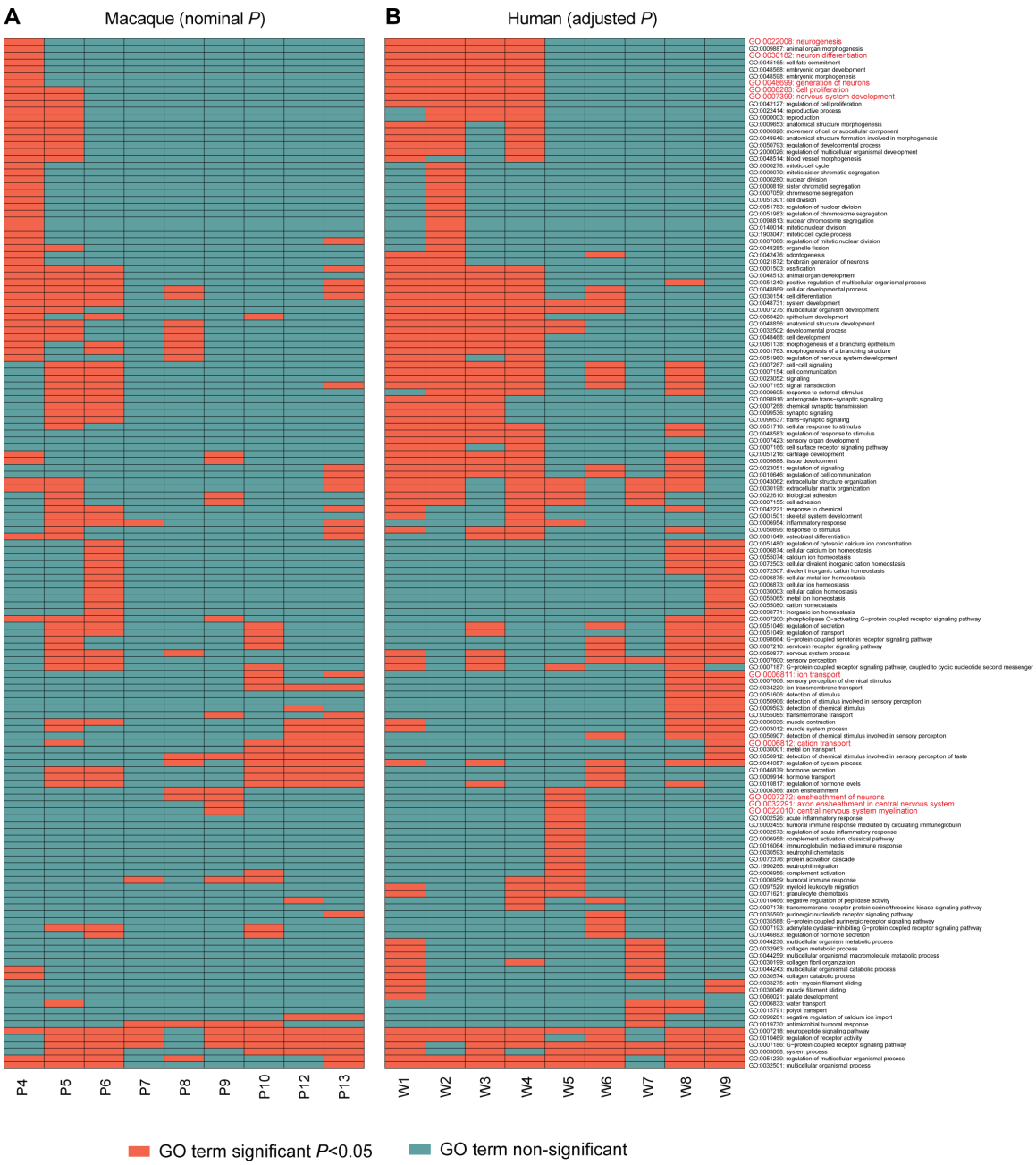


Fig. S28. Gene Ontology enrichment analysis of top variable genes per window. (A-B) The top 30 FDR-corrected significant GO terms are shown for each period in (**A**) macaque and for each developmental window (*33*) for (**B**) human, clustered by their patterns of significance across periods/windows. A selection of GO terms of special interest are highlighted in red.

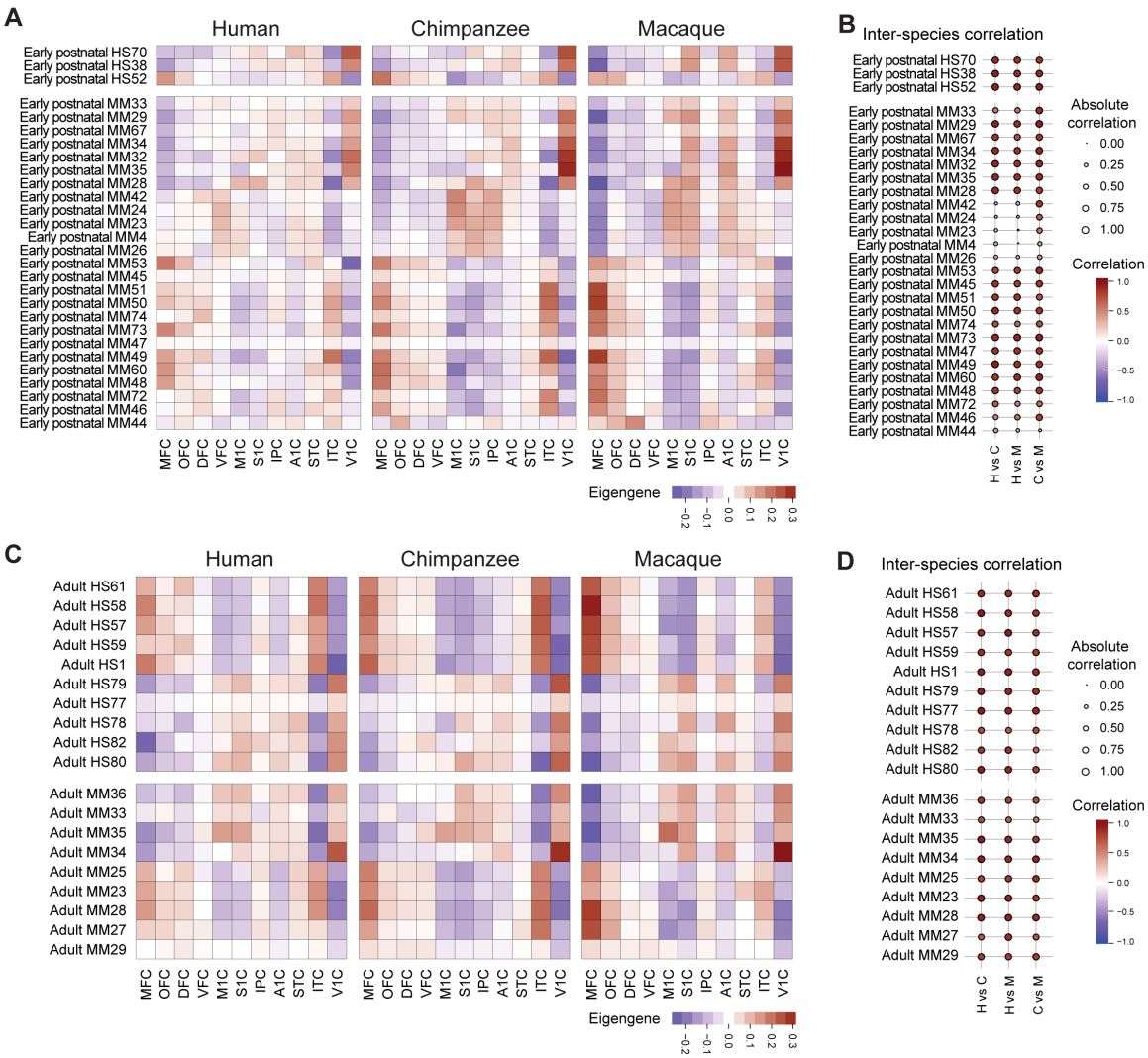


Fig. S29. Evolutionary divergence of inter-areal differentially expressed early postnatal and adult modules. (A) Heatmap displaying the eigengenes of inter-areal differentially expressed early postnatal modules calculated for human, chimpanzee, and macaque using the data set from Sousa et al (*34*). **(B)** Correlation of regional expression patterns of early postnatal modules between each pair of species. HC: human vs. chimpanzee, HM: human vs. macaque, and CM: chimpanzee vs. macaque. **(C)** Heatmap displaying the eigengenes of inter-areal differentially expressed adult modules calculated for human, chimpanzee, and macaque using a data set from Sousa et al. (*34*). **(D)** Correlation of regional expression pattern of adult modules between each pair of species.

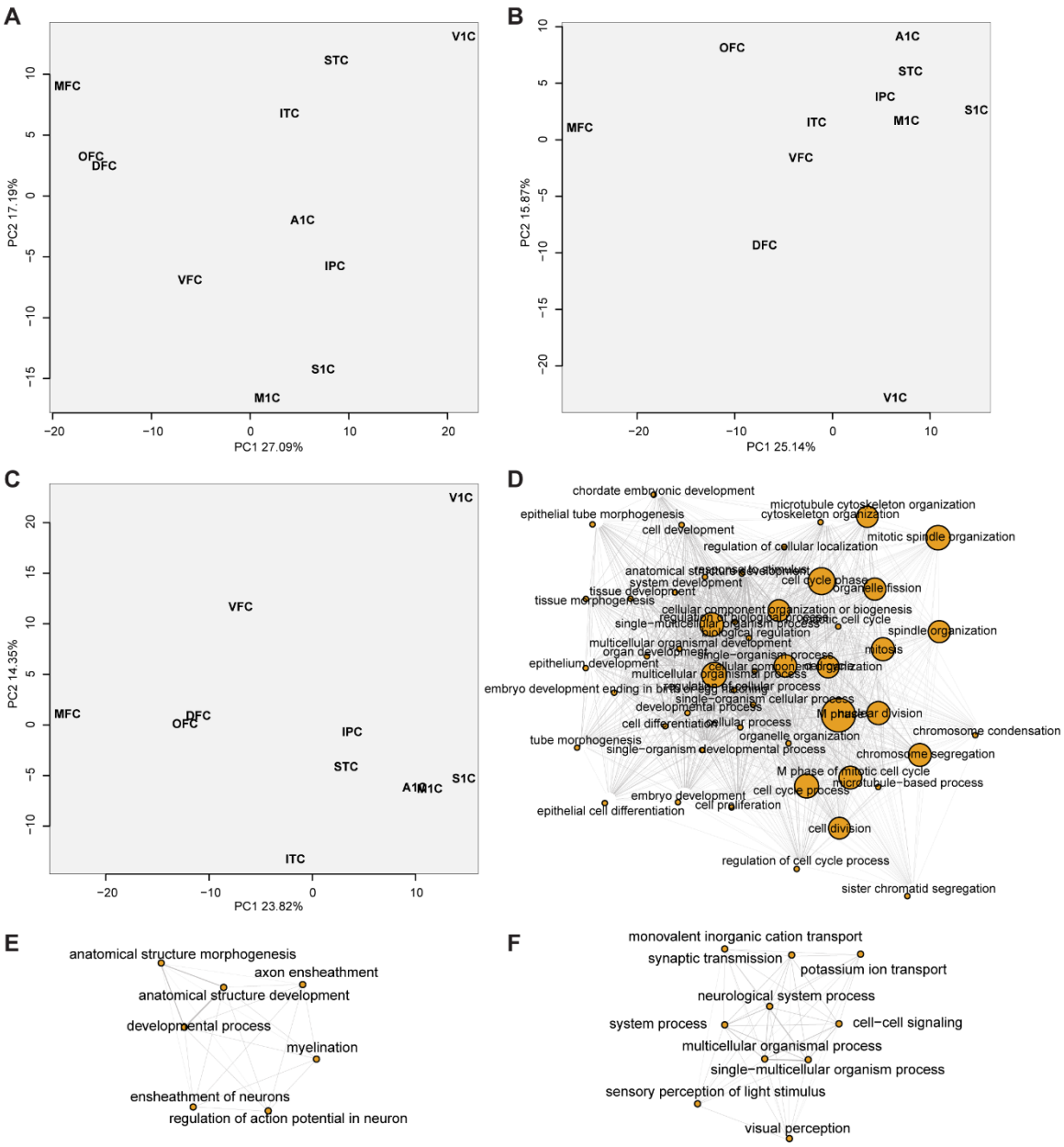


Fig. S30. Principal component analysis of the inter-species divergence among neocortical areas. **(A-C)** Principal component analysis based on inter-species divergence of genes in **(A)** prenatal, **(B)** early postnatal and **(C)** adult. **(D-F)** Networks of GO terms (FDR < 0.05) significantly enriched among the top 100 genes **(D)** negatively correlated with PC1 of prenatal PCA or **(E)** negatively and **(F)** positively correlated with PC1 of early postnatal PCA. Only top 50 GO terms were plotted for visualization in **(D)**. No significant GO terms were identified for the top 100 genes positively correlated with PC1 of prenatal PCA or the top 100 genes correlated

with either direction of PC1 of adult PCA. The size of nodes represents the level of significance (-log10[*P* value]) and the thickness of edges displays the number of genes overlapped between the pair of terms.

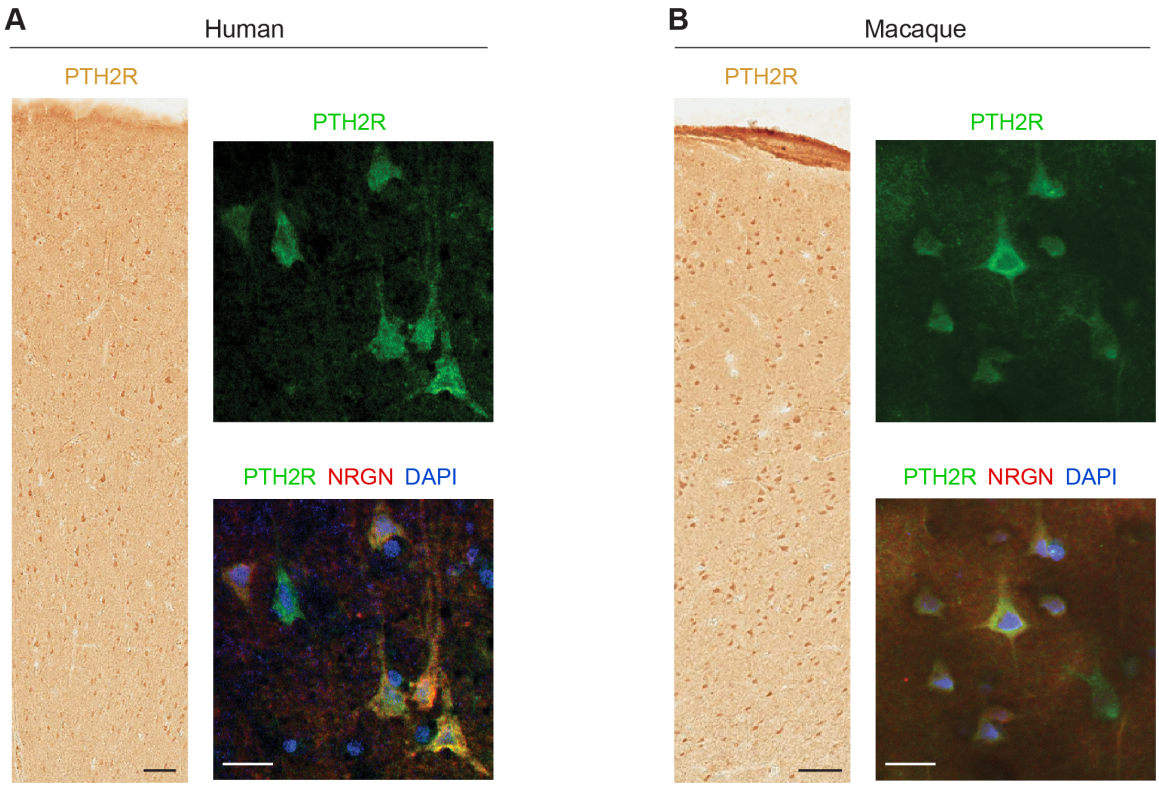


Fig. S31. PTH2R is expressed in both human and macaque excitatory pyramidal neurons. **(A, B)** Immunohistochemical detection of Parathyroid Hormone 2 Receptor (PTH2R) revealed that this protein is expressed in neocortical pyramidal neurons in adult (**A**) human and (**B**) macaque (left panels of (**A**) and (**B**); scale bar corresponds to 100μm). Furthermore, co-localization with Neurogranin (NRGN) corroborated PTH2R localization in pyramidal neurons. Note that some cells that are pyramidal neurons positively labeled by PTH2R in both (**A**) human and (**B**) macaque that do not co-localize with NRGN (right panels of (**A**) and (**B**); scale bar corresponds to 20μm). All the neocortical sections used in immunohistochemistry are from adult human and macaque dorsolateral prefrontal cortex.

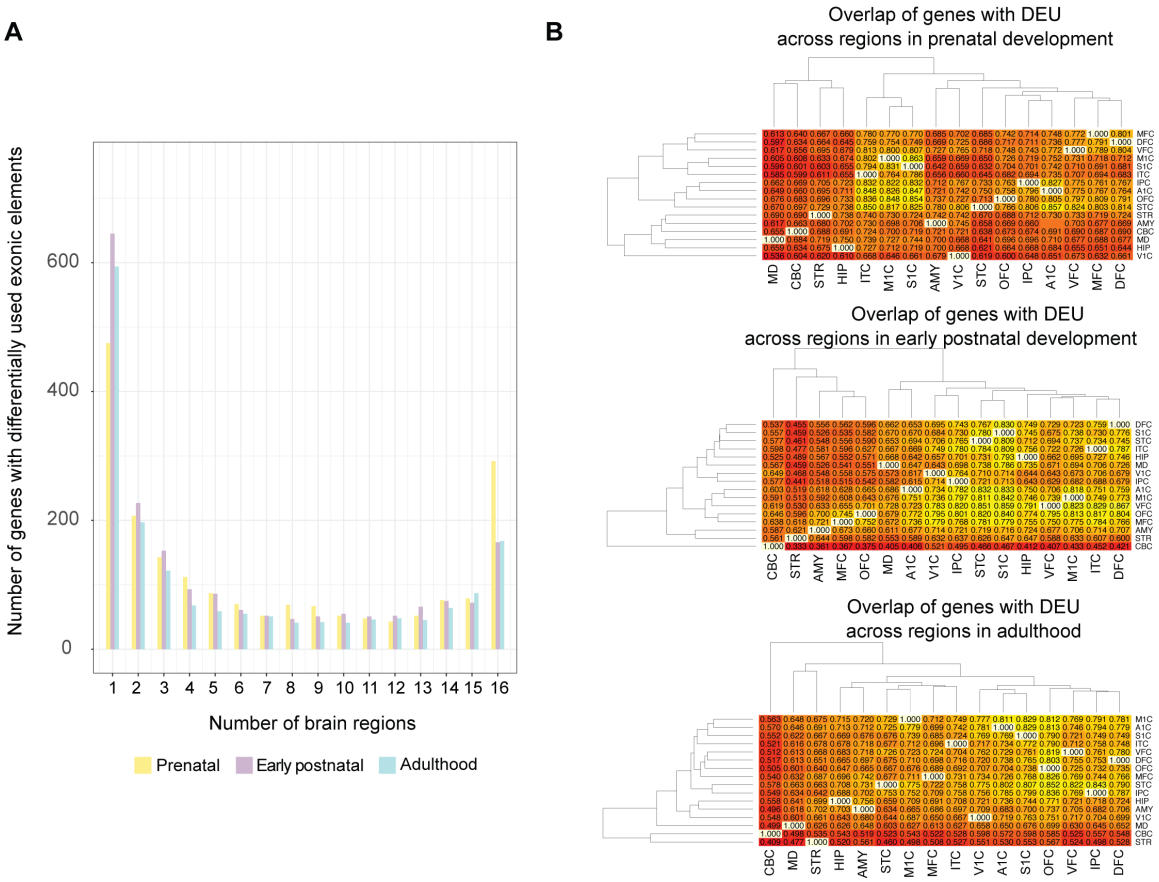


Fig. S32. Differential exon usage between human and macaque. (A) Distribution of human-macaque differential exon usage. Comparisons from different developmental stages are represented in different colors (yellow for prenatal development; purple for early postnatal development; blue for adulthood). **(B)** Heatmap for overlapping of genes with differentially used exonic elements across brain regions within prenatal development (top), early postnatal development (middle) and adult stages (bottom). Cell labels show the overlapping ratio for each pairwise comparison to the number of genes with differentially used exonic elements in the brain region indicated in the row.

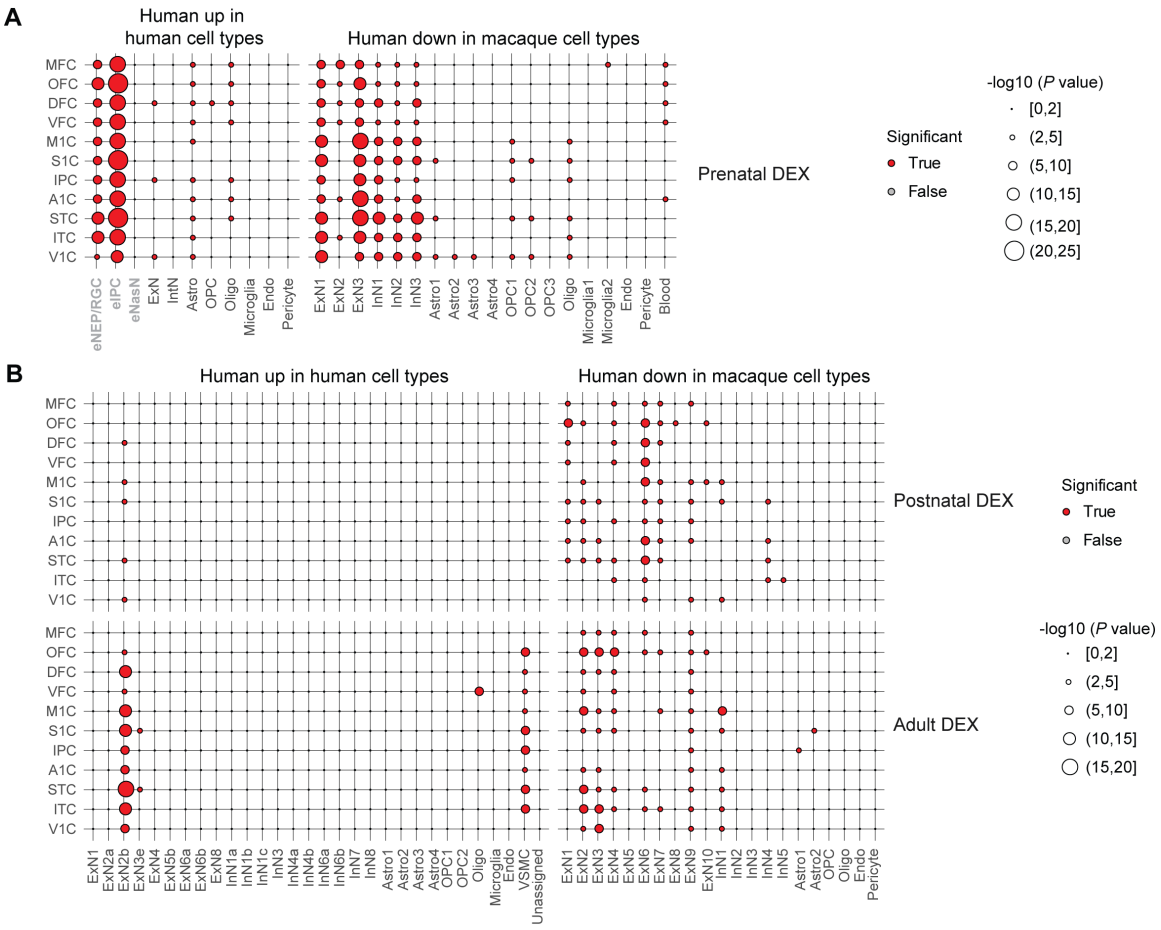


Fig. S33. Cell type enrichment of inter-species differentially expressed genes. (**A**) Cell-type enrichment for genes up- or down-regulated in human neocortical areas. Enrichment of genes up-regulated in human or macaque (down-regulated in human) were tested using single cells from prenatal (embryonic [grey], and fetal) human neocortex (*33*) or fetal macaque DFC, respectively. *P* values adjusted by Benjamini–Hochberg procedure were plotted (size of dots) and significance was labeled by color (True: red; False: grey). (**B**) Cell-type enrichment for genes up- or down-regulated in human neocortical areas. Enrichment of genes up-regulated in human or macaque (down-regulated in human) were tested using single nuclei from adult human DFC (*33*) or macaque DFC, respectively. *P* values adjusted by Benjamini–Hochberg procedure were plotted (size of dots) and significance was labeled by color (True: red and False: grey). (eNPC – embryonic neuroepithelial progenitor; eIPC – embryonic intermediate progenitor cell; eNasN –

embryonic nascent neuron; ExN – excitatory neuron; InN – interneuron; Astro – astrocyte; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell; VSMC – visceral smooth muscle cell.

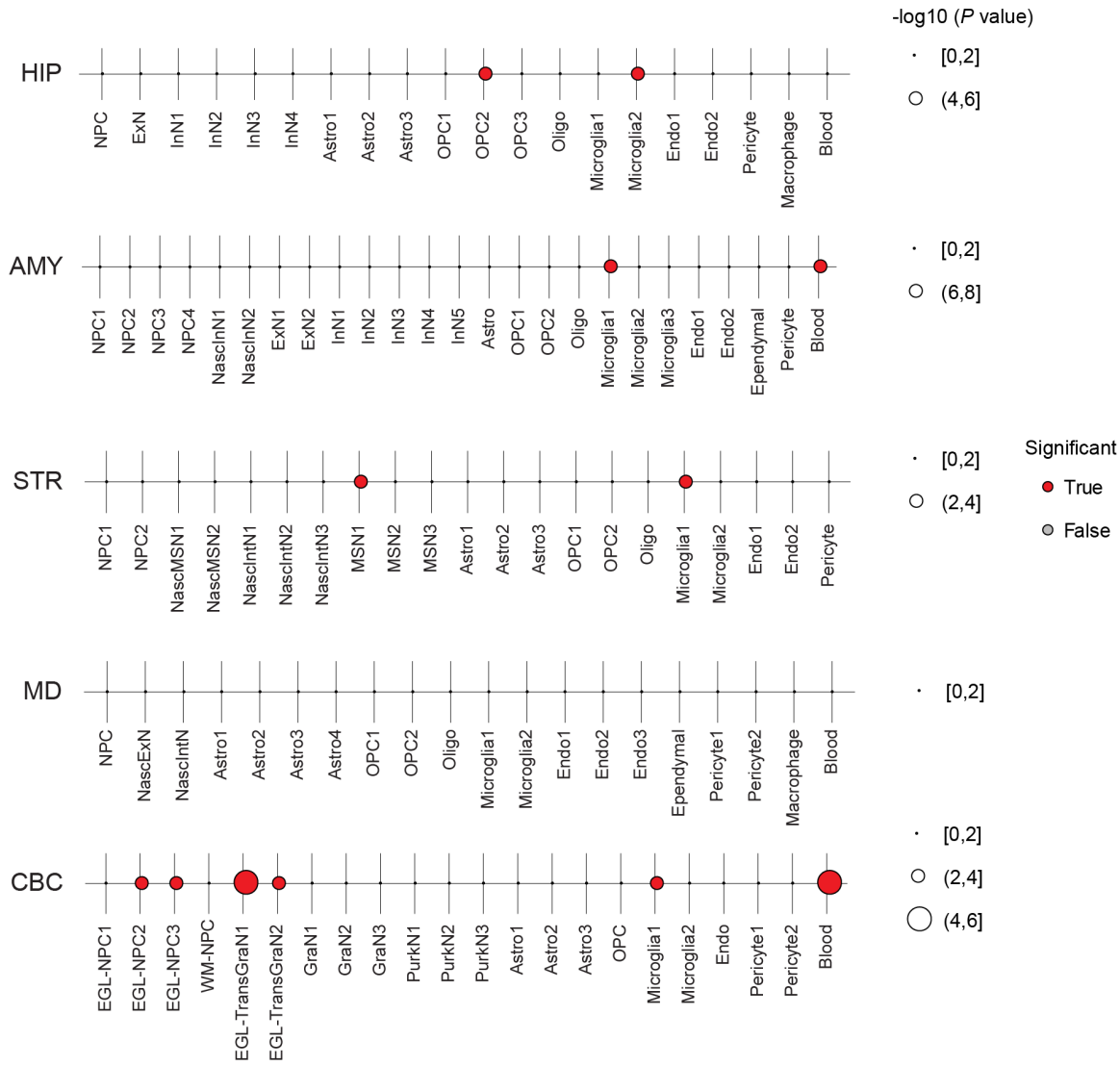


Fig. S34. Cell type enrichment of inter-species differentially expressed genes in prenatal non-neocortical regions. (**A**) Cell-type enrichment for genes exhibiting differential expression between human and macaque in HIP, AMY, STR, MD, and CBC. *P* values adjusted by Benjamini–Hochberg procedure were plotted (size of dots) and significance was labeled by color (True: red; False: grey).

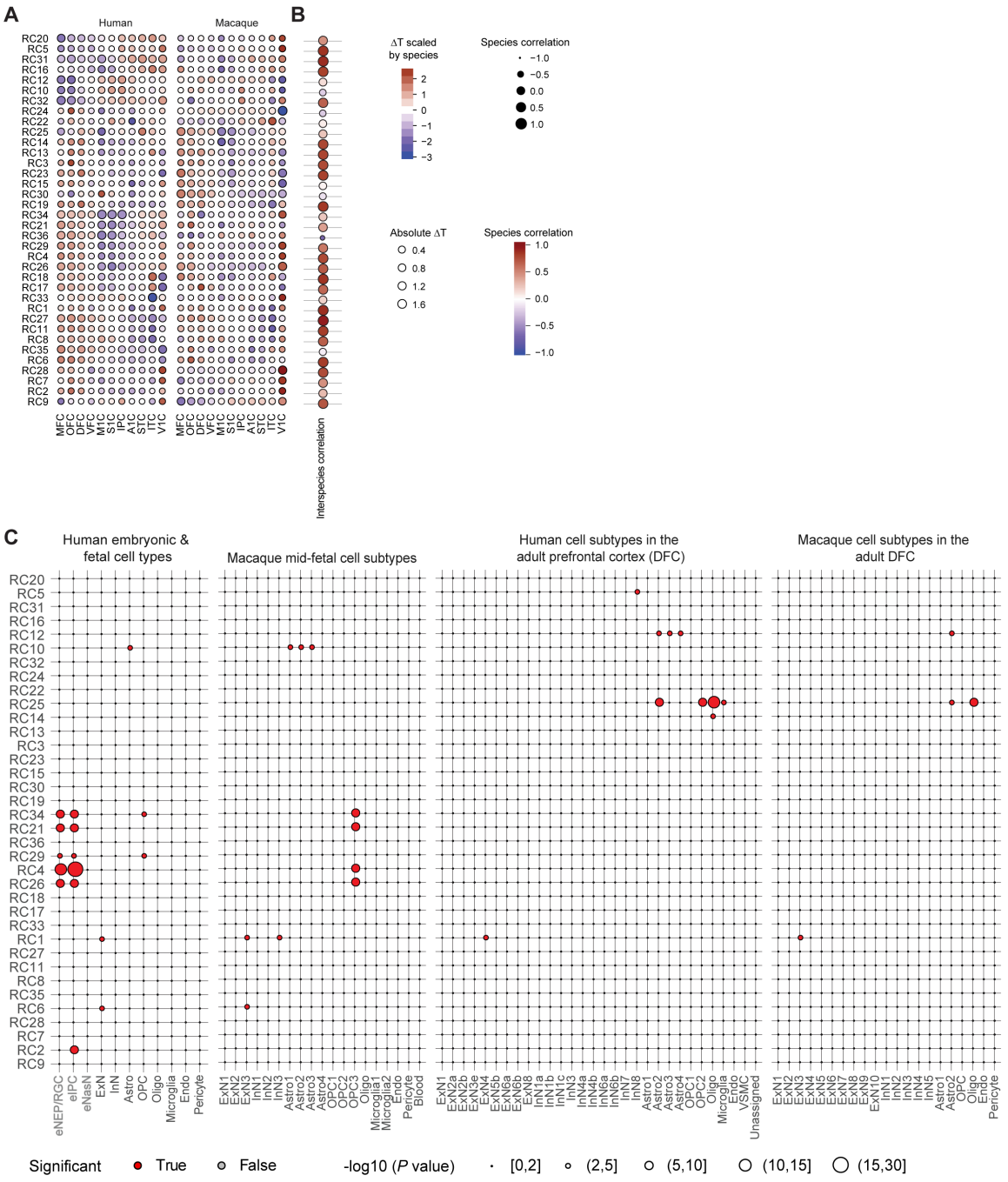


Fig. S35. Inter-species divergence and cell-type enrichment of clusters of genes with heterochronic regional expression. (A) Bubble matrix showing heterochronic expression regional clusters (RC) comprised of genes clustered by their areal heterochronic expression patterns in both species. **(B)** Inter-species correlation of regional heterochrony. **(C)** Cell type enrichment of regional clusters. Significance is based on the Kolmogorov–Smirnov test ($P <$

0.01). ΔT: temporal difference in gene expression among regions. eNEP/RGC – embryonic neural epithelial progenitor/radial glial cell; eIPC – embryonic intermediate progenitor cell; eNasN – embryonic nascent neuron; ExN – excitatory neuron; InN – interneuron; Astro – astroglial lineage; OPC – oligodendrocyte progenitor cell; Oligo – oligodendroglial lineage; Endo – endothelial cell; VSMC – vascular smooth muscle cell.

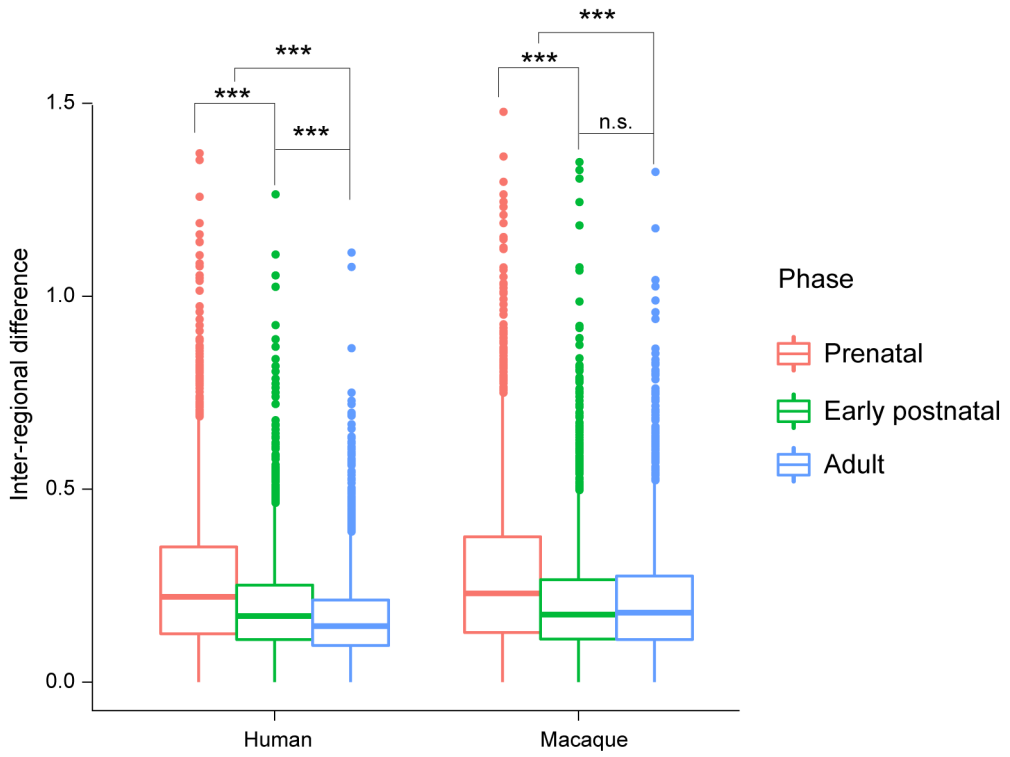


Fig. S36. Extent of inter-areal differences among inter-regional heterochronic genes across time. Boxplot displays the distribution of expression differences among genes showing inter-areal temporal differences. The expression of all inter-areal heterochronic genes were predicted at 100 time points as described in "Analysis of inter-regional differences" in (*32*). For each gene, the pairwise Manhattan distance between neocortical areas was calculated by phase and normalized by the number of time points in the phase. The inter-areal difference for a gene was defined as the average of the pairwise distance. *** - $P < 0.001$ in pairwise paired *t*-test, with multiple comparisons adjusted by the Bonferroni correction. n.s. – not significant.

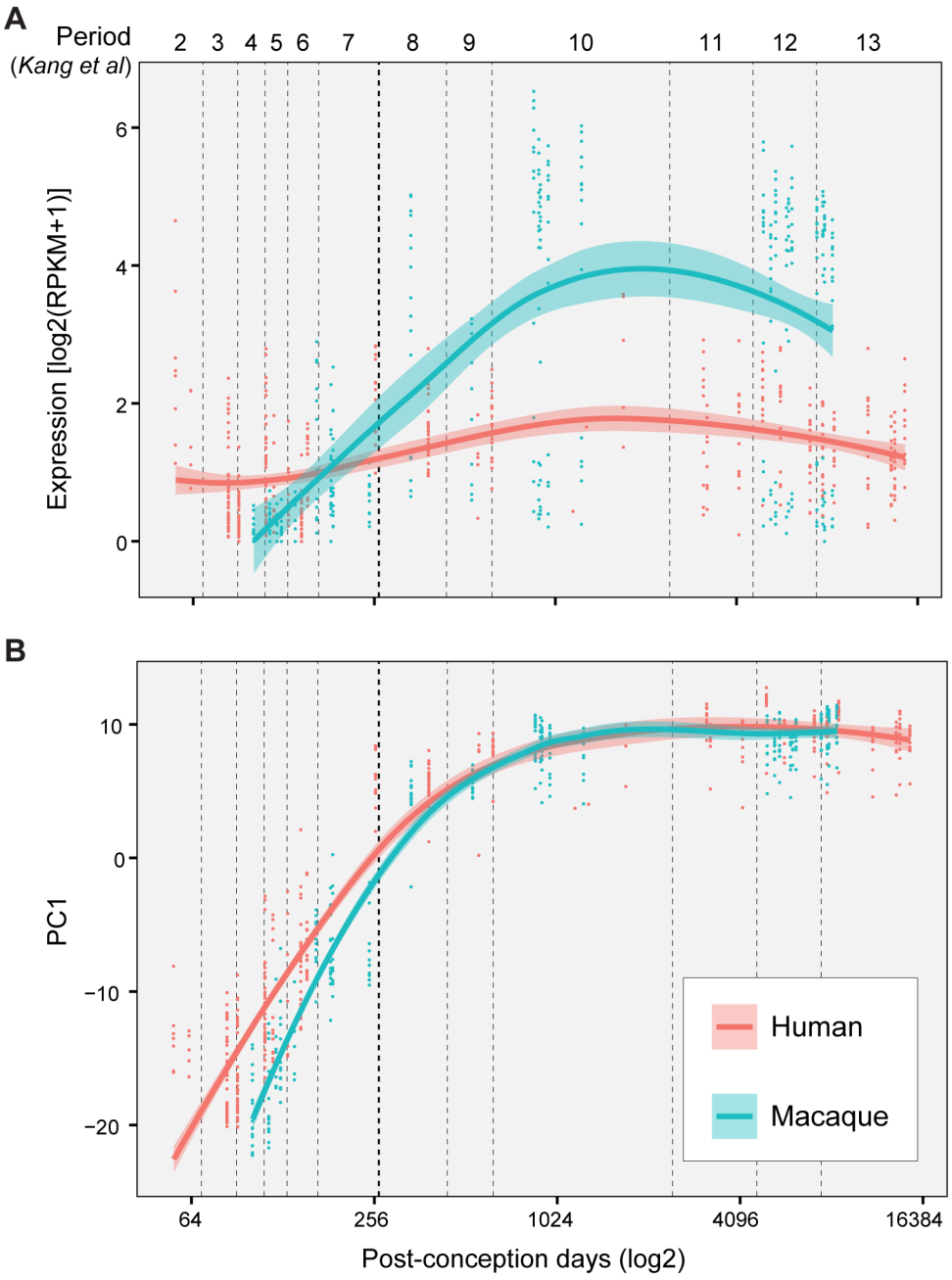


Fig. S37. Expression trajectories of *TWIST1* and heterochronic genes in the transcriptional regulatory network. (A) The expression trajectory of *TWIST1* in human (red) and macaque (green). **(B)** The expression trajectories of heterochronic genes in the transcriptional regulatory network in Fig. 6B, represented by PC1. The curves were fitted by LOESS and the shades represent the 95% confidence interval.

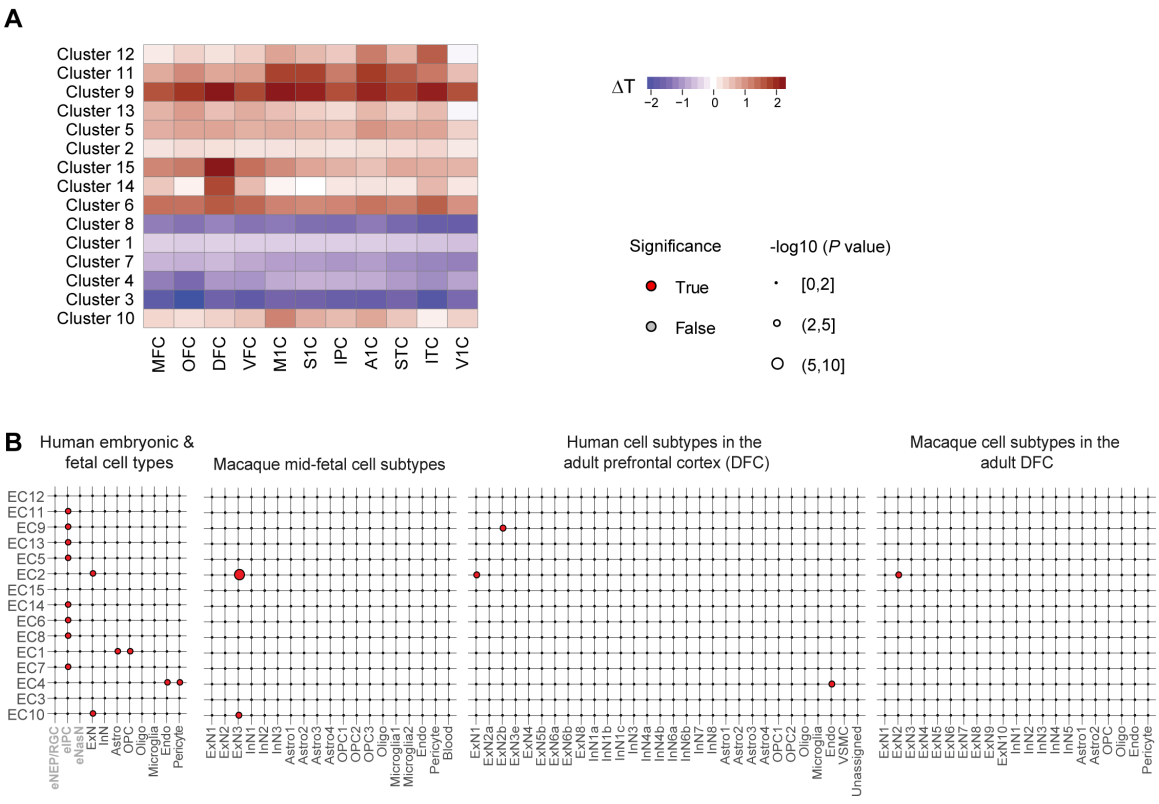


Fig. S38. Inter-species temporal differences and cell-type enrichment of clusters of genes with heterochronic expression between species. (A) Heatmap showing the heterochrony of genes clustered by their inter-species heterochronic expression patterns in each brain region. Blue indicates clusters composed by genes expressed earlier in a given region in human than in macaque; red indicates later expression in human than in macaque. **(B)** Cell-type enrichment of the clusters. Significance is based on the Kolmogorov–Smirnov test ($P < 0.01$). eNPC/RGC – embryonic neural progenitor cell/radial glial cell; eIPC – embryonic intermediate progenitor cell; eNasN – embryonic nascent excitatory neuron; ExN – excitatory neuron; InN – interneuron; Astro – astroglial lineage; OPC – oligodendrocyte progenitor cell; Oligo – oligodendroglial lineage; Endo – endothelial cell; VSMC – vascular smooth muscle cell.

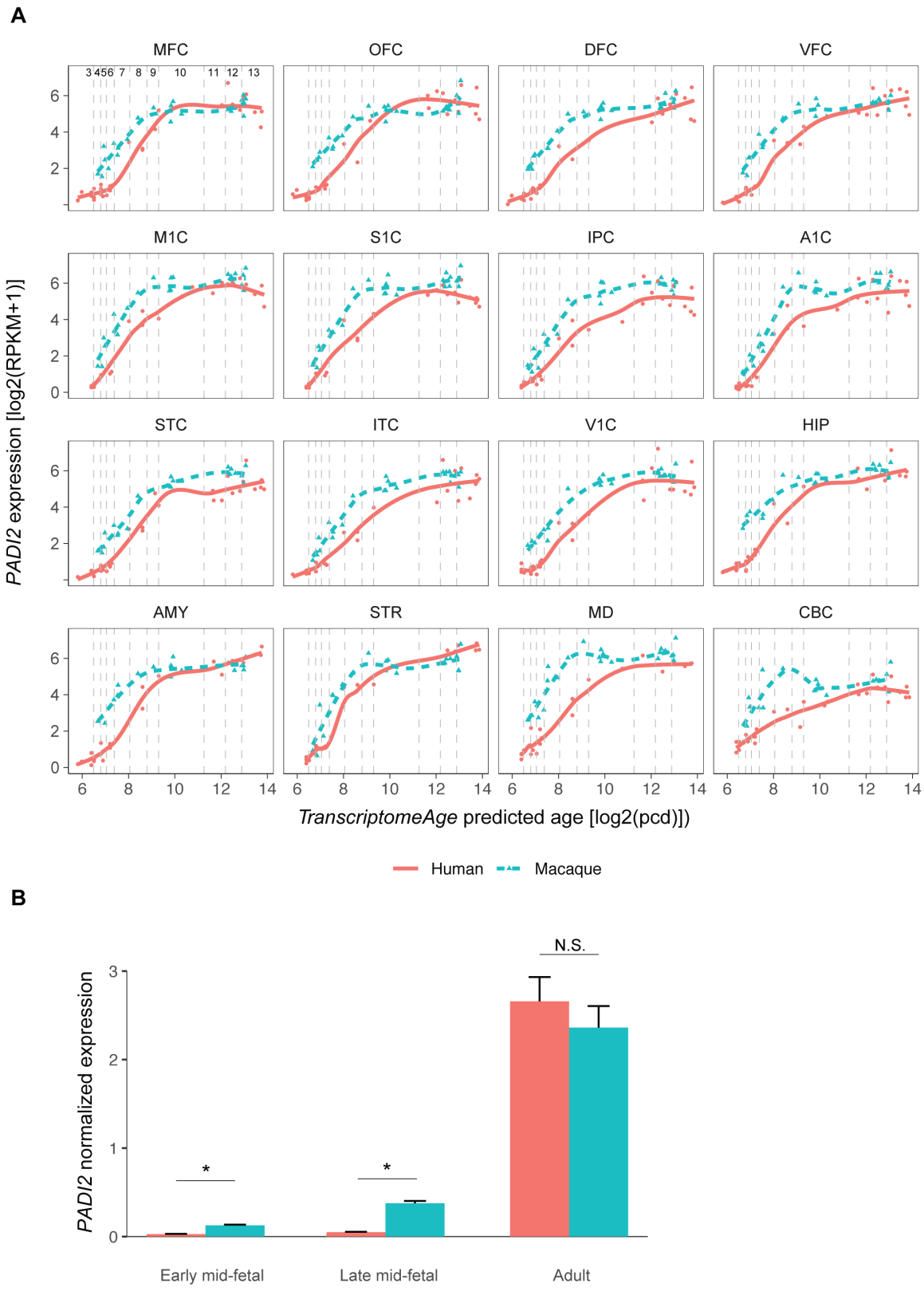


Fig. S39. Validation of *PADI2* expression profile. (A) Expression profile of *PADI2* throughout development in the profiles 16 brain regions shows that this gene has a delayed expression in

human, when compared to macaque. This gene is part of the evolutionary cluster 15 and regional cluster 2. **(B)** ddPCR validation of *PADI2* expression corroborates the expression profile assessed with RNA-seq. Expression was normalized to the expression of the housekeeping gene *TBP*. ddPCR was performed in ventrolateral prefrontal cortex in each developmental period in human and macaque (2 biological replicates; 2 technical replicates). * t-test $P < 0.05$; N.S. – non-significant.

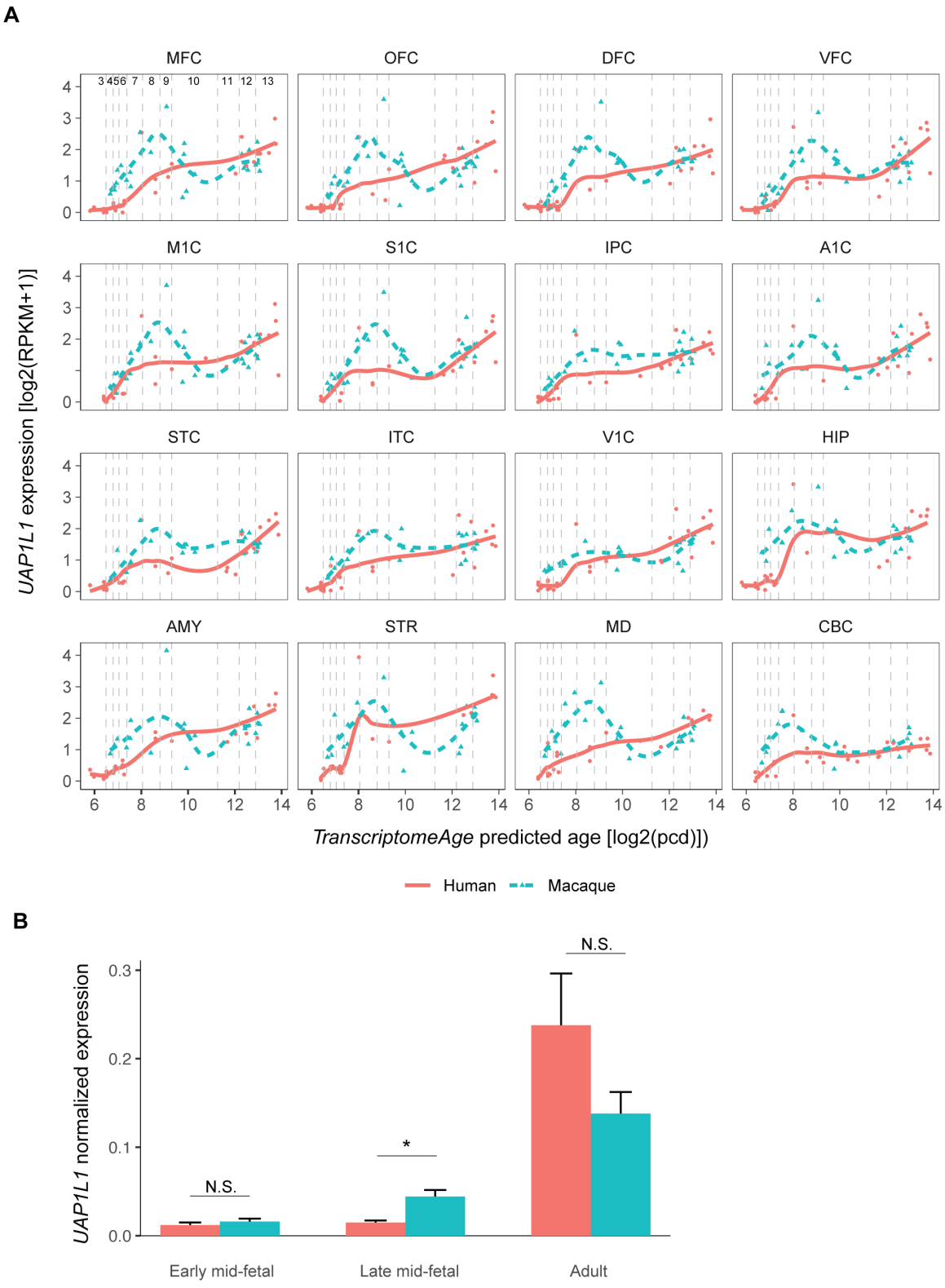


Fig. S40. Validation of *UAP1L1* expression profile. (A) Expression profile of *UAP1L1* throughout development in the profiles 16 brain regions shows that this gene is up-regulated in

macaque from midfetal development to early postnatal periods. This gene in part of the evolutionary cluster 9 and the regional cluster 3. **(B)** ddPCR validation of *UAP1L1* expression corroborates the expression profile assessed with RNA-seq. Expression was normalized to the expression of the housekeeping gene *TBP*. ddPCR was performed in ventrolateral prefrontal cortex in each developmental period in human and macaque (2 biological replicates; 2 technical replicates). * t-test $P < 0.05$. N.S. – non-significant.

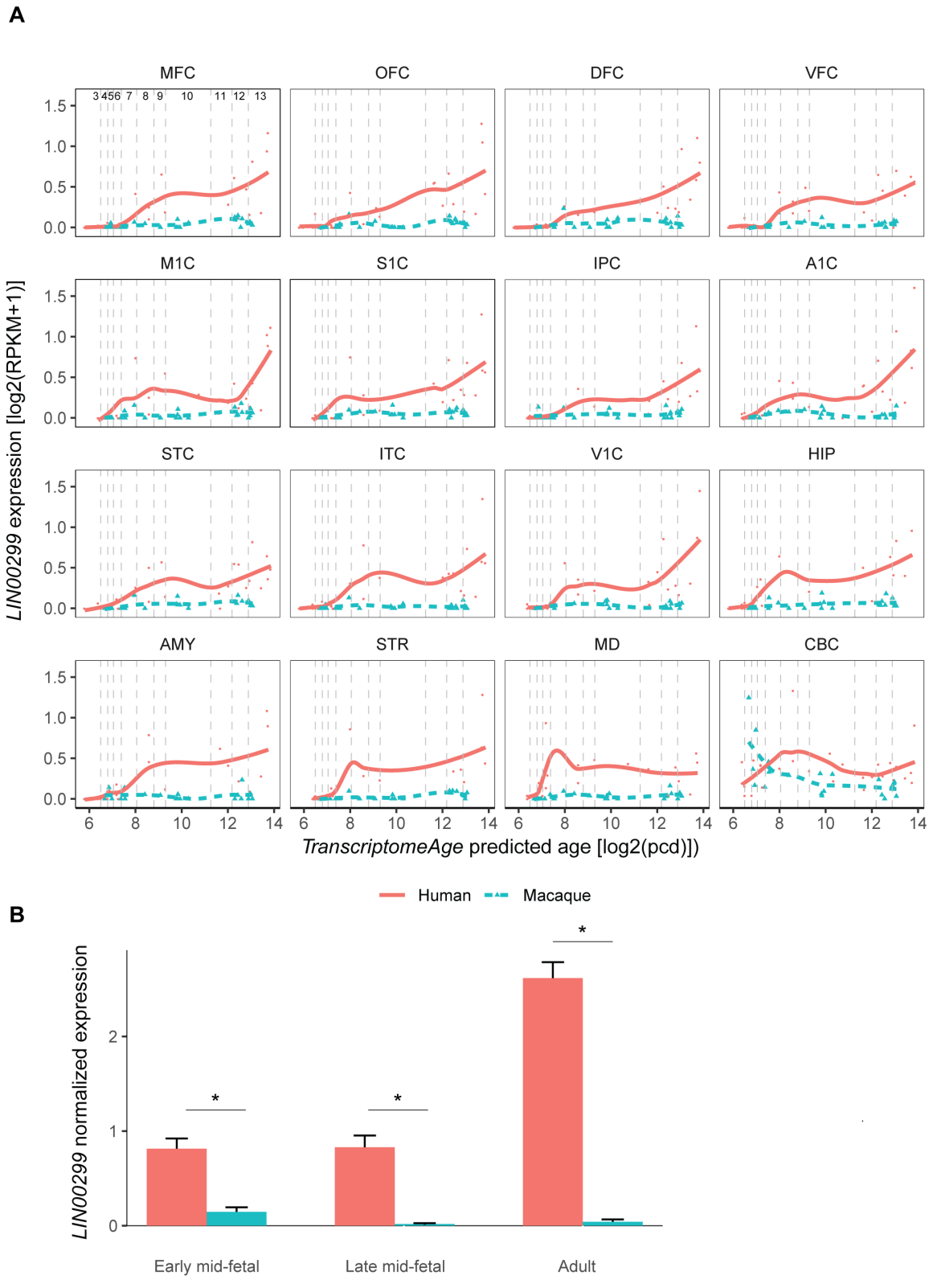


Fig. S41. Validation of *LINC00299* expression profile. (A) Expression profile of *LINC00299* throughout development in the profiles 16 brain regions shows that this gene up-regulated in

human, when compared to macaque, from midfetal development through adulthood. **(B)** ddPCR validation of *LINC00299* expression corroborates the expression profile assessed with RNA-seq. Expression was normalized to the expression of the housekeeping gene *TBP*. ddPCR was performed in ventrolateral prefrontal cortex in each developmental period in human and macaque (2 biological replicates; 2 technical replicates). * t-test $P < 0.05$. N.S. – non-significant.

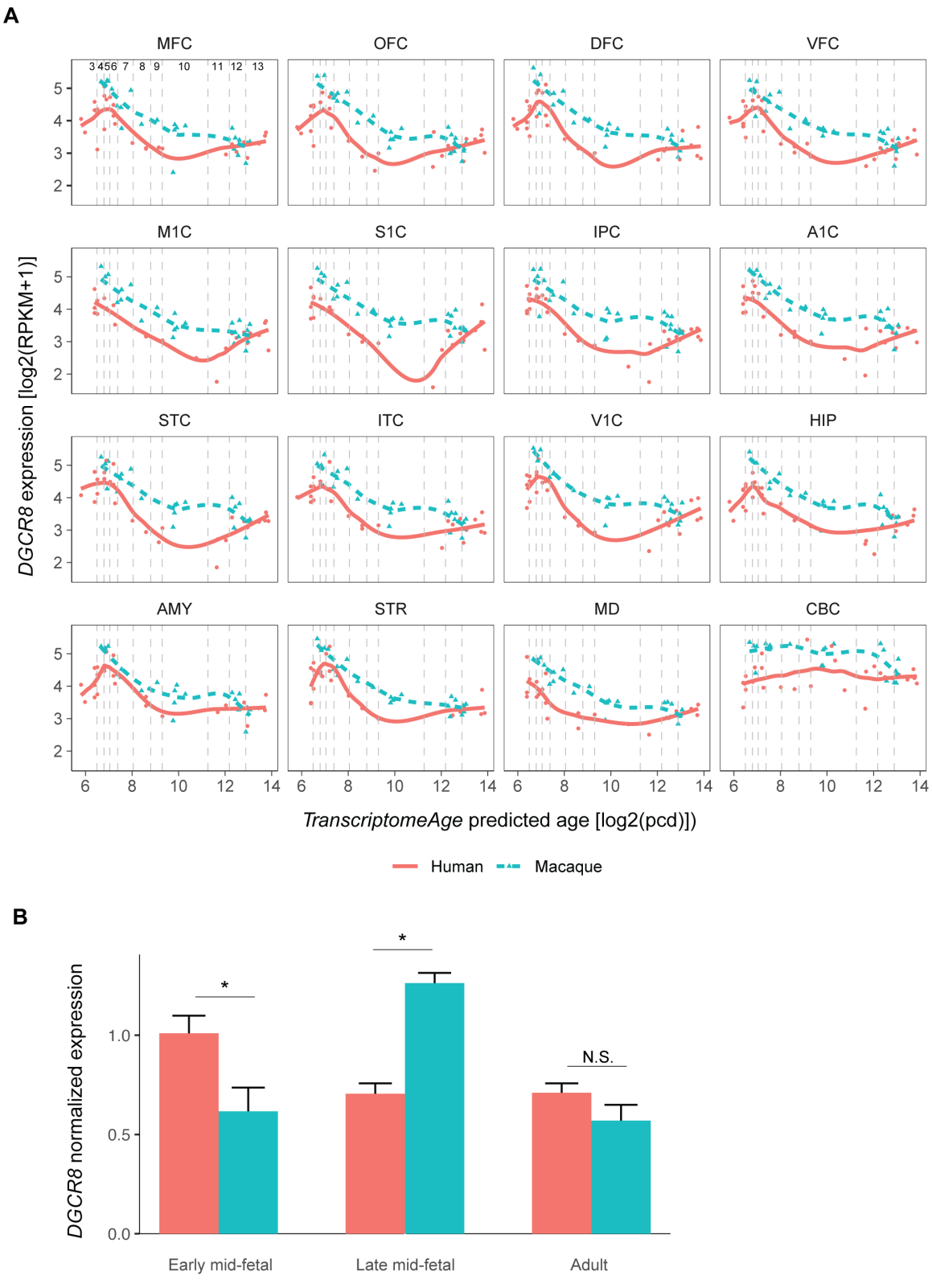


Fig. S42. Validation of *DGCR8* expression profile. **(A)** Expression profile of *DGCR8* throughout development in the profiles 16 brain regions shows that this gene is up-regulated in

macaque, especially from midfetal development to adulthood. This gene is part of the evolutionary cluster 3. **(B)** ddPCR validation of *DGCR8* expression corroborates the expression profile assessed with RNA-seq. Expression were normalized to the expression of the housekeeping gene *TBP*. ddPCR was performed in ventrolateral prefrontal cortex in each developmental period in human and macaque (2 biological replicates; 2 technical replicates). * t-test $P < 0.05$. N.S. – non-significant.

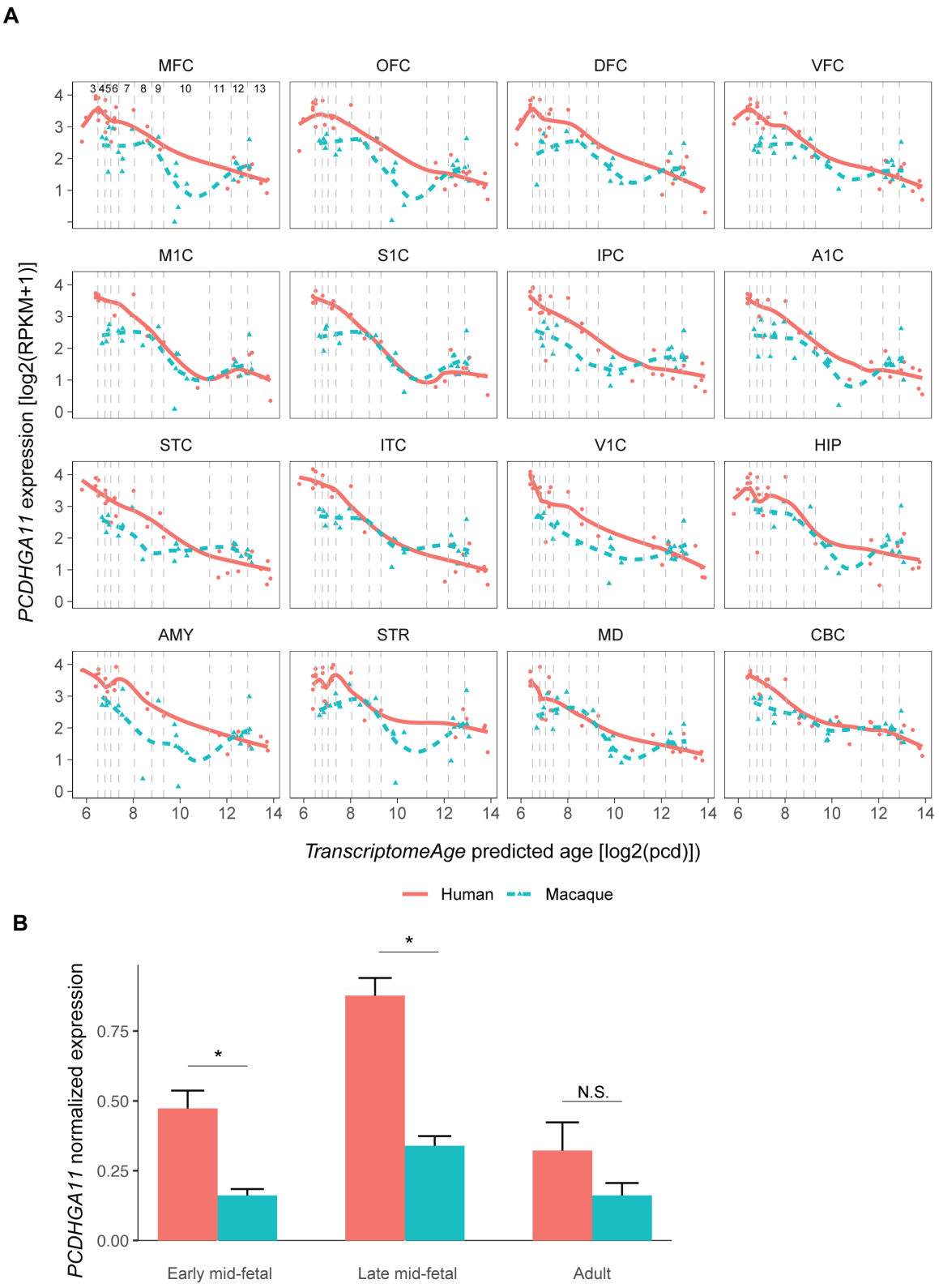


Fig. S43. Validation of *PCDHGA11* expression profile. (A) Expression profile of *PCDHGA11* throughout development in the profiles 16 brain regions shows that this gene is more expressed in

human in prenatal and early postnatal development. This gene is part of the evolutionary cluster 9. **(B)** ddPCR validation of *PCDHGA11* expression corroborates the expression profile assessed with RNA-seq. Expression was normalized to the expression of the housekeeping gene *TBP*. ddPCR was performed in ventrolateral prefrontal cortex in each developmental period in human and macaque (2 biological replicates; 2 technical replicates). * t-test $P < 0.05$. N.S. – non-significant.

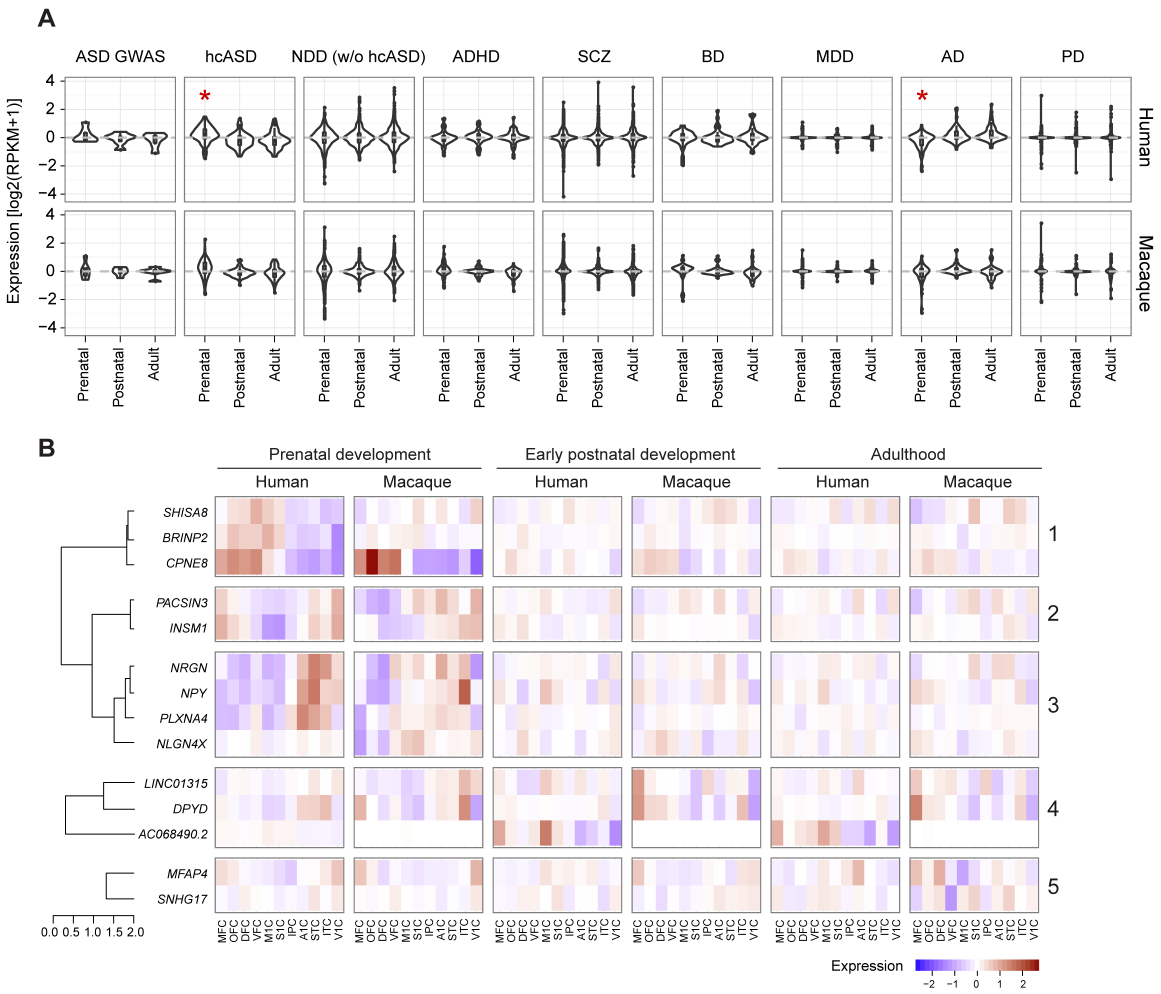


Fig. S44. Expression pattern of genes associated with neuropsychiatric disorders. (A) Violin plot displaying the expression of genes associated with autism spectrum disorder (ASD; high confidence [hc]), neurodevelopmental disorders (NDD), attention deficit hyperactivity disorder (ADHD), schizophrenia (SCZ), bipolar disorder (BD), major depressive disorder (MDD), Alzheimer's disease (AD), and Parkinson's disease (PD), in the three developmental phases. Each gene was scaled by mean and standard deviation across samples and averaged by phase. Asterisk indicates significant difference (Tukey's test $P < 0.05$) between the labeled developmental phase compared with both other phases. **(B)** Heatmap of SCZ-associated genes that exhibit heterotopic expression between human and macaque. Genes were clustered by their inter-species regional expression patterns: heterotopic expression in prenatal development (Cluster 1-3), postnatal development and adult (Cluster 4), and adult only (Cluster 5). Genes showing heterotopic

expression in prenatal development were further clustered based on their expression pattern in human prenatal development (Cluster 1-3). The expression was scaled by mean across all neocortical areas in each species. The hierarchical clustering for Cluster 4 and 5 was conducted based on expression in early postnatal and adult human, respectively.

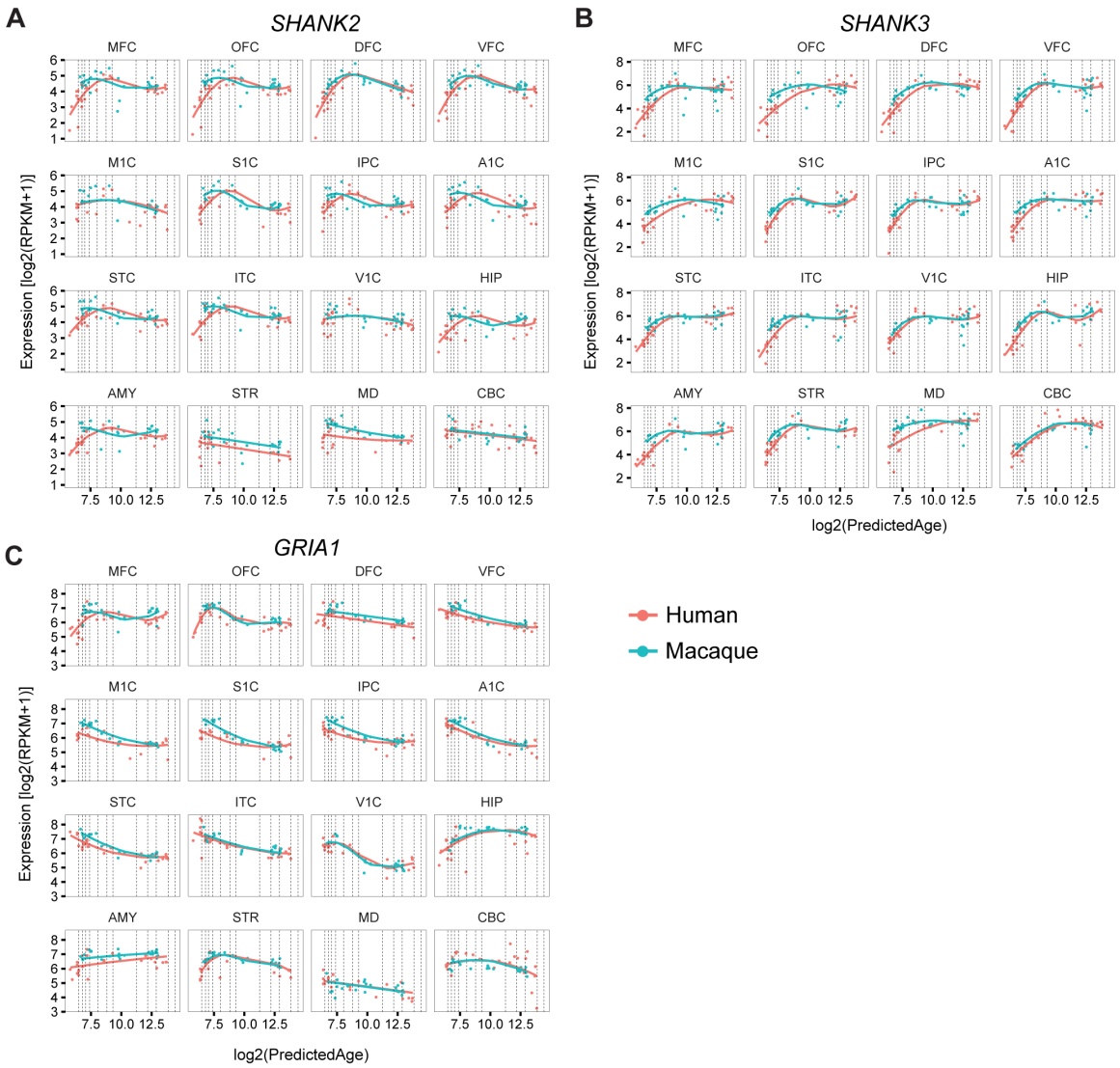
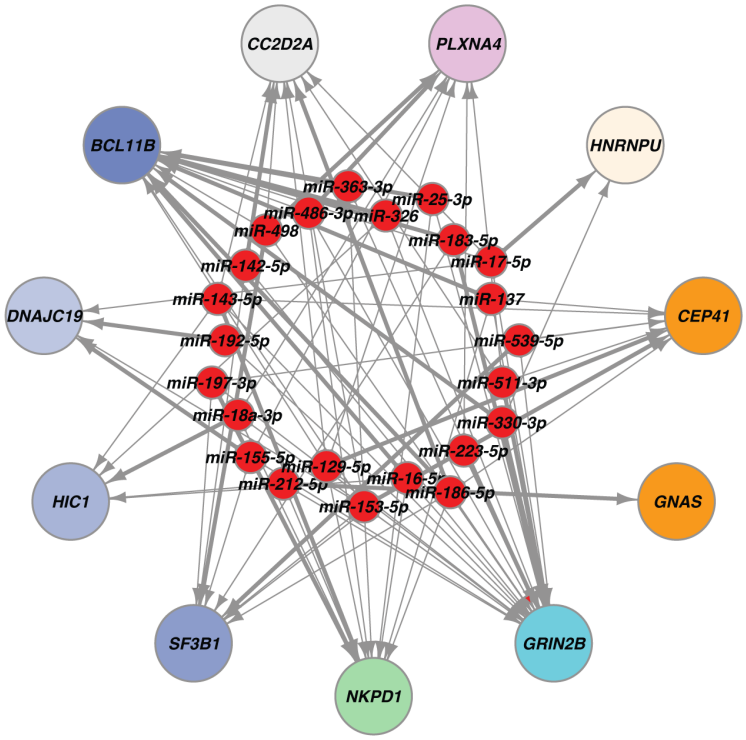


Fig. S45. Regional heterochronic expression between species of genes associated with neuropsychiatric disorders. (A-C) Plots depicting the expression of genes associated with **(A-B)** autism spectrum disorder and **(C)** schizophrenia that exhibit heterochronic expression between human and macaque. The curves were fitted by shift models of *TempShift*.

A


CC2D2A
PLXNA4
BCL11B
HNRNPU
miR-363-3p
miR-25-3p
miR-486-3p
miR-326
miR-498
miR-183-5p
miR-142-5p
miR-17-5p
miR-143-5p
miR-137
DNAJC19
CEP41
miR-192-5p
miR-539-5p
miR-197-3p
miR-511-3p
miR-18a-3p
miR-330-3p
miR-223-5p
miR-155-5p
miR-129-5p
miR-16-5p
miR-212-5p
miR-186-5p
HIC1
GNAS
miR-153-5p
SF3B1
GRIN2B
NKPD1

B

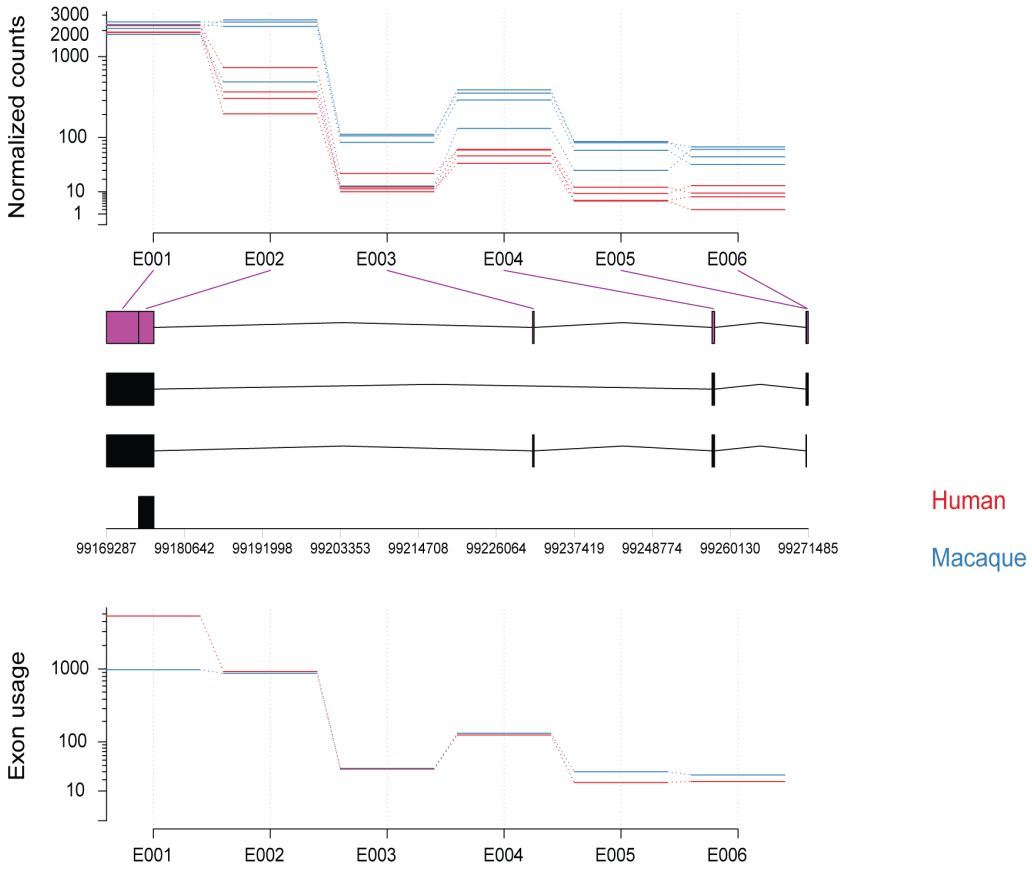

Normalized counts
3000
2000
1000
100
10
1
E001
E002
E003
E004
E005
E006
99169287
99180642
99191998
99203353
99214708
99226064
99237419
99248774
99260130
99271485
Human
Macaque
Exon usage
1000
100
10

Fig. S46. Differential exon usage in genes associated with neuropsychiatric disorders.

(A) Disease-associated miRNA:gene interactions based on the binding sites detected in the differentially used exonic elements and experimental evidence for gene-miRNA interaction. Genes and miRNAs with at least one validated interaction are shown. Validated miRNA:gene interactions are represented by thick edges. Small red nodes represent miRNAs while bigger colored nodes represent target genes. Genes associated with different neuropsychiatric disorders are coded by different color nodes: dark blue for schizophrenia, orange for neurodevelopmental disorders, green for Alzheimer's disease, grey for Parkinson's disease, light blue for autism spectrum disorder, and magenta for genes associated with several diseases. Node darkness reflects the number of miRNAs interacting with the corresponding gene. **(B)** *BCL11B* normalized counts (top) and exon usage pattern (bottom) or human (red) and macaque (blue) in striatum during adulthood stage. Exonic elements and the correspondence with orthologous transcript models are also represented (middle).

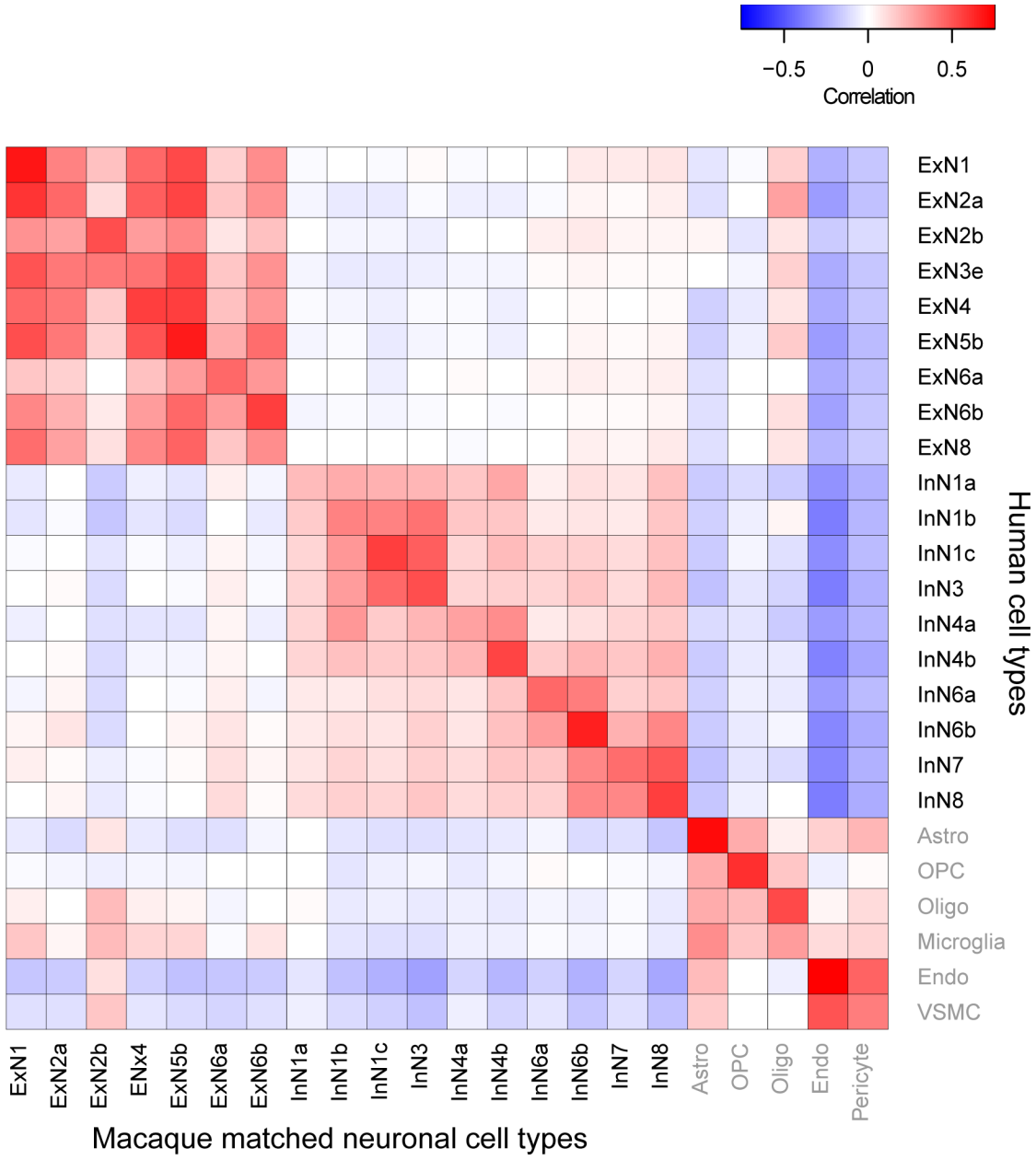


Fig. S47. Correlation of matched subtypes between species. For single-cell level inter-species differential expression analysis, we reclassified the adult macaque excitatory neurons and interneurons to match the adult human subtypes. Non-neuronal cells are labeled in grey and were not reclassified. Top 100 genes of each cell types/subtypes were selected per cell based on Preferential Enrichment Measure (PEM) and combined. The correlation matrix was plotted based on PEM of the selected genes.

Supplementary Tables:

Supplementary Tables are provided in a single Microsoft Excel file.

Table S1. Sample metadata

This table summarizes specimen information, including species, sex, brain region/neocortical area, age, developmental period, predicted age based on *TranscriptomeAge*, RIN, and number of total and uniquely mapped reads. This table also includes the batch information, such as RNA extraction method and sequencing platform.

Table S2. Specimen/sample metadata and quality control metrics for single-cell and single-nucleus RNA-seq

This table summarizes specimen and sample information of single-cell and single-nucleus RNA-seq, including species, sex, brain region/neocortical area, age, sex, and quality control metrics.

Table 3. List of identified cell types and number of cells in human and macaque single-cell and single-nucleus RNA-seq

List of cell types and the number of cells identified in each brain region of fetal macaque single-cell RNA-seq and adult human and macaque DFC single-nuclei RNA-seq.

Table S4. Fetal macaque single-cell subtypes and genes ranked by specificity score

Lists of fetal macaque single-cell subtypes and genes ranked by specificity score in corresponding cell types.

Table S5. Adult human single-nucleus subtypes and genes ranked by specificity score

Lists of adult human single-nuclei subtypes and genes ranked by specificity score in corresponding cell types.

Table S6. Adult macaque single-nucleus subtypes and genes ranked by specificity score

Lists of adult macaque single-nuclei subtypes and genes ranked by specificity score in corresponding cell types.

Table S7. Intra-species regional WGCNA modules in each developmental phase

List of all genes used in the study and all the WGCNA modules identified in each developmental phase in each species. The intramodular connectivity (modKME) of each gene is shown for the respective WGCNA module.

Table S8. Prenatal WGCNA modules

List of prenatal modules in human and macaque. Statistical significance for regional differences and inter-species correlation are shown.

Table S9. Early postnatal WGCNA modules

List of early postnatal modules in human and macaque. Statistical significance for regional differences and inter-species correlation are shown.

Table S10. Adult WGCNA modules

List of adult modules in human and macaque. Statistical significance for regional differences and inter-species correlation are shown.

Table S11. Genes differentially expressed between species during prenatal development, early postnatal development, and adulthood

List of genes differentially expressed in at least one region of the human brain. For each gene in each region, we listed whether a gene is significantly differentially expressed between human and macaque (TRUE, FDR < 0.01; fold change > 1.5) in each of the developmental phases.

Table S12. Functional enrichment of inter-species differentially expressed genes

List of functional enriched terms of genes with differential expression between human and macaque in each brain region and each developmental phase.

Table S13. Genes with inter-species differential exon usage

List of differential exon usage events between species.

Table S14. Inter-species differentially expressed (DEX) transcription factors with binding sites enriched in promoters of inter-species DEX genes

The number of transcription factors and corresponding binding sites enriched in promoter of inter-species differentially expressed genes.

Table S15. Interactions between inter-species differentially expressed (DEX) genes and DEX-transcription factors enriched in their promoters

Table of interactions between inter-species differentially expressed (DEX) genes and DEX-transcription factors enriched in their promoters

Table S16. Inter-species differentially expressed (DEX) transcription factors with binding sites enriched in other elements of inter-species DEX genes

Transcription factors and corresponding binding sites enriched in regulatory elements of inter-species differentially expressed genes.

Table S17. Interactions between inter-species differentially expressed (DEX) genes and DEX-transcription factors enriched in differentially activated regulatory elements down-regulated in human

Table of interactions between inter-species differentially expressed (DEX) genes and DEX-transcription factors enriched in regulatory elements down-regulated in human.

Table S18. Interactions between inter-species differentially expressed (DEX) genes and DEX-transcription factors enriched in differentially activated regulatory elements up-regulated in human

Table of interactions between inter-species differentially expressed (DEX) genes and DEX-transcription factors enriched in regulatory elements up-regulated in human

Table S19. Regional heterochronic clusters

List of all the intra-species heterochronic regional clusters (RC). The gene composition, functional annotation enrichment, and the enrichment in WGCNA modules are shown for each regional cluster.

Table S20. Evolutionary heterochronic clusters

List of all the inter-species heterochronic evolutionary clusters (EC). The gene composition, and functional annotation enrichment are shown for each regional cluster.

Table S21. Genes associated with neurological and psychiatric disorders

List of genes associated with neuropsychiatric disorders in our annotation. For each gene, the following information is listed: disease that the gene is associated with, difference in regional pattern between species (distance of centered expression between species > 3), WGCNA modules that the gene is in, inter-species differential expression, and regional and evolutionary heterochronic clusters that the gene is in.

Table S22. Genes associated with diseases with putative miRNA binding to differentially used exonic elements

List of genes associated with diseases and the information of putative miRNA binding to differentially used exonic elements of these genes.

Table S23. Experimental evidence of interaction between miRNAs and DEU genes associated with diseases

Experimental evidence showing interaction between miRNAs and DEU genes associated with diseases and the linkage between the gene and the miRNA to these traits.

Table S24. Comparison between adult macaque subtypes with macaque subtypes matched to human

The number of cells overlapped between adult macaque neuronal subtypes and matched neuronal subtypes to human.

Table S25. Nomenclature of cell types used in cell-type deconvolution

Nomenclature of the cell types from Zhong et al. (*35*) and human adult single-nuclei RNA-seq used in cell-type deconvolution. NPC – neural progenitor cell; ExN – excitatory neuron; InN – interneuron; Astro – astrocytes; OPC – oligodendrocyte progenitor cell; Oligo – oligodendrocyte; Endo – endothelial cell.

Table S26. Age of samples included in each WGCNA network

Table with the ages of samples used to build the WGCNA networks in human or macaque prenatal development, early postnatal development, and adulthood.

Table S27. Samples used for ddPCR validation

List of human and macaque brains used for ddPCR validation of genes with heterochronic expression.