Published in final edited form as:

Nature. 2016 July 21; 535(7612): 367–375. doi:10.1038/nature18637.

Comprehensive transcriptional map of primate brain development

A full list of authors and affiliations appears at the end of the article.

Abstract

The transcriptional underpinnings of brain development remain poorly understood, particularly in humans and closely related non-human primates. We describe a high resolution transcriptional atlas of rhesus monkey brain development that combines dense temporal sampling of prenatal and postnatal periods with fine anatomical parcellation of cortical and subcortical regions associated with human neuropsychiatric disease. Gene expression changes more rapidly before birth, both in progenitor cells and maturing neurons, and cortical layers and areas acquire adult-like molecular profiles surprisingly late postnatally. Disparate cell populations exhibit distinct developmental timing but also unexpected synchrony of processes underlying neural circuit construction including cell projection and adhesion. Candidate risk genes for neurodevelopmental disorders including primary microcephaly, autism spectrum disorder, intellectual disability, and schizophrenia show disease-specific spatiotemporal enrichment within developing neocortex. Human developmental expression trajectories are more similar to monkey than rodent, and approximately 9% of genes show human-specific regulation with evidence for prolonged maturation or neoteny.

Corresponding author: Ed S. Lein, Ph.D., Allen Institute for Brain Science, 615 Westlake Ave N, Seattle, WA 98109, Tel: 206.548.7039, Fax: 206.548.7071, EdL@alleninstitute.org.

*These authors contributed equally

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions

T.E.B., J.A.M., S.-L.D., L.N., A.S., R.A.D., J.J.R., Sh.S., M.J.H., D.G.A., A.Be. and E.S.L. contributed significantly to the analysis. K.A.S., L.N., A.S., R.A.D., J.G.H., R.F.H., Z.M., C.D., D.G.A., A.Be. and E.S.L. contributed significantly to the experimental and technical design. T.E.B., J.A.M., T.L., Sh.S. and A.Be. contributed significantly to generation of the data. S.M.S., K.A.S., L.N., T.L., P.W., J.G.H., M.J.H., J.W.P., C.D., A.R.J., D.G.A., A.Be. and E.S.L. contributed significantly to supervision and management of the project. K.A., N.De., T.D., J.G., G.Gu., C.L.K., C.L., C.-K.L., P.D.P., T.R., K.R. and D.S. contributed to the analysis. N.De., T.A.D., A.E., J.G., R.A.G., G.Gu., D.R.H., A.H.-S., J.J., C.L.K., C.L., C.-K.L., F.L., J.N., J.R., Z.L.R. and W.B.W. contributed to the experimental and technical design. K.A., J.A., C.B., J.L.B., D.B., K.Bi., A.Bo., K.Br., S.B., E.B., S.Cal., A.C., S.Cat., M.C., J.C., N.De., T.A.D., N.Do., G.Ge., T.L.G., J.G., L.G., B.G., G.Gu., J.H., Z.H., N.H., R.H., M.K., A.K., C.L.K., C.L., C.-K.L., F.L., N.Ma., R.M., J.M., N.Mo., E.M., K.N., J.N., A.O., E.O., J.Pa., S.P., J.Pe., L.P., M.R., Z.L.R., T.R., B.R., K.R., D.R., M.S., N.Sh., S.S., N.Sj., A.J.S., R.T., L.V., U.W., W.B.W., C.Wh., J.W., R.Y. and B.L.Y. contributed to generation of the data. A.Bo., E.B., M.C., T.D., T.A.D., A.E., E.F., B.G., M.K., C.L., L.L., N.Ma., S.P., M.R. and A.J.S. contributed to supervision and management of the project. T.E.B., J.A.M. and E.S.L. wrote the manuscript.

Author Information

Detailed technical white papers describing tissue processing and microarray profiling are available at the Allen Brain Atlas portal (www.brain-map.org) through the Non-Human Primate link, or directly from the NIH Blueprint NHP Atlas website (www.blueprintnhpatlas.org), under the 'Documentation' tab. Microarray data can be viewed online by selecting 'Microdissection' under the 'Microarray' tab and can be downloaded under the 'Download' tab. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests.

The primate brain develops through a series of stereotyped processes that are conserved across mammals[1], including the specification, migration and maturation of diverse cell types and the formation and refinement of functional neuronal circuits. There are also primate-specific features of brain development that increase anatomical[2], cognitive[3], and behavioral complexity[4] and may explain why many human neurological and neuropsychiatric diseases are not well modeled in rodents[5]. These features include a larger cortical progenitor pool in the outer subventricular zone not present in rodents[2] and protracted myelination[6], synapse production[7], and pruning[8]. Consequently, rhesus monkey and human share a greatly expanded neocortex and specialization of areas (most notably primary visual cortex[2]), compared to mouse, reflecting the more recent common ancestor of human and rhesus monkey (25 million years) than human and mouse (70 million years)[9]. Likewise, rhesus monkey and human brain[10] share more similar patterns of gene expression than do mouse and human brain[11]. Rhesus monkey thus provides a valuable proxy for human brain development, particularly during prenatal and early postnatal development that are difficult to sample in human, and also provides a comparator to study human-specific features such as prolonged maturation or neoteny[6,12,13].

Molecular programs responsible for brain development remain incompletely understood in any species, due to the spatiotemporal complexity of these processes and the resource intensity of methods needed to probe them. Transcriptome-based approaches have dramatically accelerated understanding of variation in gene expression programs related to brain structure and function in adult and developing human[14–17] and rhesus monkey[10,18], albeit with limited anatomical resolution and developmental coverage. Recent studies in adult mouse[19,20] and human[21] have profiled individual cortical cells and demonstrated robust transcriptional differences between neuronal and non-neuronal cell types and large-scale changes over the course of development. While these approaches offer cellular resolution in targeted domains, a global picture of gene expression changes over development across the complete cellular milieu has been lacking

This project aimed to create a transcriptional atlas of non-human primate brain combining fine anatomical precision with dense temporal coverage across prenatal and postnatal development in regions associated with neuropsychiatric disease. These transcriptional data are complemented by imaging and histology-based anatomical reference and cellular resolution *in situ* hybridization (ISH) datasets, all publicly accessible through the NIH Blueprint Non-Human Primate (NHP) Atlas (www.blueprintnhpatlas.org; www.brain-map.org).

Spatiotemporal transcriptome analysis

This resource characterizes rhesus monkey forebrain development with three modalities: 1) anatomical reference datasets consisting of magnetic resonance imaging (MRI) and corresponding densely sampled Nissl stains; 2) cellular resolution gene expression data generated with a high throughput ISH platform; and 3) fine anatomical resolution transcriptome timeseries data generated with a combination of laser microdissection (LMD) and DNA microarrays (Fig. 1a). Molecular analyses focused on discrete progenitor and postmitotic cell populations in neocortex (medial prefrontal and visual cortical areas),

hippocampus, striatum, and amygdala (Fig. 1b, Supplementary Tables 1–2). Ten developmental timepoints were chosen to correspond to peak periods of neurogenesis for neurons destined for different layers and glial cell types of V1 (prenatal)[22] and major developmental epochs: neonate, infant, juvenile, and young adult (postnatal; Fig. 1c). Anatomical parcellation at each time point was based on prior work in monkey and human brain development[2,17] (Extended Data Fig. 1). Canonical genes marking different neural cell types showed expected spatiotemporal patterning across development, shown as heatmaps (Fig. 1d and Extended Data Fig. 2) laid out following the anatomical organization in Fig. 1c.

To gain a high level understanding of transcriptional dynamics during cortical development, we represented the similarity between cortical samples using multidimensional scaling (MDS) based on correlated expression (Fig. 1 e–g). Layer and age explained almost 90% of the variation across all samples (n = 922; Fig. 1e) and prenatal samples (n = 542; Fig. 1f). To examine smaller differences between neuronal populations (grey circles in Fig. 1f), we repeated the MDS analysis only on V1 cortical layers (containing different types of excitatory neurons; n = 175; Fig. 1g). Transcriptional similarity of layers reflects spatial proximity and birth date, as described in adult rhesus monkey[10]. Moreover, continuous variation across prenatal development suggests gradual changes in gene expression.

There were few sex-related differences in expression, although the study was not powered to detect subtle differences restricted to specific brain regions or developmental stages. Only one autosomal gene (*LOC693361*, NADH-ubiquinone oxidoreductase chain 1-like) showed robust increased expression in males across brain regions during prenatal development (Extended Data Fig. 9, Supplementary Table 3). However, the microarray probe for *LOC693361* also targets an unannotated region of rhesus monkey Y chromosome[23] so there are likely no autosomal genes with persistent male- or female-specific expression differences across development.

Developmental transcriptional dynamics

Previous work has demonstrated that different cortical layers have distinct transcriptomic profiles both in adult and developing cortex, where the largest differences are seen between germinal and postmitotic cell populations[17]. In grossly dissected developing human prefrontal cortex, gene expression changes fastest prenatally and slows sharply just after birth[14]. If dividing cells drive most expression change, then depletion of the progenitor pool would explain decreased rates. In contrast, we find that all layers of V1 and prefrontal cortex (anterior cingulate gyrus; ACG) have similar rates of expression change across development (Fig. 2a). Thus transcriptional programs change dramatically both in progenitor cell populations as they give rise to different neuronal and non-neuronal cell types, as well as in postmitotic neuronal cell types as they differentiate and mature into adulthood.

Similar developmental trajectories were seen in hippocampus, basal ganglia and amygdala (Extended Data Fig. 4a). However, the genes changing dynamically could differ by region or age. We identified the top 1000 genes with the largest increase and decrease in relative expression between all pairs of ages for each structure and visualized overlap of these gene lists as a heatmap (Fig. 2b). This threshold was selected because at least 1000 genes

significantly changed expression (ANOVA FDR < 0.05, fold change > 1.25) between all adjacent ages in the majority of regions (Extended Data Fig. 4b). Genes changing over time were remarkably synchronized across different cell populations. At every age, approximately half of the top increasing and decreasing genes were shared by most brain regions, and unsupervised clustering grouped samples almost perfectly by age (age index bar in Fig. 2b). Genes increasing with age showed substantial overlap across ages within, but much less between, prenatal and postnatal ages. This suggests that distinct transcriptional programs are active in prenatal and postnatal development, and that many of these programs progress gradually. At early stages, samples grouped by proliferative state independently of brain region (mitotic index bar in Fig. 2b) reflecting a common set of cell division genes. Postnatally, samples grouped more by anatomical structure, reflecting increased regional identity during brain maturation (structure index bar in Fig. 2b).

To interpret these dynamic gene expression patterns, we searched for enrichment of gene ontology (GO) terms at each region and age (Supplementary Tables 4–5) and represented significant enrichments as heatmaps ordered as above (Fig. 2c). For example, genes related to onset (positive regulation) of axonal projection were enriched during prenatal periods as expected. Surprisingly, this process appears to be actively repressed during late postnatal development, as genes associated with offset (negative regulation) of axonal projection showed increased expression from juvenile (12 months) to young adult (48 month) stages. Processes associated with synapse development appear to be temporally enriched in late prenatal and early postnatal periods, and also synchronized between presynaptic (synaptic vesicle localization) and postsynaptic (dendrite development) neurons. Interestingly, the region-specific blocks of coordinated expression in postnatal visual cortex (V1 and V2) and hippocampus (Fig. 2b, arrows) were enriched for autophagy (Extended Data Fig. 4c), which may reflect early dendritic spine pruning[24] and grey matter volume reduction[25] in visual cortex.

Regional timing of biological processes

Some biological processes, such as synaptogenesis and myelination, are protracted over development[6,24], and we find that process activation (increasing genes) is longer than inactivation (decreasing genes; Fig. 3a). Biological processes persist longer in all regions pooled together (black lines) than in any individual region (colored lines) so developmental timing must vary between regions. We quantified synchrony between pairs of regions as the proportion of ages during which a process is active in either region, for processes active in all regions at some age (e.g., Extended Data Fig. 5a). Remarkably, despite major differences in cellular and functional makeup of brain regions sampled, all pairs of regions showed similar degrees of synchrony across all GO terms. This synchrony was centered on overlap of approximately half of all ages (Extended Data Fig. 5b), far greater than expected by chance. Few processes – cell proliferation, projection, and adhesion – were synchronized across all regions and ages (Supplementary Table 5).

Processes usually started before birth in all regions, but sometimes at strikingly different times in different regions (Fig. 3b). Consistent with known early maturation of subcortical circuits[1], rank ordering age of onset for each process across regions showed earlier

subcortical onset, particularly in amygdala, and similar timing for neocortex and hippocampus (Extended Data Fig. 5c). For example, subcortical globus pallidus is myelinated before cortical white matter[26] and shows early increased expression of genes specific to myelinating oligodendrocytes (*MOG*, *MOBP*, and *ERMN*) (Fig. 3c and Extended Data Fig. 5d). Interestingly, three other oligodendrocyte markers (*MAL*, *ASPA*, and *OPALIN*) turned on synchronously in all regions, suggesting a temporal and regional dissociation of different stages of myelination.

Late emergence of mature cortical identity

The neocortex is composed of diverse neuronal subtypes organized into layers that have distinct connectivity and transcriptional signatures in the adult[10,20,27] with graded and discrete variation between cortical areas[15,16,28]. A neuron develops its mature molecular identity through cell-autonomous processes and environmental interactions during migration, process extension, and synapse formation. If early generated neurons closely resemble their mature states, then environmental interactions must have small effects on their terminal molecular identities. Since we found dramatic expression changes within postmitotic cortical layers (see Fig. 2a), there is a significant role for either or both of these mechanisms.

To examine the emergence of mature laminar signatures, we compared layer-enriched expression patterns across development from the earliest stage at which all postmitotic layers are present (E80) to adulthood. We calculated the average expression correlation, across layers and between all pairs of ages, of the top 2.5% of genes most specific for each layer in V1 at each age (see **Methods**, Supplementary Table 6). Adult identity emerged gradually (Fig. 4a), as evidenced by decreased expression similarity with increased age separation. In fact, laminar expression patterns in young adulthood barely resembled mid-fetal cortex ($R < 0.1$).

These correlation trends result from nearly complete shifts in genes showing robust layer-specific expression (fold change > 1.5; Fig. 4b). Most genes were enriched only for a subset of contiguous ages at early, middle, and late stages, while few genes showed persistent laminar enrichment (Fig. 4b,c, Supplementary Table 7). Sets of laminar genes tile smoothly across development, suggesting a graded emergence of functional characteristics of neuronal subtypes as well as shifting proportions of subtypes within layers. Genes with early laminar expression are enriched for GO terms related to early developmental events such as neuron differentiation ($P = 1.6 \times 10^{-7}$) and axon generation ($P = 6.7 \times 10^{-6}$), whereas later laminar genes are enriched for axon guidance ($P = 6.5 \times 10^{-5}$) and innervation ($P = 9.8 \times 10^{-6}$).

Different anatomical and functional neocortical areas have distinct gene expression patterns in the adult, most prominently in V1[15,16,28]. During development, intrinsic (protomap)[29] and extrinsic (protocortex)[30] factors shape cortical area identity. To ask whether the temporal dynamics of areal expression support intrinsic versus extrinsic areal specification, we first identified genes differentially expressed between caudal (V1, blue) and rostral (ACG, red) cortex in different layers at each developmental stage (Fig. 4d, Supplementary Table 8). More genes were differentially expressed between cortical areas in postmitotic layers during

late postnatal than prenatal development. These genes were enriched for cell adhesion ($P = 3.9 \times 10^{-6}$) and synaptic transmission ($P = 7.3 \times 10^{-4}$) GO terms, suggesting a role for extrinsic factors related to the functional maturation of cortical circuitry. Developmental process timing varies across primate cortex, with neurogenesis occurring earlier in frontal than caudal cortex but circuit maturation persisting later in frontal cortex[6,8], and this was recapitulated by an even more protracted emergence of enriched genes in ACG (12 to 48 months) compared to V1 (3 to 12 months).

To test for genes that may contribute to intrinsic specification, we searched for genes with persistent areal enrichment beginning prior to the arrival of thalamic[31] or other extrinsic inputs. While 38% of genes showed regional differences, only 1% were persistent over development (Supplementary Table 8). These genes were primarily expressed in postmitotic cortical layers, for example *CBLN2*[17], and only 20 genes (0.2%) were expressed in germinal zones, including the well-known caudal-to-rostral gradient genes *FGFR3*[17,29] and *NR2F1* (*COUP-TFI*)[29]. Thus, while there is evidence for a small number of genes clearly reflecting early intrinsic areal specification, there are many more adult differences that could be due to extrinsic interactions and activity-dependent mechanisms[32].

Some areal differences are likely due to differences in the proportion and maturation state of specific cell types, such as dividing neural progenitors and postmitotic neurons and glia. Indeed, we identified spatiotemporal locations with significant enrichment in V1 or ACG for markers of cell cycle (i.e. actively dividing progenitor cells), GABAergic and glutamatergic neurons in adult mouse V1, and astrocytes (Fig. 4e–h). Cell cycle markers were enriched in V1 SZ from E80-E90 (Fig. 4e), reflecting extended generation of superficial excitatory neurons in V1 relative to ACG. Likewise, markers for excitatory neurons (Fig. 4g) were enriched in ACG prenatally, corresponding to earlier generation and maturation of these cells. Postnatal enrichment in V1 is likely secondary to using markers derived from adult mouse V1. Astrocyte marker enrichment in prenatal ACG progressed from proliferative to postmitotic layers (Fig. 4h), tracking astrocyte generation and migration in these layers. Finally, markers of inhibitory GABAergic interneurons showed early enrichment in V1 SZ, followed by ACG enrichment in several layers (Fig. 4f), a potential signature of early interneuron migration from medial ganglionic eminence to caudal cortex.

The majority of areal genes showed intermediate expression in primary somatosensory cortex (S1), located approximately midway between rostral ACG and caudal V1 (Extended Data Fig. 6). Therefore, rostro-caudal expression gradients, rather than specific features of ACG and V1, likely drive many of these areal expression differences.

Mapping disease genes to development

Highly heritable but genetically complex neurodevelopmental disorders will be more tractable if we can identify common pathways linking many candidate risk genes. For example, recent work on autism spectrum disorder (ASD) demonstrated that candidate genes are co-expressed in mid-fetal human cortical neurons[17,33,34] and converge on synaptic development pathways. Large-scale genetic association studies have identified high confidence genes associated with primary autosomal recessive microcephaly (MCPH), ASD,

intellectual disability (ID), and schizophrenia (SCZ). We localized these disease genes in developing cortex using this new high resolution map.

We used weighted gene co-expression network analysis (WGCNA) to identify sets of genes with correlated patterning (modules) in cortical samples at each age, linked modules with overlapping genes between adjacent ages, and annotated modules for enrichment in different cortical layers and cell types (Fig. 5a; Supplementary Table 9). Next, we tested disease gene sets for significant enrichment (hypergeometric test, corrected $P < 0.1$). As expected, MCPH genes showed early- to mid-fetal enrichment in neuronal progenitor-enriched modules (Fig. 5b) where they are positioned to alter cell cycle kinetics to reduce neuron numbers and brain size. Many ASD genes showed coordinated expression in post-mitotic neurons prenatally (Fig. 5c), as previously reported[17,33,34], but also postnatally. In contrast, SCZ genes were selectively enriched only in postmitotic neuronal modules from infancy through adulthood (Fig. 5d).

ID-associated genes were not enriched in any gene modules, consistent with the view that disruption of many biological pathways can undermine cognitive development[34]. To investigate whether these diverse molecular pathways converge on intermediate phenotypes within ID such as structural brain malformations, we looked for correlated expression among the 71 ID genes across prenatal and postnatal cortex. We identified four major expression patterns: 1) postnatal cortical plate, 2) prenatal germinal layers, 3) prenatal subventricular zone and cortical plate, and 4) more complex patterns spanning development (Fig. 5e). Consistent with the module analysis above, MCPH-associated genes had significantly similar expression profiles (multivariate distance matrix regression, $P < 0.001$) and were enriched in prenatal germinal zones in clusters 2 and 3 in either VZ or SZ. In contrast, other phenotypes did not share expression patterns, including cortical atrophy, seizures and even microcephaly with other brain malformations. Furthermore, genes with similar patterns can give rise to different phenotypes: *ASXL3* and *GRIP1* are co-expressed ($R = 0.80$) in cluster 3 but are associated with MCPH and corpus callosum agenesis, respectively.

Human-specific developmental trajectories

Humans exhibit unique cortical and cognitive developmental trajectories compared to other mammals[5] that are likely correlated with differences in gene expression. To identify conserved and divergent genes, we compared gene expression trajectories across prenatal and postnatal development in frontal cortex between rat, rhesus monkey, and two independent human data sets, normalizing timescales by developmental event scores[1]. Expression conservation varied across species (Kruskal-Wallis rank sum test, $P < 10^{-21}$), with greater similarity within humans than within primates (Wilcoxon signed rank test, $P < 10^{-6}$), as expected based on evolutionary distance (Fig. 6a). Surprisingly, there was no significant difference in conservation of rat trajectories with either human or rhesus monkey, indicating transcriptional patterning has evolved at the same average rate in human and rhesus monkey since our last common ancestor. On average, 69% of genes had conserved expression trajectories ($R > 0.5$) across all three species, and genes with the most dynamic expression change over development were even more highly conserved (Extended Data Fig.

7a,b). For example, *EMX2*, a transcription factor critical to cortical arealization[35], and *CNTN1*, a cell adhesion gene that guides neuronal migration and connectivity (Fig. 6b), are highly conserved.

A significant minority (13%) of genes differed between primates and rat (validated for a subset of genes in developing mouse; Supplementary Table 10) and approximately 9% had human-specific expression trajectories. The proportion of genes showing different conservation patterns was robust to gene selection and correlation threshold (Extended Data Fig. 7b). Some of these differences were dramatic; for example, *BMP3* decreases over primate development but increases in rodents (Fig. 6b). *BMP3* is a *WNT* signaling growth factor that has been linked to craniofacial variation in dogs[36]. *CNTN2*, a close family member of the conserved adhesion gene *CNTN1*, likewise shows opposite developmental trajectories in primates and rodents. In both human data sets, *LGALS1* and *LIN7A* increase or decrease expression, respectively, in opposite directions to other species. *LGALS1* further has a different adult laminar pattern of cortical expression in human compared to mouse[11], while deletion of *LIN7A* is associated with human intellectual disability and disrupted neuronal migration and axonal pathfinding in mouse[37].

Some aspects of cortical development, such as myelination[6] and synaptic pruning[8], are protracted or neotenous between cortical areas or between species. Reasoning that abrupt changes in expression may represent important developmental milestones, we compared developmental timing by identifying distinct changes in expression trajectories, or breakpoints. 179 increasing and 179 decreasing genes met these criteria in all three species, being well fit by segmented linear regression (Supplementary Table 11). Consistent with different developmental rates of these species, breakpoints for increasing (Fig. 6c) and decreasing (Extended Data Fig. 7c) genes often occurred earliest in rat, intermediate for rhesus monkey, and latest in human. Almost half (81) of increasing genes were synapse related (dashed lines, Fig. 6c), and breakpoints coincided with ages of peak synaptic density estimated in these species (shaded rectangles, Fig. 6c, Extended Data Fig. 7d). There was no significant difference in the breakpoints for synapse-related genes in human V1 versus prefrontal cortex and only a short delay in rhesus monkey V1 (42 days, $P = 9.6 \times 10^{-9}$, Extended Data Fig. 7e), consistent with synchronous synaptogenesis reported in primates[7,38–40]. One study later reported protracted synaptogenesis in human[41], but this study suffered from sparse sampling and no statistical analysis and does not appear consistent with the bulk of evidence.

While breakpoint timing was largely conserved between species, particularly between rhesus monkey and rat (Extended Data Fig. 7f), human breakpoints were clearly bimodally distributed (Fig. 6c) suggesting that some genes had different developmental timing. For example, *OLIG1* is expressed in maturing oligodendrocytes and had a late breakpoint relative to rhesus monkey (Fig. 6d), consistent with prolonged myelination in several human cortical areas compared to rhesus monkey and chimpanzee[6]. Surprisingly, many more genes had an early rather than late breakpoint in human prefrontal cortex (Fig. 6e), in contrast to previous work[12,13] showing human-specific delayed peak expression in this brain region. However, we found that many early genes continued to increase expression through adulthood (Fig. 6e), such as *SYT7* (Fig. 6d), a presynaptic calcium-binding protein

important for modifying neurotransmitter release[42]. In rhesus monkey, gene expression often did not change significantly after breakpoints. Genes with an early breakpoint and late maximal expression may mark developmental processes that are protracted in human relative to non-human primates and that underlie our extended cognitive development. Interestingly, we found this transcriptional signature of neoteny between species in both prefrontal cortex and V1, but much less so in evolutionarily older structures including hippocampus, amygdala and striatum (Fig. 6e and Extended Data Fig. 7g).

Discussion

The current project transcriptionally profiled primate brain development with fine anatomical detail from early gestation to young adulthood. While expression rates of change decreased more than 100-fold over development, we observed small but coherent changes from juvenile to adult, a period of enormous cognitive change and susceptibility to neuropsychiatric disease. These changes related to the late emergence of mature laminar and areal signatures, biological pathways such as negative regulation of axonal pathfinding, and the appearance of gene modules significantly associated with ASD and SCZ.

Two unexpected characteristics of expression trajectories support extrinsic influences on developmental transcriptional regulation. First, there was a striking synchrony among genes changing in disparate brain regions and cell populations, suggesting a mechanism for global regulation such as circulating hormones. Second, the surprisingly late acquisition of adult-like cortical areal and laminar molecular phenotypes points to an important role for contextual and activity-dependent mechanisms in sculpting mature cellular phenotypes[43,44].

We show that 22% of genes have different developmental trajectories in rat and human, comparable to 25% of genes that have different laminar patterns in adult mouse and human cortex[11]. In contrast, approximately 9% of genes have different trajectories in rhesus monkey and human including genes with delayed peak expression solely in human cortex. Therefore, rhesus monkey is an important comparator for understanding human-specific features of brain development but cannot fully model protracted circuit formation and associated diseases that are seen only in humans.

A recently expanded set of ASD candidate genes[45] were significantly enriched in newborn neurons during prenatal development, as previously reported[17,33,34], but also enriched throughout postnatal development. SCZ risk genes were also enriched in neurons but not until infancy, suggesting a larger role for dysfunction in circuit refinement than prenatal processes such as neurogenesis. SCZ and ASD were enriched in the same neuron-enriched module in infancy despite only 5% overlap in candidate genes, pointing to a shared etiology that is being increasingly appreciated[46].

This data resource (www.blueprintnhpatlas.org; www.brain-map.org) has many potential applications; for example, establishing an *in vivo* baseline against which to compare the identity and maturity of *in vitro* stem cell-derived neurons[47]. Recent technical advances enable profiling the full transcriptome[19–21], epigenome[48], and ultimately proteome of single cells. These techniques promise to refine this broad survey to a causal understanding of

molecular programs driving the complete lineage of primate brain cells and the maturation of specific neuron types in functional circuits.

Methods

No statistical methods were used to predetermine sample size, and the experiments were not randomized or blinded. More complete descriptions of the experimental and data processing methods are available in white papers at the NIH Blueprint NHP Atlas website (www.blueprintnhpatlas.org; “Documentation” tab).

Animals

Frozen postmortem tissue samples from prenatal rhesus macaque (*Macaca mulatta*) of predominantly Indian origin (Extended Data Table 1) were provided by the Time-Mated Breeding Program at the California National Research Primate Center (CNPRC; www.cnprc.ucdavis.edu). Prenatal brain material was acquired following fetectomy using Standard Operating Procedures (SOPs) at the CNPRC. Timed pregnancy derived biological replicate specimens (2 males, 2 females) at each of six prenatal developmental stages (E40, E50, E70, E80, E90, and E120) were profiled. These timepoints were selected to coincide with peak periods of neurogenesis for the different layers of primary visual cortex based on birthdating experiments[22]. Frozen postmortem tissue samples from postnatal rhesus macaque were also provided by CNPRC. Brain regions were systematically collected from well-characterized rhesus monkeys born and raised at the CNPRC in outdoor, half-acre enclosures that provide a naturalistic setting and normal social environment. For transcriptome analysis, three male specimens at each of four postnatal developmental stages representing the neonate (0 months), infant (3 months), juvenile (12 months) and post-pubertal adult (48 months) were profiled. Extensive health, family lineage and dominance information were maintained on all animals. All procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at UC Davis.

Postnatal MRI and histological reference series

An interleaved anatomical atlas of magnetic resonance imaging (MRI), histology, and block face photographs was created from four male rhesus macaques at 2 weeks, 3, 12 and 48 months after birth as described in the “Reference Series” white paper. In addition, ISH data were generated serially across complete hemispheres of three post-pubertal adult (48 months) male specimens. Genes were selected to mark specific cell populations and cortical areas and to represent gene families important to neural function and associated with neuropsychiatric disease and brain evolution. Several histological stains (Nissl, acetylcholinesterase and SMI-32) of complete hemispheres were combined with ISH data to provide an unprecedented cellular resolution view of gene expression in the whole post-pubertal adult brain. High-throughput colorimetric ISH methods are described in detail elsewhere[49] and in the “*In Situ* Hybridization” white paper.

Tissue processing

Following collection, brain tissues were partitioned in a manner dependent on specimen stage, gradually frozen, and then stored at −80°C until processing. For a subset of E40

specimens, following specimen collection, the calvarium was frozen intact in an isopentane/dry ice slurry maintained at −40°C to −45°C, gradual freezing at a steady rate. For the majority of E40 specimens, the specimens were embedded in OCT (optimal cutting temperature compound) during the freezing process. In brief, chilled OCT was placed around the calvarium. A disposable embedding chamber was filled with approximately 5 mm^3 chilled OCT. The specimen was carefully oriented and centered in the OCT, posterior surface down in the OCT. Then, the specimen was aligned along the medial/lateral axis using the bilateral ocular fiduciaries as a frame of reference. Next, the specimen was aligned in the coronal plane. After alignment along all axes, OCT was added to encase the specimen in its entirety. The top of the specimen was covered with approximately 3 mm of OCT. The chamber containing the specimen was directly placed onto a level bed of dry ice. The specimen and OCT were allowed to freeze completely. After demarcation of the orientation of the brain in the OCT block, the frozen tissue block was stored at −80°C.

For the E50 specimens, following removal of the brain from the skull, the whole brain was frozen intact in an isopentane/dry ice slurry maintained at −40°C to −45°C, gradual freezing at a steady rate. For a subset of E50 specimens, the specimens were embedded in OCT during the freezing process. In brief, chilled OCT was placed around the intact brain. Freestanding aluminum bars were assembled onto a Teflon coated plate and sized to the appropriate specifications for the E50 brain. The internal chamber was filled with approximately 5 mm^3 chilled OCT. The specimen was carefully oriented and centered in the OCT dorsal surface down in the OCT. Then, the specimen was aligned along the medial/lateral axis using the longitudinal fissure as the frame of reference. Next, the specimen was aligned in the coronal plane. After alignment along all axes, OCT was added to encase the specimen in its entirety. The top of the specimen was covered with approximately 3 mm of OCT. The Teflon plate containing the specimen was directly placed onto the level bed of dry ice. The specimen and OCT were allowed to freeze completely. After demarcation of the orientation of the brain in the OCT block, the aluminum bars were removed and the frozen tissue block was stored at −80°C.

For the E70, E80, E90, and E120 specimens, the hemispheres were bisected along the midline and individually frozen by placing the medial aspect of each hemisphere down onto an aluminum-Teflon coated plate that was slowly lowered into an isopentane/dry ice slurry maintained at −40°C to −45°C. Only approximately a third of the tissue was submerged in the slurry to allow the tissue to gradually freeze and to keep freezing artifacts to a minimum. Frozen hemispheres were stored at −80°C.

Depending on the prenatal timepoint, different approaches were taken for generating coronal slabs. When possible, the number of slabs per specimen was kept to a minimum. The E40, E50, and E70 specimens were not slabbed. For E80, the first slab contained up through the temporal pole and the second slab contained the occipital pole. For E90, the first slab contained the frontal lobe anterior of the temporal pole and the second slab contained temporal pole posterior through the occipital lobe. For E120, three coronal slabs were made. The first slab consisted of the frontal lobe anterior of the temporal pole. The second slab consisted of the temporal pole posterior to the cerebellum including all of the mid-brain. The third slab included primarily the occipital lobe.

For postnatal brains, after dissection brains were sectioned into coronal slabs approximately 1 to 1.5 cm in thickness and the left hemisphere was prepared for sectioning onto microscope slides for ISH. Structures for microarray analysis were isolated from the right hemisphere slabs, and these samples were then frozen at −80°C until processed further.

Laser microdissection and RNA isolation

Tissue spanning five anatomically distinct brain regions — frontal cortex, visual cortex, hippocampus, striatum, and amygdala (Figure 1; Supplementary Table 1) — was selected from each specimen and processed for further thin sectioning and LMD using a standard protocol. Specifically, frozen tissue was cryosectioned at 14 μm onto polyethylene naphthalate (PEN) slides (Leica Microsystems, Inc., Bannockburn, IL) and a 1:10 Nissl series was generated for neuroanatomical reference for all prenatal timepoints. In addition, for the E40, E50, E70, E80, and E90 timepoints, a 1:10 *GAP43* and 1:10 *ENC1 in situ* hybridization (ISH) series was generated for neuroanatomical reference, as they often clearly delineate nuclei and layers at early developmental stages. For E120, a 1:10 acetylcholinesterase series was generated for neuroanatomical reference. In total 127 transient and terminal anatomical structures were isolated using this strategy[50–53].

After drying for 30 minutes at room temperature, PEN slides were frozen at −80°C. Slides were lightly Nissl stained with cresyl violet to allow cytoarchitectural visualization. Slides were fixed in ice-cold 70% ethanol for 30 seconds, washed 15 seconds in nuclease-free water, stained in 0.7% cresyl violet in 0.05% NaOAc, pH 3.4 for 2 minutes, nuclease-free water for 15 seconds, followed by 15 seconds each in 50%, 75%, and 95% ethanol, followed by 20 seconds in 100% ethanol, and then a final 100% ethanol wash for 25 seconds. Slides were air-dried for 2 minutes, and desiccated in a vacuum for 1 hour, then frozen at −80°C until microdissection. Laser microdissection was performed on a Leica LMD6000 (Leica Microsystems, Inc., Bannockburn, IL), using the Nissl stain and GAP43 and ENC1 ISH or acetylcholinesterase histological staining as a guide to identify target brain regions in prenatal samples.

Microdissected tissue was collected directly into RLT buffer from the RNeasy Micro kit (Qiagen Inc., Valencia, CA) with β-mercaptoethanol. Samples were volume-adjusted with water to 75 μl, vortexed, centrifuged, and frozen at −80°C.

RNA was isolated for each structure following the manufacturer's directions for the RNeasy Micro kit. RNA samples were eluted in 14 μl and 1 μl was run on the Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA) using the Pico assay. Due to low sample volume and incompatibility of the eluant with the Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE), samples were quantitated using the Bioanalyzer concentration output. This was done by running a 1ng/μl RNA standard on the same Pico chip and then dividing the sample concentration output by the output of the standard concentration. For prenatal samples, 2 ng of total RNA was almost always used as the input amount for the labeling reaction, and the average RNA Integrity Number (RIN) of passed samples was 7.5, with RINs typically lower than 4.5 failing. For postnatal samples, 5 ng of total RNA was almost always used as the input, and the average RIN of passed samples was 7.5, with RINs typically lower than 5 failing.

Histological stains

Nissl—For the Nissl neuroanatomical reference slides, slides were stored at 37°C for 1–5 days prior to staining. Sections were defatted with xylene or the xylene substitute Formula 83, and hydrated through a graded series containing 100%, 95%, 70%, and 50% ethanol. After incubation in water, the sections were stained with 0.213% thionin, then differentiated and dehydrated in water and a graded series containing 50%, 70%, 95% and 100% ethanol. Finally, slides were incubated in xylene or Formula 83, and coverslipped with the mounting agent DPX. After drying, slides were cleaned prior to digital imaging.

Acetylcholinesterase—A modified acetylcholinesterase protocol was used to assist in neuroanatomical delineations in E120. Unlike acetylcholinesterase staining in fixed tissue, staining in fresh frozen tissue does not elucidate cholinergic fibers commonly seen in the cortex, but instead provides demarcation of various subcortical nuclei. Acetylcholinesterase staining was performed using a direct coloring thiocholine method combined with a methyl green nuclear counterstain to improve tissue visibility[54]. Glass slides with fresh frozen tissue sections were removed from 4°C, allowed to equilibrate to room temperature, fixed in 10% neutral buffered formalin (NBF) and washed briefly in ultra-pure water. Sections were then incubated for 25 minutes in a solution of acetylthiocholine iodide, sodium citrate, cupric sulfate, and potassium ferricyanide in a 0.1M sodium acetate buffer (pH 6.0), washed in 0.1M Tris-HCl buffer (pH 7.2), and incubated with 0.5% diaminobenzidine (DAB) in 0.1M Tris-HCl with 0. 03% hydrogen peroxide for 8 minutes. Slides were incubated in 0.2% methyl green, briefly dipped in 100% ethanol, cleared with Formula 83 and coverslipped with DPX. After drying, slides were cleaned prior to digital imaging.

In situ hybridization (ISH)—High-throughput colorimetric ISH methods are described in detail elsewhere (see Lein et al. 2007) and in the *In situ* Hybridization white paper located under the “Documentation” tab. In brief, fresh frozen tissue sections (from either E40, E50, E70, E80, or E90) on slides were fixed in 4% PFA in PBS, acetylated, and dehydrated through graded alcohols. Endogenous peroxidase activity was blocked with 3% hydrogen peroxide in methanol. Proteinase K digestion at 0.0135 unit/ml was done. Digoxigenin (DIG)-labeled riboprobes (either *GAP43* or *ENC1*) were hybridized at 63.5°C for 5.5 hr, followed by stringency washes and a series of enzymatic reaction steps for detection and amplification of DIG signal. Sections were washed with EDTA, fixed in 4% PFA, and washed with acid alcohol (70% ethanol, adjusted to pH = 2.1 with 12 N HCl) to reduce background signal. Slides were coverslipped with Hydromatrix, subjected to quality control analysis and cleaned prior to digital imaging. The complete list of 46 ISH gene targets for ISH in postnatal rhesus monkey is also available in the “Documentation” tab.

Digital imaging and processing of histologically stained sections

Digital imaging was done using the ScanScope XT (Aperio Technologies Inc., Vista, CA). Final resolution of images was 1 μm/pixel. All images were databased and preprocessed, then subjected to quality control analysis to ensure optimal focus and that no process artifacts were present in the images. Images that passed this initial QC were further assessed to ensure that the staining data were as expected. Once all QC criteria were met, images became available for annotation of anatomic structures.

mRNA profiling

Prenatal samples passing RNA quality control (QC) were amplified and profiled. Briefly, samples were amplified and labeled using a custom 2 cycle Ultra Low Input procedure, using components of MessageAmp II aRNA Amplification kit (AM1751) for the first amplification cycle (using oligodT), and components of MessageAmp II Biotin Enhanced Single Round aRNA Amplification kit (AM1791) for the second amplification cycle (using both random hexamers and oligodT). For prenatal samples, 2ng of total RNA was added to the initial reaction mix together with 250ng of pBR322 (Invitrogen). Following the first cycle of IVT, the plasmid carrier was removed with a DNaseI (Qiagen) treatment. The first cycle IVT products were purified using the QiagenMinElute Kit (Qiagen). Input into the second cycle was normalized to 400ng. Hybridization was to catalog GeneChip Rhesus Macaque Genome Arrays from Affymetrix containing 52,803 probe sets/sequences. For detailed information about this macaque microarray, see the Affymetrix web site (http://www.affymetrix.com/browse/products.jsp?productId=131496&navMode=34000&navAction=jump&aId=productsNav#1_3). Labeling and scanning were completed following the manufacturer's recommendations. QC assessment failed a small number of microarray samples. Sample amplification, labeling, and microarray processing were performed by Covance in Seattle, WA.

Postnatal samples passing RNA quality control (QC) were amplified and profiled as described in Winrow *et al.*[55] with a few modifications. Briefly, samples were amplified and labeled using a custom 2 cycle version, using 2 kits of the GeneChip HT One-Cycle cDNA Synthesis Kit from Affymetrix. For postnatal samples, 5 ng of total RNA was added to the initial reaction mix together with 250ng of pBR322 (Invitrogen). *In vitro* transcription (IVT) for the first cycle was performed using a 5X MEGAscript T7 Kit (Ambion). Following the first round of IVT, the plasmid carrier was removed with a DNaseI (Qiagen) treatment. The first round IVT products were purified using the QiagenMinElute Kit (Qiagen). Input into the second round was normalized to 150ng for postnatal samples. Hybridization was to catalog GeneChip Rhesus Macaque Genome Arrays. Labeling and scanning were completed following the manufacturer's recommendations and profiles were normalized using robust multi-array (RMA). Sample amplification, labeling, and microarray processing were performed by Covance in Seattle, WA.

Microarray processing and normalization

To account for differences in mRNA processing, prenatal and postnatal samples were first normalized separately using a standard procedure, and then scaled together to minimize the effects of technical biases (e.g. due to differences in mRNA processing of prenatal and postnatal samples) while preserving biological variance. For prenatal samples, the BioConductor "affy" package was used to read in the Affymetrix microarray data and RMA method[56], consisting of background normalization, $\log_2$ transformation, and quantile normalization, was applied to summarize the probe level data into gene expression measurements for each batch. Control samples from macrodissected adult cortex were included in each batch to normalize the data across batches by aligning their mean gene expression values. Postnatal samples from each batch were preprocessed using the same RMA method as for prenatal samples. ComBat (http://statistics.byu.edu/johnson/ComBat/)[57]

was then applied to reduce more severe non-biological batch-to-batch bias. The postnatal microarray data are unchanged since the original public release. Finally, the normalized prenatal data was scaled to postnatal data by sorting probes by average expression across developed regions in E80 and older animals, finding the difference in average expression levels (postnatal – prenatal) of each group of 100 probes, and then scaling each probe by the corresponding value.

The purpose of normalization was to minimize the effects of these non-biological biases while keeping biological variance intact so that within and across brain comparisons primarily reveal differences and similarities that are biologically relevant. As a data-driven QC process, for each batch, we applied clustering/MDS (multi-dimensional scaling) to detect any outlier in the batch by checking whether samples from the same structure/age were grouped together. IAC (inter-array-connectivity) was also calculated to numerically measure how one microarray was similar to the other microarrays in the batch. The same QC process was applied over multiple batches to identify outliers. After normalization and outlier removal, a total of 1212 prenatal and 724 postnatal samples remained.

Probe set filtering and gene symbol assignments

From the 52,865 probe sets included on the microarray, we identified 12,441 high-confidence probe sets mapping to unique genes (Supplementary Table 2). First, to assess the targeting specificity of each probe set, we pooled mRNA from 20 prenatal and 20 postnatal discrete anatomical and temporal populations and ran these mRNA pools both on Affymetrix microarrays and using RNA-Seq as previously described[58]. Since a high quality transcriptome was not available for rhesus macaque, we used GSNAP to map RNA-seq reads to the rheMac2 (Jan. 2006) genome and identified expression values for 50,648 putative transcripts. We used Galaxy (usegalaxy.org)[59–61] to intersect rheMac2 genomic coordinates of probes (Affymetrix annotation version 34; 1/30/2014), GSNAP transcripts and NCBI rhesus gene information (downloaded 2/06/2014). 34,646 probe sets mapped unambiguously to 15,577 genes (i.e. rhesus Entrez ids) and were further considered. For each probe set, we calculated the correlation between probe set intensity and FPKM values for all RNA-seq transcripts corresponding to the same gene and retained the most highly correlated probe set for each gene. We omitted 3,136 genes that did not have any probe sets significantly correlated (FDR < 0.1) between methods. We annotated the final set of 12,441 genes with 10,715 human and 10,403 mouse orthologs using NCBI Homologene (downloaded 1/7/2014).

Comparing gene expression across neocortex

Heatmaps were independently generated for each gene by mapping expression levels in neocortex to a color vector of white (low expression) to red (high expression). Samples were organized in a grid corresponding to layer (vertical axis, from ventricular to pial surface), age (horizontal axis, from young to old), and cortical region (ordered rostral to caudal within each age). Each sample is labeled on the axes as well as within the corresponding boxes, with missing data points (either due to incomplete sampling or if a structure does not exist) left blank. The size of each box is chosen for clarity and is unrelated to gene expression.

We calculated the correlation between all neocortical samples based on the expression correlation of 12,441 genes. We converted correlations into distances with the transformation 1 - (0.5 * (1 + correlation)), applied metric multi-dimensional scaling (MDS) using the *cmdscale* R function to represent these distances in two dimensions, and estimated percent variance explained for the first two principal coordinates. We plotted samples with point color representing cortical layer and point size representing age. For clarity, samples that were annotated as cortical plate were assigned to postmitotic layers (L2-L6) that were most transcriptionally similar. We repeated this analysis for two subsets of samples: prenatal cortical layers and V1 postmitotic layers (layers 2–6).

Sex differential expression

Brain samples with data from both male and female donors included all prenatal ages and were included in a linear mixed model with sex as a fixed effect and brain region, age and donor as random effects. One model was fit for each gene individually using the "nlme" R package and P values were estimated using Satterthwaite's approximation to degrees of freedom as implemented in the "lmerTest" R package (see Supplementary Table 3). A permutation based correction for multiple comparisons was then performed. For each of 6 prenatal ages, donor sex labels (2 males, 2 females) were shuffled to generate 64 permuted data sets. For each of these permutations, an identical linear mixed model was fit to all genes, and genes were ordered by the magnitude of the fixed effect of sex. Median and 95% confidence interval (i.e. 2.5% and 97.5% quantiles) were calculated for each rank ordered fixed effect. Finally, a quantile-quantile plot (Extended Data Fig. 3) was created to compare observed fixed effects to median expected (permuted) effects. 11 genes had sex effects lying outside of the 95% confidence interval (i.e. greater expression in males), and 7 of these genes were nominally significant (mixed model P value < 0.05).

For probes that showed differential sex expression greater than expected by chance or for Y chromosome probes, target sequences were downloaded from Affymetrix (https://www.affymetrix.com/analysis/index.affx). These sequences were compared to a primate nucleotide database using NCBI Blastn[62], and the best match (Query coverage > 80%, Sequence identity > 90%) was reported in Supplementary Table 3. Note that 5 of 6 Y-chromosome probes likely measured expression of highly similar paralogs on the X-chromosome or autosomes, a consequence of the Affymetrix Rhesus Macaque microarray having been developed several years before the rhesus monkey Y chromosome was sequenced[23].

Analysis of dynamic genes using GO

We first calculated average gene expression in all brain regions (n = 3–4 monkeys) at 10 ages, and then calculated rates of expression change as the difference in log2-transformed expression between adjacent ages in each region and expressed as doublings or halvings per year. Note that rates of expression change could not be calculated for the earliest time point (E40) or for regions without sampling at adjacent ages. For each of 419 distinct brain regions, we rank ordered genes by rates of expression change, and selected the top 1000 increasing and decreasing genes. We compared all of these genes lists (419 * 418 = 175,142 comparisons). For each gene list comparison, we found the number of overlapping genes and

calculated an overlap similarity score based on the weighted rank order of the lists ("OrderedList" package in R). We then generated a null distribution of similarity scores by shuffling gene lists 10 million times and calculated P values for observed similarity scores. We found that 56,540 gene lists were significantly (Bonferroni corrected $P < 0.05$) overlapping.

We visualized the number of overlapping genes as a heatmap and reordered rows and columns based on the correlated patterns of overlapping increasing genes. Specifically, we calculated the distance between 419 brain samples as (1 – correlation) / 2 and used unsupervised hierarchical clustering with complete linkage to order samples. The dendrogram was reordered (while preserving the tree structure) by weighting branches by median ages of samples and rotating branches to group similar ages. Samples from the same age were highlighted with grid lines aligned to this dendrogram (Fig. 2b). The heatmap was also annotated with brain structure and a mitotic index approximated by mean expression of *MKI67* (Ki-67 protein).

We performed GO enrichment with the Bioconductor "GOstats" package in R with a background set of 10,715 human orthologs included in this study. We repeated GO enrichment analysis on each of 56,540 gene lists (32,053 increasing and 24,487 decreasing), and found 11,899 biological process terms were nominally significant ($P < 0.05$) for at least one list. To reduce false positive associations, we calculated the level of statistical support for each GO term as the proportion of children of that term (based on the GO hierarchy) that were nominally enriched. For example, if a GO term had 10 associated children and four were enriched, then the term had a support of 0.4. For each term, we plotted a heatmap of statistical support for all 56,540 gene lists (on a scale of 0 to 1), and many of these heatmaps showed distinct patterns of enrichment (e.g. Fig. 2c and Extended Data Fig. 4). We summarized GO enrichment as a table by calculating the average support for all 419 regions and considered a GO term active if mean support was greater than 0.01. We transformed this table into a binary matrix showing when and where GO terms were active. We removed terms associated with fewer than 10 of 419 regions since these terms likely represented noise, and this resulted in 3,704 processes increasing and/or decreasing over development (Supplementary Table 4).

We repeated the previous gene overlap and GO enrichment analyses with a set of five major brain regions including neocortex, hippocampus, basal ganglia and amygdala at nine ages and thalamus at four ages (40 total brain regions). We found that 556 of 1,600 gene list comparisons were significantly overlapping and 1,012 biological processes were active in these regions (Supplementary Table 5). For each GO term, we calculated commonality as the number of ages (maximum nine) that term was active in each brain region and in any region. We plotted the distribution of commonalities across all terms in different regions as density line plots separately for increasing and decreasing processes (Fig. 2d).

Relative timing of developmental processes

We calculated the relative timing of GO term activity using the binary matrix representing when and where terms were active in the four brain regions (neocortex, hippocampus, basal ganglia and amygdala) for which we had information across development. For each of six

possible region comparisons, we divided the number of ages a GO term was active in both regions by the number of ages it was active in either region (e.g. Extended Data Fig. 5a). For each term, we averaged the six age overlap calculations to summarize overall timing differences. We repeated these calculations for all 1,012 GO terms and plotted the distribution of age overlaps for increasing and decreasing processes (Extended Data Fig. 5b).

For 837 increasing processes, we plotted the earliest age of activity in four brain regions and ordered terms by age of onset in amygdala, basal ganglia, hippocampus and neocortex (Fig. 3c). We characterized timing as either simultaneous (equal age of onset in all regions), subcortex early (either amygdala or hippocampus onset earlier than both hippocampus and neocortex), or cortex early (either hippocampus or neocortex onset earlier than both amygdala and hippocampus). Next, for each term, we rank ordered the four brain regions based on age of onset. Ties were broken by assigning the average rank to all tied regions. For example, if a process was first active at E50 in amygdala and at E70 in the other three regions, then amygdala was assigned a rank of 1 while the other regions were assigned a rank of 3 (average of ranks 2, 3 and 4). This tie-breaking method highlighted terms that were earlier and later than all other regions. We calculated region ranks for 837 terms and plotted a heatmap of the proportion of terms with each rank for each region (columns sum to one in Extended Data Fig. 5c) and used these proportions to calculate the weighted mean rank of each region.

We selected the top 40 genes enriched in myelinating oligodendrocytes in adult mouse[63] and identified 23 orthologs with expression data in rhesus. Six genes (*MOG*, *MOBP*, *ERMN*, *MAL*, *ASPA* and *OPALIN*) were highly correlated (R> 0.9) with each other across all brain regions over development (Extended Data Fig. 5d). We calculated average expression of these genes (N= 3–4 subjects) in the highly myelinated regions that were sampled in this study.

Transcriptional signatures of V1 layers

We compared gene expression across postmitotic layers in V1 for each age starting from E80, the earliest age at which all layers were sampled. Samples from L2 and L3 were considered together as L2/3, and samples from sub-layers of L4 were considered together as L4. For each subject and age, we calculated the average log2 scaled gene expression of layers 2/3, 4, 5 and 6, as well as adjacent layers marginal zone/L1 and intermediate zone/white matter for comparison. For each age, we fit a linear model to the expression data with a dummy variable coding for layer. The proportion of variance explained by each layer was estimated using the CAR metric[64] in the “relaimpo” R package. For each layer and age, we rank ordered genes based on the proportion of variance explained (Supplementary Table 6).

Layer signatures between ages were compared by identifying the top 10% most laminar genes evenly allocated between layers (top 2.5% each in L2/3, L4, L5, and L6), and then calculating their average across-layer correlation between ages. Specifically, for each “source” list of marker genes, we found the average expression by layer of these genes at the “source” age and at every other “target” age, resulting in two expression tables of 1224 genes by 4 layers. Gene by gene, expression across layers was compared using Pearson correlation and the average correlation across all 1224 genes was determined.

Differential expression between the layer and all other layers was calculated using the "limma" R package. Genes were considered layer-enriched if they explained at least 25% of the expression variance in all layers (see description of analysis above) and had expression > 1.5-fold higher than other layers combined. Genes marking the same layer in at least two ages were further defined as "persistent" if they were laminar during at least 5 out of 7 ages, or "early", "middle", or "late" if their median age of laminar enrichment was <E120, between E120 and 3M, or >3M, respectively (Supplementary Table 7). Selected marker genes were visualized using heatmaps as described above, but only showing the subset of data considered in this analysis (L1-WM in V1 from E80-48M). To characterize the function of genes from each laminar timing group, we grouped genes from all layers with the same timing and performed gene ontology (GO) enrichment using all "Biological Process" terms with fewer than 500 associated genes. We identified significantly associated terms (nominal $P < 10^{-4}$ and five or more genes enriched for at least one timing group) and visualized–log10P values using a heatmap.

Differential expression between neocortical regions

We compared gene expression in V1 and ACG independently for each age in the outer (L2/L3) and inner (L5/L6) cortical layers, in the subventricular zone (SZ), and in the ventricular zone (VZ). First, we averaged gene expression (logarithmically scaled) within specimen across all relevant samples for each layer in V1 and ACG independently (i.e., L2 and L3 in V1 were averaged in each brain). For each age and layer, nominal P values for differential regional expression were calculated using a paired test (n = 3–4 monkeys), and genes with P <0.05 and average fold change >2 were considered significantly enriched in ACG or V1 (Supplementary Table 8). We then defined persistent regional markers as genes significantly enriched in either ACG or V1 in at least 2/3 of the ages in which a layer was present (N>=5 in L2/L3; N>=6 in L5/L6; N>=4 in SZ and VZ). We visualized the expression patterns of select regional markers using line plots, where each point shows the average (± standard deviation) log2-scaled gene expression levels of that gene in ACG (red) or V1 (blue) for a particular layer (specified to the right of the plot).

We quantified regional enrichment of several cell types in each layer and at each age using a hypergeometric test. Specifically, we compared 242 lists of marker genes derived from the literature against each set of V1 and ACG enriched genes described above, calculated a P value based on the significance of overlap using the *phyper* function in R. P values were Bonferroni corrected for multiple comparisons and visualized as boxes in a grid of layers (vertical axis) and ages (horizontal axis). Cell type lists were derived as follows: dividing cell markers were defined as any marker of any phase of cell cycle as presented in Supplemental Figure 3 of Bar-Joseph et al.[65]. Markers for excitatory neurons in adult mouse V1 were derived from Tasic et al.[66] by identifying all genes expressed in at least 2.2X as many excitatory as inhibitory cells (based on proportion) and that have expression in the 80th percentile excitatory cell that is both 2.2X higher than the 80th percentile inhibitory cell and also higher than RPKM=2.5. Inhibitory neuron markers were defined the same way, but swapping excitatory and inhibitory above. Finally, astrocyte marker genes were defined as genes which showed at least 10-fold enrichment in sorted astrocytes compared with both neurons and oligodendrocytes in Cahoy et al.[67]

To address the issue of timing of gene expression in prenatal cortical development, we extended these analysis to S1 by considering the set of genes showing rostral enrichment during the phase in which V1 corticogenesis lags ACG (E70-E90), and asked to what extent these genes also show higher expression in S1 vs. V1. We do this in two ways. First, we re-plot genes presented in Fig. 4 with S1 also shown for direct comparison. Second, we calculate the average enrichment of these genes in S1 relative to V1 (using log2 fold change across all samples from the relevant layers between E70-E90), and compare it directly to the average enrichment in ACG relative to V1 (Extended Data Fig. 6).

Network profiling of disease genes

We used weighted gene co-expression network analysis (WGCNA)[68,69] to generate 10 unbiased gene co-expression networks (one at each age) using cortical structures (ACG and V1) with the goal of quantifying the distribution of disease-related genes throughout these networks. Initial networks were generated using an automated strategy with the following function call in R:

```
blockRun = blockwiseModules(datExprRun, checkMissingData = TRUE,
 maxBlockSize = 17500,power = 14, networkType = "signed", deepSplit = 1,
 minModuleSize = 50, minCoreKMESize = 17, minKMEtoStay = 0.4,
 mergeCutHeight = 0.1, numericLabels = TRUE, verbose = 1)
```

where `datExprRun` is the log2 normalized microarray data from the top 9331 (75%) most variable genes (at each time point). For postnatal time points these values in V1 and ACG were scaled to the mean interquartile expression of samples from layers 2, 3, 5, and 6. The goal of this scaling was to deemphasize regional patterning differences or batch effects and maximize laminar and cell type patterning. After running the network (and after performing the merging steps described below) module eigengenes (ME; the first principal component of genes in the module) were calculated for each module, and all genes were reassigned to the module to which it is most highly correlated to the ME (referred to as the gene's module membership, or kME). Genes with kME<0.4 for all modules in a given age are left unassigned (defined as module 0). While there is no agreed upon value for a kME cutoff in the literature, we have successfully used a cutoff of kME < 0.4 in the past as the cutoff for assigned vs. unassigned genes (e.g. Hawrylycz et al. 2015[70]).

If the resulting network at any time point contained more than 15 modules, the module pair with the most highly correlated ME were iteratively merged until 15 modules remained. Since networks are unaffected by changes in labeling, modules were then re-ordered manually such that modules with strong enrichment for cortical plate neurons had low numbers, modules with strong enrichment for progenitor or glial cells had higher numbers, and modules between ages with a highly significant number of overlapping genes were ordered similarly across development. To do this third step, every pair of modules in networks from adjacent ages were compared using a hypergeometric test, and pairs with overlap significance of $P < 10^{-50}$ were linked (using grey lines in Fig. 5). We note that several additional pairs of modules show significant overlap (but with $P >= 10^{-50}$)—these

modules pairs are not shown in the plot for readability and because we are interesting in finding only the most significant overlaps. Finally, modules which were split into two modules at a single age but not at both adjacent ages were merged. At each age, modules are numbered by row number in Fig. 5a with the top row being Module 01 at each age. Module assignments for each gene at each age are provided in Supplementary Table 9.

We have broadly annotated many of these modules based on cortical layer, cell type, and disease enrichments, and provide the annotated network as a resource to the community. Module assignments of postmitotic vs. progenitor cell layer in prenatal neocortex was done by comparing mean expression in cortical plate (CP) and subplate (SP) vs. all other layers, and noting which mean value is highest. Similarly, assignments of cortical vs. glial cell layers in postnatal neocortex was done by comparing mean expression in post-mitotic layers 2–6 vs. white matter and layer 1. Module assignments based on cortical layer were performed by finding the average expression of all module genes in all layers included in them and finding the layer in which expression is highest (select layer enrichments are labeled in Fig. 5a). Additional assignments for primarily postmitotic cortical layers (from E80-48M) were found by performing a hypergeometric test on every module against layer markers identified in the analysis in Fig. 4b for each age. Significant enrichments ($P < 10^{-30}$) are color-coded in Fig. 5a. This p-value cutoff ensured that modules selected were only enriched with markers for a single layer, whereas with lower cutoffs some modules had markers for multiple layers. Module enrichments for astrocytes and oligodendrocytes were calculated by performing hypergeometric test comparisons, and plotting these P values as colored discs in Fig. 5a for modules that are highly significant ($P < 10^{-15}$). Astrocyte marker genes were defined as genes which showed at least 10-fold enrichment in sorted astrocytes compared with both neurons and oligodendrocytes in Cahoy et al.[67] (likewise for oligodendrocyte markers).

We used empirical rather than Bonferroni correction of the significance of disease gene enrichments because the tests were correlated due to high module gene overlap. Hypergeometric P values were calculated by comparing each module against genomics-based disease genes for autosomal recessive primary microcephaly (13 genes)[71], autism spectrum disorder (*a posteriori* probability > 0.8; 171 genes)[45] and schizophrenia (206 genes)[72] with homologs in this data set. Module assignments were then permuted 999 times per disease, and the minimum P value across all modules for each of the permuted comparisons was recorded. Nominal P values were corrected with this empirical null distribution of minimum P values. For example, a nominal P value less than the minimum P value in 90% of the permutations had an empirically corrected P value equal to 0.1 that controlled the family-wise error rate at 10%. This is a much more stringent criterion than FDR < 0.1 because it means that there is a less than 10% chance that any of the enriched modules is a false positive enrichment. Modules that were significantly enriched for disease-associated genes were plotted as red discs in Fig. 5b–d. For each disease, expression heatmaps for genes present in at least two enriched modules were plotted to show different developmental profiles.

A recent study of intellectual disability with and without brain malformations in consanguineous families identified 89 known and candidate disease related genes[73]. For

each pair of 71 genes profiled in this data set, we calculated the Pearson correlation of expression in all ACG and V1 cortical samples across development. We used hierarchical clustering of the Euclidean distance of correlations to group genes with similar spatiotemporal expression profiles into 4 major clusters and visualized correlations as a heatmap with the “pheatmap” R package. Genes were annotated with color bar tracks indicating phenotypes present in one or more individuals with likely pathological mutations in these genes. Primary microcephaly was defined as microcephaly without other brain malformations and corpus callosum abnormalities included agenesis and various hypogenesis and dysgenesis phenotypes. Multivariate distance matrix regression (MDMR) was used to test whether each phenotype explained a significant proportion of variation in the expression based distance matrix, and p-values were calculated based on a permutation test as previously described[74]. Expression heatmaps for a representative gene from each cluster were plotted to show different developmental profiles.

Conservation of developmental expression trajectories

Developmental expression data were downloaded for rat frontal cortex[75] and five human brain regions (medial frontal cortex, MFC; primary visual cortex, V1; hippocampus; amygdala; and striatum)[16] homologous to regions profiled in this study. To compare rhesus monkey to gross cortical samples in rat and human, we calculated average expression across all sub-regions of anterior cingulate gyrus (ACG, partially overlapping with MFC in human), V1, hippocampus, amygdala and striatum at each age. We aligned ages across species by estimating event scores based on the conserved timing of developmental events[1] and found that prenatal and postnatal development were well sampled in all species.

For frontal cortical samples, 4,125 orthologous genes were identified using NCBI Homologene, and we selected 2,062 genes that were among the top 50% most variable genes over development in rhesus. This filter removed genes that had relatively constant expression and therefore would be difficult to compare trajectories between species. Independently for each gene and species, we estimated the shape of expression trajectories by smoothed local linear regression (“loess” R function) with degree zero to allow for complex dynamics while reducing overfitting. With these smoothed fits, we estimated expression at 8 approximately evenly spaced event scores (ranging from 0.27 to 1.32) for each species. We correlated expression trajectories between each pair of species and plotted a Venn diagram showing the number of conserved (Pearson correlation $R > 0.5$) genes between each pair of species. Genes were assigned to one of five sets based on the pattern of species conservation. “Conserved” genes were conserved between at least two species pairs and “Not conserved” were not conserved between any species pair and may represent noise. “Primate-specific” genes were conserved between human and rhesus but not between rat and either primate. “Human-specific” genes were conserved between rhesus and rat but not between human and either other species. “Rhesus-specific” genes were conserved between human and rat but not between rhesus and either other species.

Next, we downloaded developmental expression profiles from an independent human data set (dorsolateral prefrontal cortex, dlPFC, GEO: GSE30272)[14]. For 1,848 of 2,062 variable genes with dlPFC data, we calculated Pearson correlations of expression trajectories

between human MFC and dlPFC as described above. Expression correlations between all pairs of species were compared with box plots (median ± 25th and 75th percentiles, whiskers at 1.5 times the inter-quartile range, and outliers plotted as points). Correlation distributions deviated significantly from normality based on a Shapiro-Wilk test ("shapiro.test" in R), and so differences were tested with a Kruskal-Wallis rank sum test ("kruskal.test") followed by paired Wilcoxon signed rank tests ("wilcox.test") to test for significant differences in conservation between species. Finally, we repeated the analysis described above for a larger set of 8,683 orthologous genes in human and rhesus macaque. First, we rank ordered genes based on the standard deviation of expression across rhesus monkey development. For the top 100 most variable genes, we calculated expression correlations between human data sets and between human and rhesus. We repeated this analysis 86 times, each time including the next set of 100 most variable genes. Finally, we plotted the median correlation within and between species as a function of developmental expression variability.

We downloaded expression data (average ISH density) for 2,024 genes in the dorsal pallium (which includes frontal cortex) of developing mouse[76] (http://developingmouse.brain-map.org/) using the Allen Brain Atlas Application Programming Interface (API) and calculated Spearman correlations between 305 mouse and rat orthologous genes. All pair-wise species correlations are included in Supplementary Table 10.

We used DAVID[77,78] v6.7 (https://david.ncifcrf.gov/home.jsp) to test different sets of conserved genes for significant enrichment in gene ontology categories, tissue specificity and disease association relative to a background set of all 1,848 variable genes. Gene paralogs were downloaded from Ensembl with BioMart[79]. We checked genes for evidence of positive selection on different evolutionary branches using Selectome[80,81] (http://selectome.unil.ch/).

Developmental breakpoint analysis

For 4,125 orthologous genes in homologous brain regions of human, rhesus monkey and rat, we fit expression trends (log-transformed intensity versus age in post-conceptional days) using segmented linear regression ("segmented" R package) with a single breakpoint and two slopes. Breakpoints were initialized at the mid-point of development, and standard errors for breakpoint and slope estimates were calculated using 20 bootstrap iterations (Supplementary Table 11). 95% confidence intervals (CI) were calculated for slopes based on bootstrapped standard errors (i.e. ± 1.96 × SE). Genes with a good model fit ($R^2 > 0.8$) and significantly increasing (95% CI did not include zero) or decreasing initial slopes in all three species were included for further analysis. Breakpoint densities of $\log_{10}$(Age) in frontal and primary visual cortical areas were plotted for different species for all genes and for synapse related genes downloaded from SynaptomeDB[82] (http://metamoodics.org/SynaptomeDB). For each brain region, species pairwise comparisons of breakpoints and standard errors were plotted and genes with significantly increasing or decreasing second slopes (95% CI did not include zero) were highlighted (Fig. 6e and Extended Data Fig. 7d,e). For each pair of species, ages corresponding to equivalent (i.e. equal event scores) or nearby developmental stages (event scores within ± 0.2) were plotted. Genes with significantly different breakpoints between species (95% CI of breakpoints did not overlap)

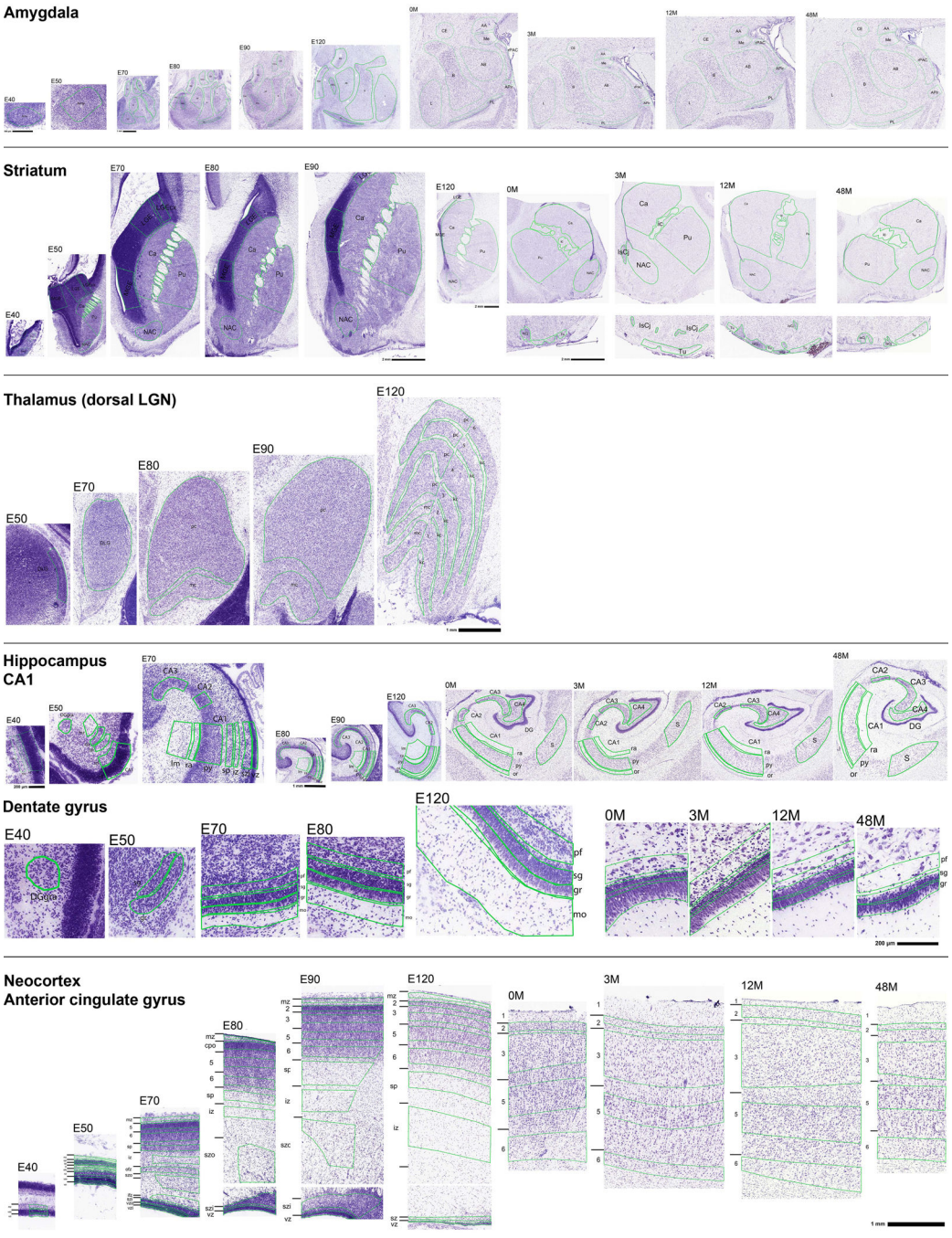
were counted and summarized for each brain region as a stacked bar plot indicating the expression slope after human breakpoints. Breakpoint timing was also compared between frontal and primary visual areas in rhesus and human with a paired Wilcoxon signed rank test and plotted (Extended Data Fig. 7c). In total, expression breakpoints were identified for 179 significantly increasing and 179 decreasing genes in human and rhesus monkey V1 and ACG.

Synaptic density measurements were obtained from previously published studies for V1 in rat[83], and V1 and prefrontal cortex in human[41] and rhesus[40]. Segmented linear regression was fit to synaptic density trends (% max synaptic density versus age in post-conceptional days), and breakpoints corresponded to ages of peak density (Extended Data Fig. 7b). Breakpoint 95% CIs were calculated and plotted as shaded bands in density plots.

Code availability

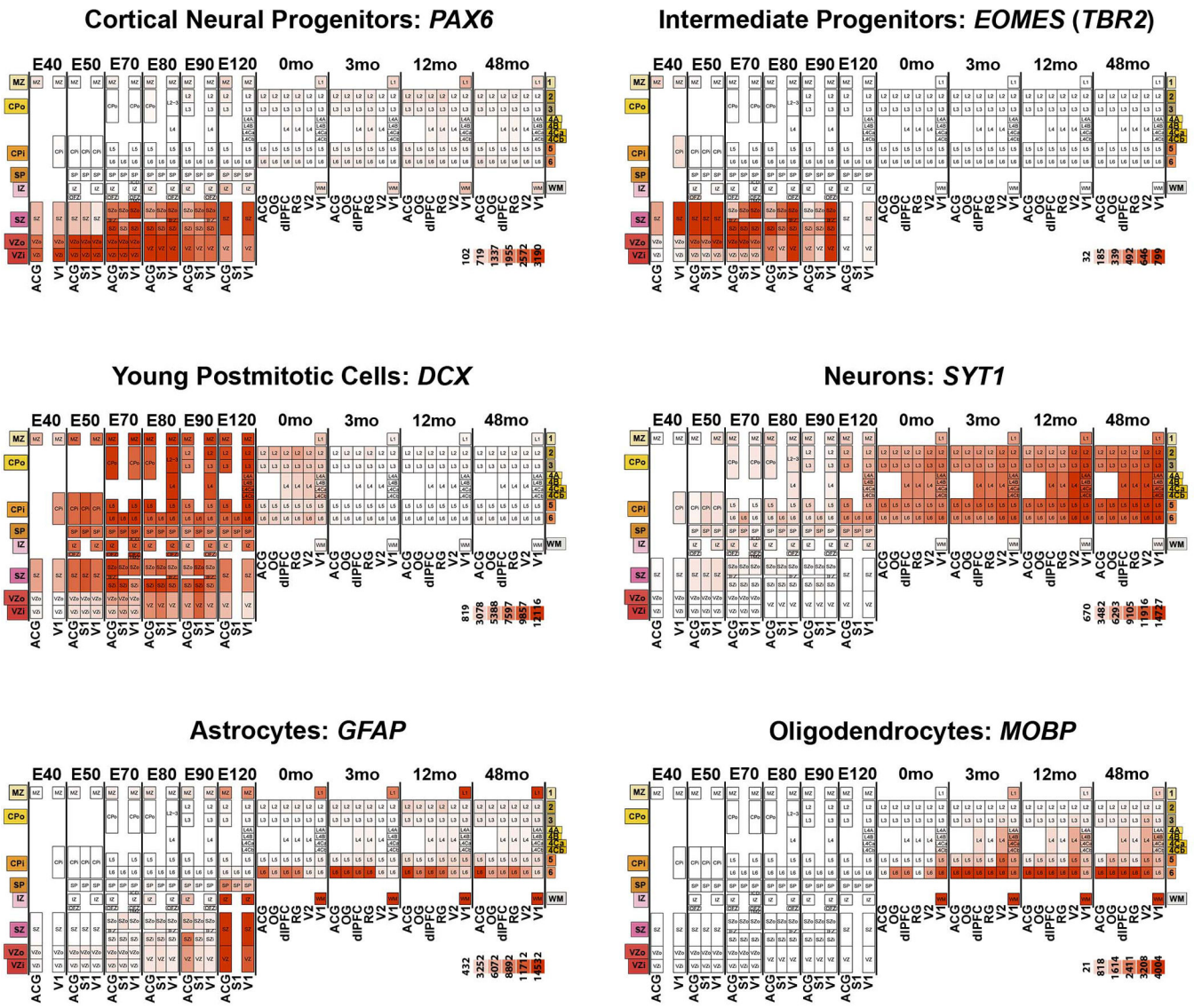
All results and figures from this study can be reproduced using R code and data files available from GitHub (https://github.com/AllenBrainAtlas/DevRhesusLMD). The software is free and can be modified and redistributed under the terms of the GNU General Public License, Version 3 (GPLv3).

Extended Data



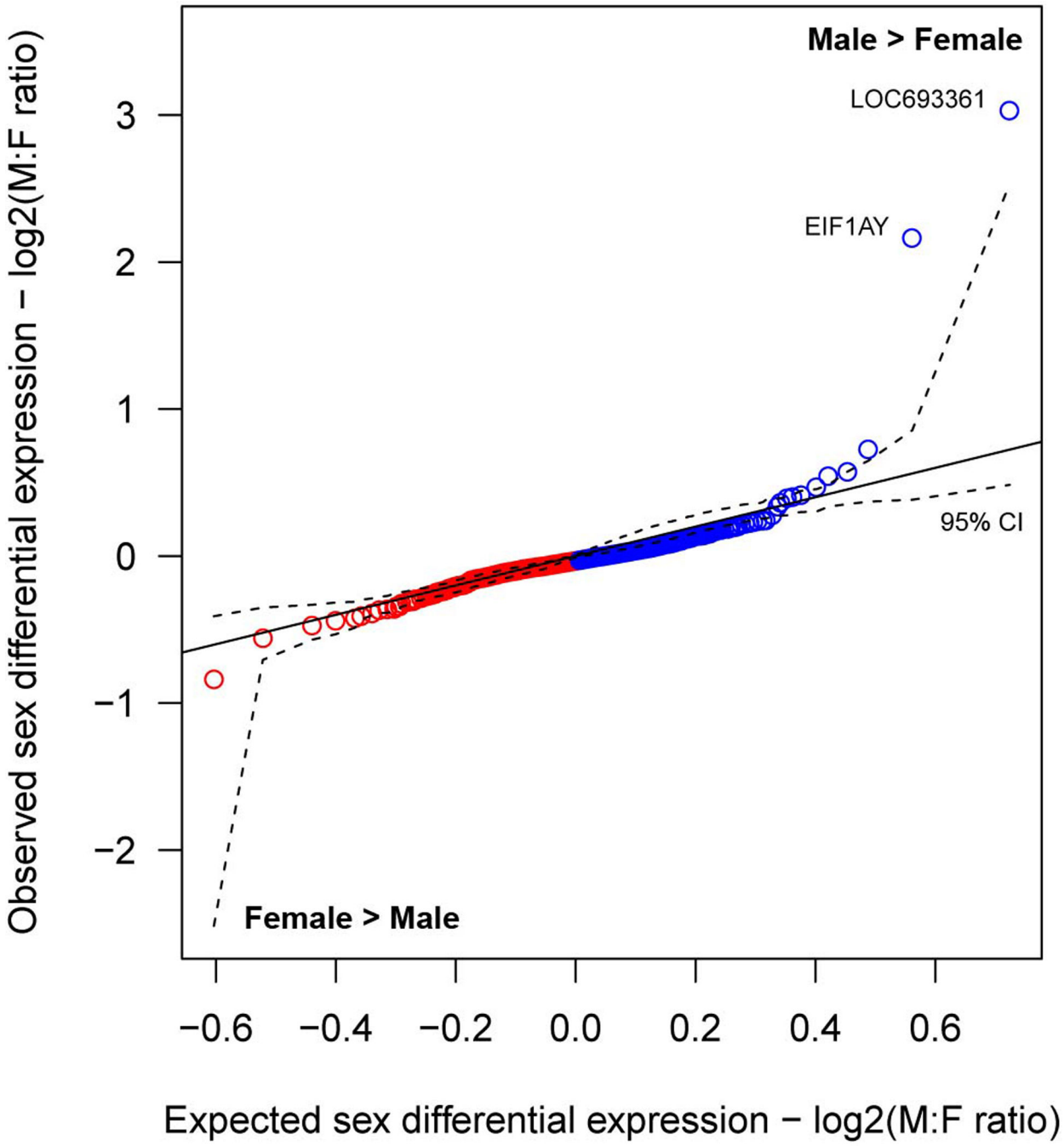
Extended Data Figure 1. Anatomical parcellations of developing cortical and subcortical regions

Nissl stained sections of major brain regions sampled in this resource. Green lines demarcate subregions that were isolated by laser capture microdissection and transcriptionally profiled.



Extended Data Figure 2. Canonical cell type marker gene expression across cortical development

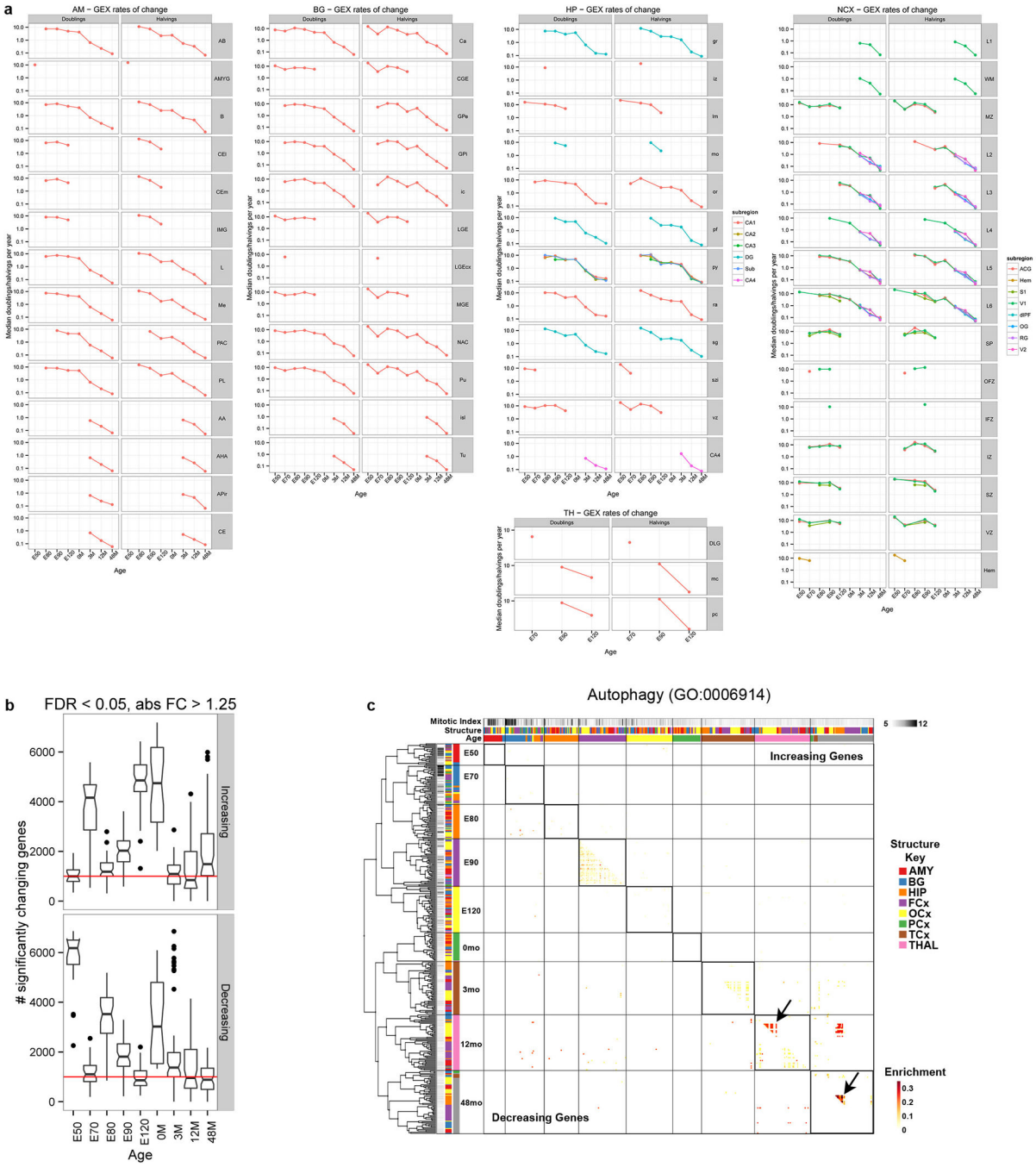
Heatmaps of average gene expression in cortical layers at different prenatal (E40 – E120) and postnatal ages (0 – 48 months). Anterior cingulate gyrus (ACG) and primary visual cortex (V1) were sampled at all ages, while primary somatosensory cortex (S1) was sampled in a limited set of layers prenatally. Several additional prefrontal and visual areas were sampled postnatally: orbital gyrus (OG), dorsolateral prefrontal cortex (dlPFC), rectal gyrus (RG) and secondary visual cortex (V2).



Extended Data Figure 3. Sex differential expression is limited to two Y chromosome genes

a, Quantile-quantile plot of observed versus expected sex differential expression for all 12,441 genes across brain regions during prenatal development. A linear mixed model was fit to all prenatal brain samples with a fixed effect for sex and random effects for brain region, age and donor (see Supplementary Table 11). Genes were ordered by the observed sex effect and plotted versus the expected sex effect based on permutation testing. A 95% confidence interval was calculated (dashed line) based on permutations, and 11 genes in males and no genes in females were more highly expressed than expected by chance. Seven

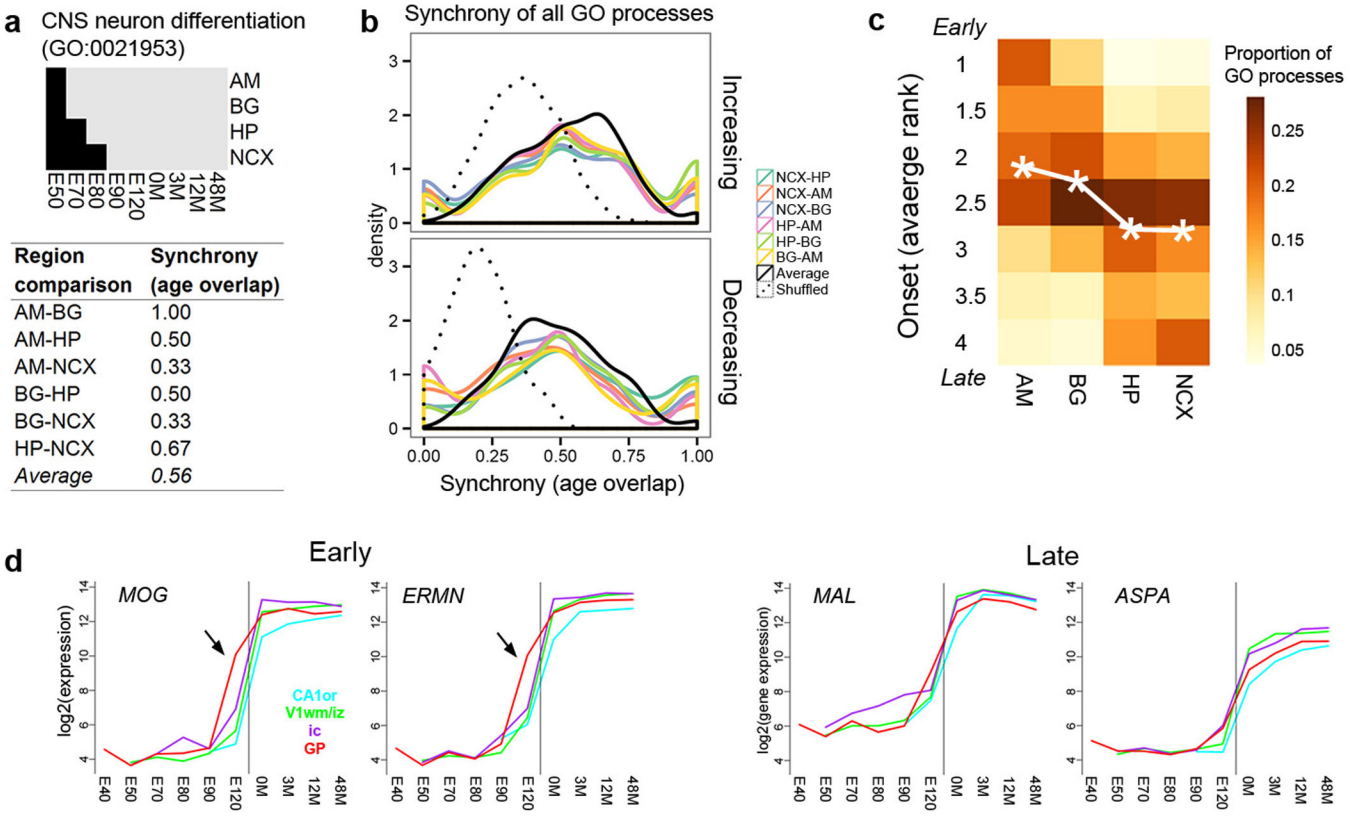
of these 11 genes were nominally significant and included at least two Y-chromosome genes (*EIF1AY*, *LOC720563*) and potentially a third gene (*LOC693361*) whose microarray probe maps to an unannotated region of the Y chromosome.



Extended Data Figure 4. Expression rates of change have similar developmental trajectories across all brain regions

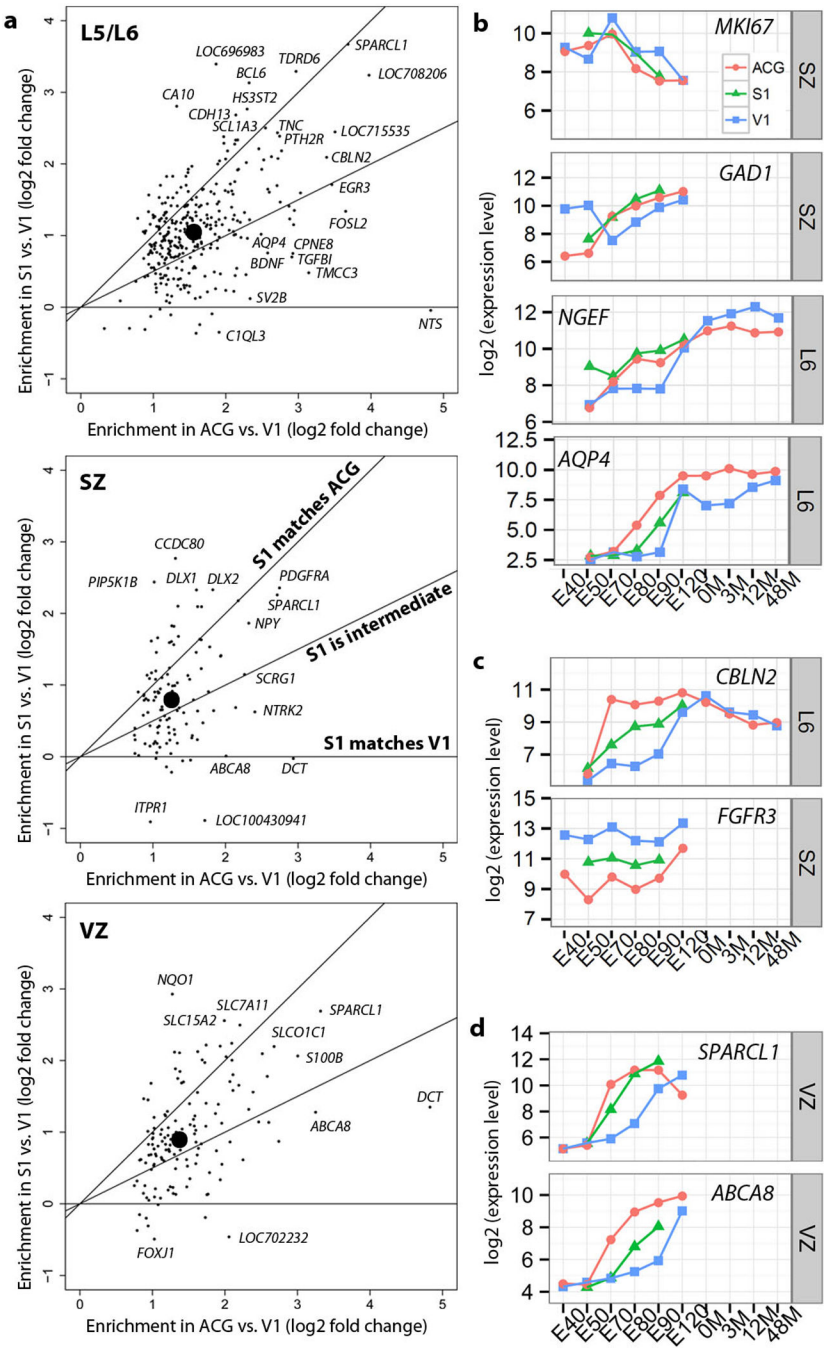
a, Rates of expression change in all available brain subregions and ages. **b**, Boxplots summarizing the number of significantly increasing or decreasing genes between adjacent time points in all subregions. At all ages, the majority of subregions had at least 1000 genes

(red line) that were significantly changing. **c**, Regional specificity of increased autophagy may reflect differential timing of synaptic pruning. Enrichment for autophagy of the most dynamically increasing (upper triangle) and decreasing (lower triangle) genes with samples ordered and labeled as in Fig. 3b. For each gene list, the color corresponds to the proportion of GO terms that are more specific (i.e. “child”) terms subsumed under autophagy (GO: 0006914) based on the GO hierarchy and that are significantly enriched (nominal $P < 0.05$). Note that autophagy was selectively turned on in occipital cortex after infancy and in hippocampus after juvenility (arrows).



Extended Data Figure 5. Variable synchrony of biological processes between brain regions

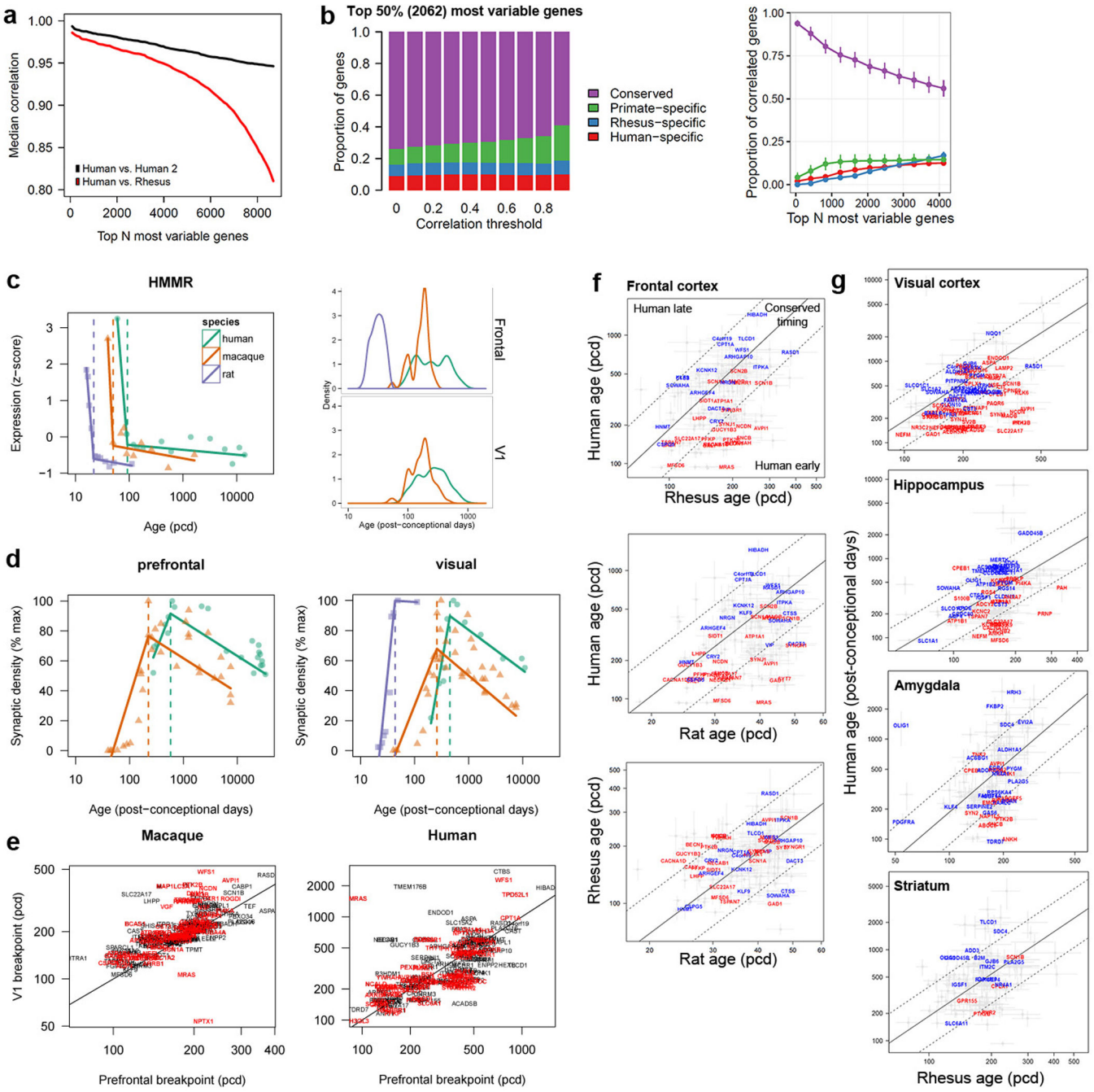
a, Example of variable timing of GO process activity (black boxes) between regions, resulting in different age overlaps (table below). Note that E50 was the earliest age for which we could calculate expression change. AM, amygdala; BG, basal ganglia; HP, hippocampus; NCX, neocortex. **b**, Average pairwise age overlaps (black, solid line) for all increasing (top) and decreasing (bottom) GO processes were greater than expected by chance (black, dotted line). **c**, Rank ordered timing of GO processes in Fig. 4c with weighted average rank for each region (asterisks). **d**, Developmental expression of mature oligodendrocyte markers in four myelin-enriched brain subregions with early increased expression in globus pallidus (arrows) for *MOG* and *ERMN* but not *MAL* and *ASPA*.



Extended Data Figure 6. Neuro- and glio-genesis in S1 occur at a time course intermediate between V1 and ACG

a, Genes with enriched expression in ACG relative to V1 between E70-E90 also show enriched expression in S1 relative to V1, suggesting that the timing of primary sensory regions is non-uniform in L5/L6 (top), SZ (middle), and VZ (bottom). Each plot shows the average enrichment (log2 fold change) in S1 vs. V1 (y-axes) compared with the average enrichment in ACG vs. V1 (x-axes) between E70-E90 for all genes significantly enriched in ACG in at least two of the three ages between E70-E90 (Fig. 5a). **b**, Marker genes for cell

types (compare with Fig. 5e–h) show expression patterns in S1 which are either consistent with ACG (i.e., *GAD1*) or intermediate between V1 and ACG (i.e., AQP4). **c**, Genes with enriched expression for V1 (*FGFR3*) or ACG (*CBLN2*) across development show intermediate expression in S1, suggesting that these genes may show cortical gradients rather than specific expression in V1 or ACG. **d**, Genes with common expression patterning in V1 and ACG can show different patterns in S1, suggesting that rate of neuron and glia development is not the whole story. For example, *SPARCL1* and *ABCA8* both show increased expression with time in VZ with V1 showing a delay relative to ACG; however, these two genes show different temporal delays in S1.



Extended Data Figure 7. Evolutionary conservation of developmental expression

a, Median pairwise correlations of expression trajectories in prefrontal cortex within human (black) and between human and rhesus monkey (red) decrease for less variable genes in rhesus monkey (genes ordered by standard deviation of expression across ages). **b**, Left: Proportion of genes assigned to different conservation categories is robust to correlation threshold. Right: Developmentally dynamic genes are more highly conserved (genes ordered by same method as in **a**). For each set of genes, the average ± standard deviation of the proportion of genes in each conservation category was estimated using correlation thresholds

ranging from 0 to 0.9. **c**, Left: Segmented linear fits of expression for example gene with estimated breakpoints in each species (dashed lines). Right: Distribution of breakpoint ages for 179 decreasing genes with good fits to the model in frontal and primary visual cortex. Colors and symbols are consistent in **c-d. d**, Segmented linear fits with breakpoint estimation of synaptic density for prefrontal and primary visual cortex based on previously published studies (see **Methods**). **e**, Breakpoint comparison of 179 increasing genes including 81 synapse related genes (red) between cortical areas within species. Genes that fall on the lines peak at the same age in primary visual and prefrontal cortex. **f-g**, Comparison of breakpoint timing between human, rhesus monkey and rat in **f**, prefrontal cortex and **g**, additional brain regions. Genes that plateau in expression after their breakpoint in human (grey points), and genes that significantly decrease (blue symbols) or increase (red symbols) expression with 95% confidence intervals (grey lines) of breakpoints. Black lines correspond to equal (solid) ± a window (dashed) of developmental ages between species. pcd, post-conceptional days.

Extended Data Table 1

Rhesus monkey donor information.

Animal ID	Age	SX date	Sex	Indian origin (proportion)	Chinese origin (proportion)
MMU37857	E40	2/9/2012	Female	15/16	1/16
MMU38481	E40	5/2/2012	Male	1	0
MMU31370	E40	3/26/2013	Male	1	0
MMU36477	E40	4/16/2012	Female	1	0
DAM37624	E50	1/13/2012	Male	5/8	3/8
MMU36725	E50	4/23/2012	Male	1	0
MMU30213	E50	3/6/2012	Male	1	0
MMU38099	E50	2/23/2012	Female	15/16	1/16
DAM35650	E70	1/23/2012	Female	1	0
MMU32143	E70	3/9/2012	Female	1	0
MMU33757	E70	3/5/2012	Male	15/16	1/16
MMU35475	E70	3/5/2012	Male	7/8	1/8
MMU35058	E80	2/2/2012	Female	15/16	1/16
MMU37852	E80	3/29/2012	Female	1	0
MMU34624	E80	4/19/2012	Male	3/4	1/4
MMU36435	E80	4/30/2012	Male	15/16	1/16
MMU34326	E90	4/16/2012	Male	1	0
MMU35674	E90	4/13/2012	Female	1	0
MMU38285	E90	4/13/2012	Female	13/16	3/16
MMU38521	E90	4/27/2012	Male	1	0
MMU27052	E120	3/23/2012	Male	1	0
MMU30021	E120	3/7/2012	Female	1	0
MMU34501	E120	7/30/2012	Male	3/4	1/4
MMU29380	E120	3/12/2012	Female	1	0

Animal ID	Age	SX date	Sex	Indian origin (proportion)	Chinese origin (proportion)
MMU39867	OM	NA	Male	NA	NA
MMU39893	OM	4/12/2013	Male	3/4	1/4
MMU40864	OM	4/21/2014	Male	1	0
MMU40944	OM	4/29/2014	Male	1	0
MMU39465	3M	11/13/2008	Male	1	0
MMU39538	3M	5/10/2013	Male	1	0
MMU39766	3M	7/15/2013	Male	7/8	1/8
MMU40525	3M	6/9/2014	Male	1	0
MMU39420	12M	6/26/2013	Male	1	0
MMU39620	12M	3/9/2014	Male	7/8	1/8
MMU39645	12M	3/9/2014	Male	7/8	1/8
MMU36322	48M	7/11/2013	Male	1	0
MMU36358	48M	7/14/2013	Male	5/8	3/8
MMU36468	48M	6/26/2013	Male	5/8	3/8

Subject information for 38 rhesus macaque monkeys included in study. Surgery (SX) date, age, sex, and ancestry based on origin of parents and grandparents.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Trygve E. Bakken[1,*], Jeremy A. Miller[1,*], Song-Lin Ding[1,*], Susan M. Sunkin[1], Kimberly A. Smith[1], Lydia Ng[1], Aaron Szafer[1], Rachel A. Dalley[1], Joshua J. Royall[1], Tracy Lemon[1], Sheila Shapouri[1], Kaylynn Aiona[1], James Arnold[1], Jeffrey L. Bennett[2], Darren Bertagnolli[1], Kristopher Bickley[1], Andrew Boe[1], Krissy Brouner[1], Stephanie Butler[1], Emi Byrnes[1], Shiella Caldejon[1], Anita Carey[1], Shelby Cate[1], Mike Chapin[1], Jefferey Chen[1], Nick Dee[1], Tsega Desta[1], Tim A. Dolbeare[1], Nadia Dotson[1], Amanda Ebbert[1], Erich Fulfs[1], Garrett Gee[1], Terri L. Gilbert[1], Jeff Goldy[1], Lindsey Gourley[1], Ben Gregor[1], Guangyu Gu[1], Jon Hall[1], Zeb Haradon[1], David R. Haynor[3], Nika Hejazinia[1], Anna Hoerder-Suabedissen[4], Robert Howard[1], Jay Jochim[1], Marty Kinnunen[1], Ali Kriedberg[1], Chihchau L. Kuan[1], Christopher Lau[1], Chang-Kyu Lee[1], Felix Lee[1], Lon Luong[1], Naveed Mastan[1], Ryan May[1], Jose Melchor[1], Nerick Mosqueda[1], Erika Mott[1], Kiet Ngo[1], Julie Nyhus[1], Aaron Oldre[1], Eric Olson[1], Jody Parente[1], Patrick D. Parker[1], Sheana Parry[1], Julie Pendergraft[1], Lydia Potekhina[1], Melissa Reding[1], Zackery L. Riley[1], Tyson Roberts[1], Brandon Rogers[1], Kate Roll[1], David Rosen[1], David Sandman[1], Melaine Sarreal[1], Nadiya Shapovalova[1], Shu Shi[1], Nathan Sjoquist[1], Andy J. Sodt[1], Robbie Townsend[1], Lissette Velasquez[1], Udi Wagley[1], Wayne B. Wakeman[1], Cassandra White[1], Crissa Bennett[1], Jennifer Wu[1], Rob Young[1], Brian L. Youngstrom[1], Paul Wohnoutka[1], Richard A. Gibbs[5], Jeffrey Rogers[5], John G. Hohmann[1], Michael J. Hawrylycz[1],

Robert F. Hevner[6], Zoltán Molnár[4], John W. Phillips[1], Chinh Dang[1], Allan R. Jones[1], David G. Amaral[2], Amy Bernard[1], and Ed S. Lein[1]

Affiliations

[1]Allen Institute for Brain Science, Seattle, Washington 98109, USA

[2]Department of Psychiatry and Behavioral Science, California National Primate Research Center, The M.I.N.D. Institute, University of California, Davis, Sacramento, CA 95817, USA

[3]Department of Radiology, University of Washington, Seattle, Washington 98195, USA

[4]Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road Oxford OX1 3QX, UK

[5]Human Genome Sequencing Center and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

[6]Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, Washington 98101, USA

Acknowledgments

The authors thank the Allen Institute for Brain Science founders, Paul G. Allen and Jody Allen, for their vision, encouragement, and support. The authors also thank Dr. Juan Montiel and Dr. Wei Zhi Wang for their advice on developmental neuroanatomy and experimental design. We also wish to acknowledge the California National Primate Research Center (NIH Award Number RR00169) for providing tissues and Covance Genomics Laboratory (Seattle, WA) for microarray probe generation, hybridization and scanning. The project was supported by NIH Blueprint for Neuroscience Research contract HHSN-271-2008-0047 (PI: Ed Lein, Allen Institute for Brain Science) from the National Institute of Mental Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or the National Institute of Mental Health.

References

1. Workman AD, Charvet CJ, Clancy B, Darlington RB, Finlay BL. Modeling Transformations of Neurodevelopmental Sequences across Mammalian Species. J. Neurosci. 2013; 33:7368–7383. [PubMed: 23616543]
2. Smart IHM, Dehay C, Giroud P, Berland M, Kennedy H. Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. Cereb. Cortex. 2002; 12:37–53. [PubMed: 11734531]
3. MacLean EL, et al. The evolution of self-control. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:E2140–E2148. [PubMed: 24753565]
4. Caldwell, CA., Whiten, Ain. Primates Perspect. Oxford University Press; 2011. p. 652-662.
5. Belmonte JCICI, et al. Brains, Genes, and Primates. Neuron. 2015; 86:617–631. [PubMed: 25950631]
6. Miller DJ, et al. Prolonged myelination in human neocortical evolution. Proc. Natl. Acad. Sci. U. S. A. 2012; 109:16480–16485. [PubMed: 23012402]
7. Rakic P, Bourgeois JP, Eckenhoff MF, Zecevic N, Goldman-Rakic PS. Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. Science. 1986; 232:232–235. [PubMed: 3952506]
8. Petanjek Z, et al. Extraordinary neoteny of synaptic spines in the human prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:13281–13286. [PubMed: 21788513]

9. Gibbs RA, et al. Evolutionary and biomedical insights from the rhesus macaque genome. Science. 2007; 316:222–234. [PubMed: 17431167]
10. Bernard A, et al. Transcriptional Architecture of the Primate Neocortex. Neuron. 2012; 73:1083–1099. [PubMed: 22445337]
11. Zeng H, et al. Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. Cell. 2012; 149:483–496. [PubMed: 22500809]
12. Somel M, et al. Transcriptional neoteny in the human brain. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:5743. [PubMed: 19307592]
13. Liu X, et al. Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. Genome Res. 2012; 22:611–622. [PubMed: 22300767]
14. Colantuoni C, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature. 2011; 478:519–523. [PubMed: 22031444]
15. Hawrylycz MJ, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012; 489:391–399. [PubMed: 22996553]
16. Kang HJ, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011; 478:483–489. [PubMed: 22031440]
17. Miller JA, et al. Transcriptional landscape of the prenatal human brain. Nature. 2014; 508:199–206. [PubMed: 24695229]
18. Bakken TE, et al. Spatiotemporal dynamics of the postnatal developing primate brain transcriptome. Hum. Mol. Genet. 2015; 24:4327–4339. [PubMed: 25954031]
19. Molyneaux BJ, et al. DeCoN: Genome-wide Analysis of In Vivo Transcriptional Dynamics during Pyramidal Neuron Fate Selection in Neocortex. Neuron. 2015; 85:275–288. [PubMed: 25556833]
20. Zeisel A, et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science. 2015; 347:1138–1142. [PubMed: 25700174]
21. Darmanis S, et al. A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. 2015; 112:7285–7290. [PubMed: 26060301]
22. Rakic P. Neurons in rhesus monkey visual cortex: systematic relation between time of origin and eventual disposition. Science. 1974; 183:425–427. [PubMed: 4203022]
23. Hughes JF, et al. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature. 2012; 483:82–86. [PubMed: 22367542]
24. Elston GN, Oga T, Fujita I. Spinogenesis and pruning scales across functional hierarchies. J. Neurosci. 2009; 29:3271–3275. [PubMed: 19279264]
25. Scott, Ja, et al. Longitudinal analysis of the developing rhesus monkey brain using magnetic resonance imaging: birth to adulthood. Brain Struct. Funct. 2015 doi:10.1007/s00429-015-1076-x.
26. Hasegawa M, et al. Development of myelination in the human fetal and infant cerebrum: a myelin basic protein immunohistochemical study. Brain Dev. 1992; 14:1–6. [PubMed: 1375444]
27. Sorensen SA, et al. Correlated Gene Expression and Target Specificity Demonstrate Excitatory Projection Neuron Diversity. Cereb. Cortex. 2015; 25:433–449. [PubMed: 24014670]
28. Pletikos M, et al. Temporal specification and bilaterality of human neocortical topographic gene expression. Neuron. 2014; 81:321–332. [PubMed: 24373884]
29. O'Leary DDM, Chou S-JJ, Sahara S. Area patterning of the mammalian cortex. Neuron. 2007; 56:252–269. [PubMed: 17964244]
30. De la Rossa A, et al. In vivo reprogramming of circuit connectivity in postmitotic neocortical neurons. Nat. Neurosci. 2013; 16:193–200. [PubMed: 23292682]
31. Rakic P. Prenatal development of the visual system in rhesus monkey. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 1977; 278:245–260. [PubMed: 19781]
32. Tropea D, et al. Gene expression changes and molecular pathways mediating activity-dependent plasticity in visual cortex. Nat. Neurosci. 2006; 9:660–668. [PubMed: 16633343]
33. Willsey AJ, et al. Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism. Cell. 2013; 155:997–1007. [PubMed: 24267886]
34. Parikshak NN, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell. 2013; 155:1008–1021. [PubMed: 24267887]

35. Bishop KM, Goudreau G, O'Leary DD. Regulation of area identity in the mammalian neocortex by Emx2 and Pax6. Science. 2000; 288:344–349. [PubMed: 10764649]
36. Schoenebeck JJ, et al. Variation of BMP3 Contributes to Dog Breed Skull Diversity. PLoS Genet. 2012; 8:1–11.
37. Matsumoto A, et al. LIN7A depletion disrupts cerebral cortex development, contributing to intellectual disability in 12q21-deletion syndrome. PLoS One. 2014; 9:3–10.
38. Bianchi S, et al. Synaptogenesis and development of pyramidal neuron dendritic morphology in the chimpanzee neocortex resembles humans. Proc. Natl. Acad. Sci. 2013; 110:10395–10401. [PubMed: 23754422]
39. Huttenlocher PR. Synaptic density in human frontal cortex - developmental changes and effects of aging. Brain Res. 1979; 163:195–205. [PubMed: 427544]
40. Rakic P, Bourgeois JP, Goldman-Rakic PS. Synaptic development of the cerebral cortex: implications for learning, memory, and mental illness. Prog. Brain Res. 1994; 102:227–243. [PubMed: 7800815]
41. Huttenlocher PR, Dabholkar AS. Regional differences in synaptogenesis in human cerebral cortex. J. Comp. Neurol. 1997; 387:167–178. [PubMed: 9336221]
42. Jackman SL, Turecek J, Belinsky JE, Regehr WG. The calcium sensor synaptotagmin 7 is required for synaptic facilitation. Nature. 2016; 529:88–91. [PubMed: 26738595]
43. Pouchelon G, et al. Modality-specific thalamocortical inputs instruct the identity of postsynaptic L4 neurons. Nature. 2014; 511:471–474. [PubMed: 24828045]
44. De Marco García NV, Karayannis T, Fishell G. Neuronal activity is required for the development of specific cortical interneuron subtypes. Nature. 2011; 472:351–355. [PubMed: 21460837]
45. Iossifov I, et al. Low load for disruptive mutations in autism genes and their biased transmission. Proc. Natl. Acad. Sci. 2015; 112:E5600–E5607. [PubMed: 26401017]
46. de Lacy N, King BH. Revisiting the Relationship Between Autism and Schizophrenia: Toward an Integrated Neurobiology. Annu. Rev. Clin. Psychol. 2013; 9:555–587. [PubMed: 23537488]
47. Stein JJL, et al. A quantitative framework to evaluate modeling of cortical development by neural stem cells. Neuron. 2014; 83:69–86. [PubMed: 24991955]
48. Cusanovich DA, et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science. 2015; 348:910–914. [PubMed: 25953818]
49. Lein ES, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature. 2007; 445:168–176. [PubMed: 17151600]
50. Bystron I, Blakemore C, Rakic P. Development of the human cerebral cortex: Boulder Committee revisited. Nat. Rev. Neurosci. 2008; 9:110–122. [PubMed: 18209730]
51. Price, J., Russchen, F., Amaral, Din. Handb. Chem. Neuroanat. Hokfelt, B., Swanson, L., editors. Elsevier; 1987. p. 279-381.
52. Berger B, Alvarez C. Neurochemical development of the hippocampal region in the fetal rhesus monkey. III: Calbindin-D28K, calretinin and parvalbumin with special mention of Cajal-Retzius cells and the retrosplenial cortex. J. Comp. Neurol. 1996; 366:674–699. [PubMed: 8833116]
53. Paxinos, G. The Rhesus Monkey Brain in Stereotaxic Coordinates. Elsevier/Academic Press; 2009.
54. Karnovsky MJ, Roots L. A 'direct-coloring' thiocholine method for cholinesterases. J. Histochem. Cytochem. 1964; 12:219–221. [PubMed: 14187330]
55. Winrow CJ, et al. Refined anatomical isolation of functional sleep circuits exhibits distinctive regional and circadian gene transcriptional profiles. Brain Res. 2009; 1271:1–17. [PubMed: 19302983]
56. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003; 19:185–193. [PubMed: 12538238]
57. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007; 8:118–127. [PubMed: 16632515]
58. Miller, Ja, et al. Improving reliability and absolute quantification of human brain microarray data by filtering and scaling probes using RNA-Seq. BMC Genomics. 2014; 15:154. [PubMed: 24564186]

59. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010; 11:R86. [PubMed: 20738864]
60. Blankenberg D, et al. Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. Chapter. 2010; 19 Unit 19.10.1–21.
61. Giardine B, et al. Galaxy: A platform for interactive large-scale genome analysis. Genome Res. 2005; 15:1451–1455. [PubMed: 16169926]
62. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J. Mol. Biol. 1990; 215:403–410. [PubMed: 2231712]
63. Zhang Y, et al. An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. J. Neurosci. 2014; 34:11929–11947. [PubMed: 25186741]
64. Zuber V, Strimmer K. Variable importance and model selection by decorrelation. Statistics. 2010; 10:1–22.
65. Bar-Joseph Z, et al. Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. Proc. Natl. Acad. Sci. U. S. A. 2008; 105:955–960. [PubMed: 18195366]
66. Tasic B, et al. Adult cortical cell taxonomy by single cell transcriptomics. Nat. Neurosci. 2016
67. Cahoy J, et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J. Neurosci. 2008; 28:264–278. [PubMed: 18171944]
68. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. [PubMed: 19114008]
69. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. J. Stat. Softw. 2012; 46
70. Hawrylycz M, et al. Canonical genetic signatures of the adult human brain. Nat. Neurosci. 2015; 18:1832–1844. [PubMed: 26571460]
71. Verloes, A., Drunat, S., Gressens, P., Passemard, Sin. GeneReviews® [Internet]. Pagon, RA., et al., editors. University of Washington, Seattle; 2013. at <http://www.ncbi.nlm.nih.gov/books/NBK9587/>
72. Ripke S, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511:421–427. [PubMed: 25056061]
73. Karaca E, et al. Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. Neuron. 2015; 88:499–513. [PubMed: 26539891]
74. Zapala, Ma, Schork, NJ. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. Proc. Natl. Acad. Sci. 2006; 103:19430–19435. [PubMed: 17146048]
75. Stead JDH, et al. Transcriptional profiling of the developing rat brain reveals that the most dramatic regional differentiation in gene expression occurs postpartum. J. Neurosci. 2006; 26:345–353. [PubMed: 16399705]
76. Thompson CL, et al. A High-Resolution Spatiotemporal Atlas of Gene Expression of the Developing Mouse Brain. Neuron. 2014; 83:309–323. [PubMed: 24952961]
77. Huang DW, Sherman BT, Lempicki Ra. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 2009; 4:44–57. [PubMed: 19131956]
78. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37:1–13. [PubMed: 19033363]
79. Smedley D, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. 2015; 43:W589–W598. [PubMed: 25897122]
80. Proux E, Studer RA, Moretti S, Robinson-Rechavi M. Selectome: a database of positive selection. Nucleic Acids Res. 2009; 37:D404–D407. [PubMed: 18957445]
81. Moretti S, et al. Selectome update: quality control and computational improvements to a database of positive selection. Nucleic Acids Res. 2014; 42:D917–D921. [PubMed: 24225318]

82. Pirooznia M, et al. SynaptomeDB: an ontology-based knowledgebase for synaptic genes. Bioinformatics. 2012; 28:897–899. [PubMed: 22285564]

83. Blue ME, Parnavelas JG. The formation maturation of synapses in the visual cortex of the rat. II. Quantitative analysis. J. Neurocytol. 1983; 12:697–712. [PubMed: 6619907]

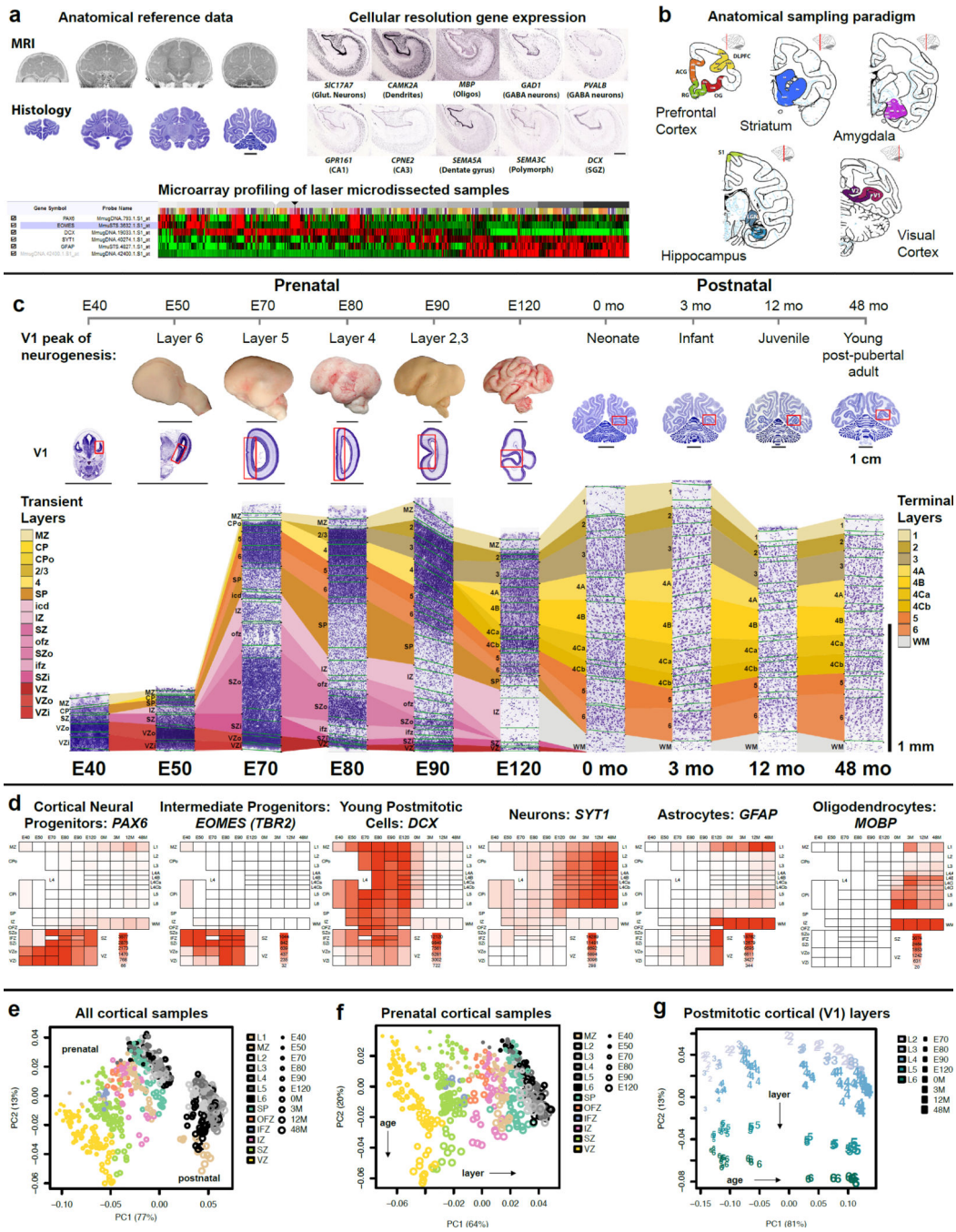


Figure 1. High-resolution transcriptional profiling of rhesus monkey brain development

a, Neuroimaging, histological and transcriptome data components. **b**, Brain regions analyzed. **c**, Primary visual cortex (V1) sampling paradigm for transcriptome analysis spanning ages from 40 post-conceptional days (E40) to 48 months after birth, with salient developmental events or stages noted (top row). Whole brain or hemisphere sections are shown for each age (V1, red box), as well as a high magnification Nissl image detailing laminar microdissections. Cortical layers profiled at each stage are color-coded by predominant mitotic (red, pink) or postmitotic (orange, yellow, tan) cell makeup. **d**,

Heatmaps of canonical cell type marker gene expression organized anatomically as in **c. e-g**, Multi-dimensional scaling (MDS) plots showing the first two principal axes of variation for all cortical samples (n = 922) (**e**), prenatal samples (n = 542) (**f**) and postmitotic layers in V1 (n = 175) (**g**). Scale bars in **c**: 1 cm (top panels) or 1 mm (bottom panel). Structure abbreviations in Supplementary Table 1.

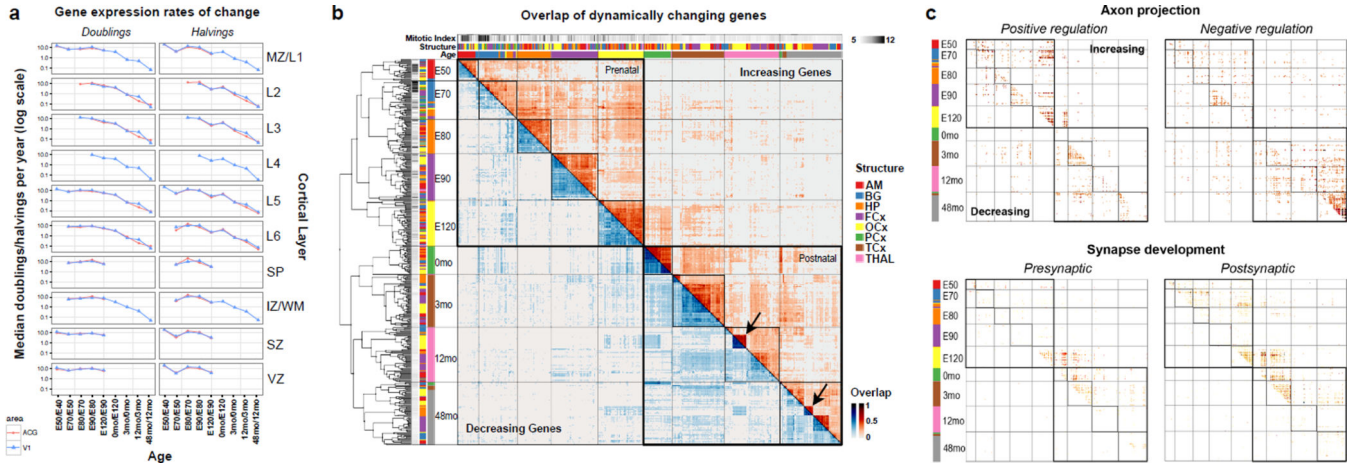


Figure 2. Transcriptional dynamics across brain regions and ages

a, Similar decreasing rates of gene expression change across cortical layers (rows, ordered from apical to basal) and areas (anterior cingulate gyrus, ACG, red; primary visual cortex, V1, blue). **b**, Overlap of temporally dynamic genes across brain regions and ages. Proportion of the top 1000 increasing (red, upper triangle) or decreasing (blue, lower triangle) genes for each region between pairs of adjacent ages (e.g. E50 compared to E40) that are shared with other ages and regions. Note large overlap between all regions at each age and within prenatal and postnatal periods. Arrows highlight region-specific rapid expression change. AM, amygdala; BG, basal ganglia; HP, hippocampus; FCx frontal, OCx occipital, PCx parietal and TCx temporal cortex; THAL, thalamus. **c**, Timing of neuronal connectivity-related events revealed by enrichment of gene ontology (GO) terms in overlapping gene lists from **b**.

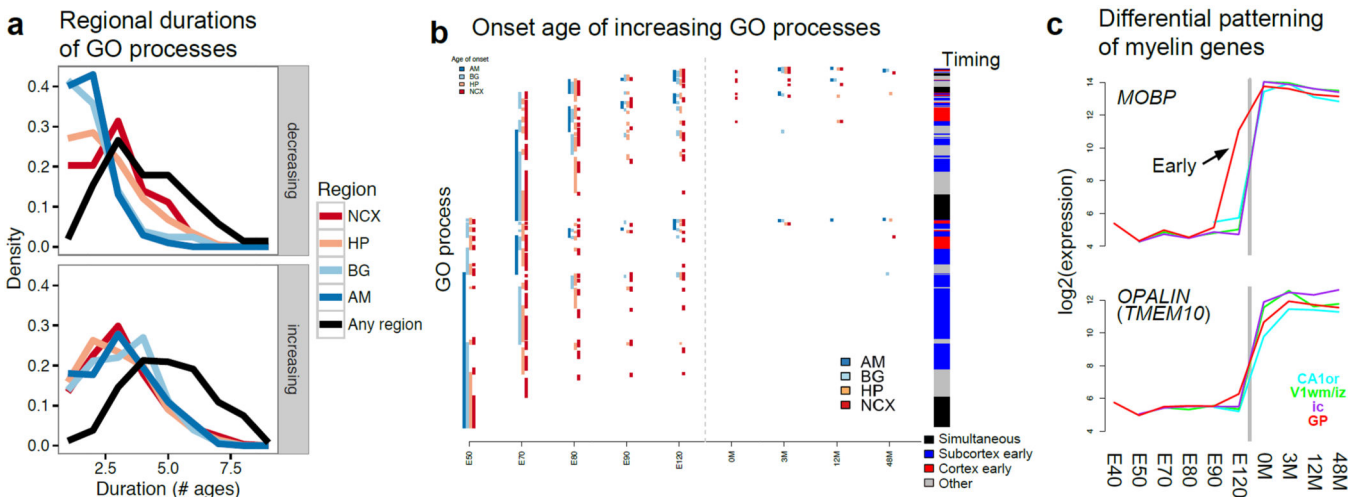


Figure 3. Variable onset of biological processes between brain regions

a, Prolonged duration of biological GO processes. Number of ages at which increasing and decreasing GO processes are active in each region (colors) or in any region (black). **b**, Synchronous and asynchronous onset of GO processes between regions, mainly beginning in prenatal development. Left: Age of onset for all increasing GO processes ordered by onset in subcortical and then cortical regions (AM, BG, HP, then NCX). Right: GO processes predominantly begin earlier in subcortical than cortical regions. **c**, Early onset (arrow) of mature oligodendrocyte marker *MOBP* in subcortical globus pallidus (arrow) compared to cortical myelin-enriched regions. Other myelin genes such as *OPALIN* show synchronous developmental timing. CA1or, hippocampal CA1 stratum oriens; V1 iz/wm, intermediate zone/white matter; ic, internal capsule.

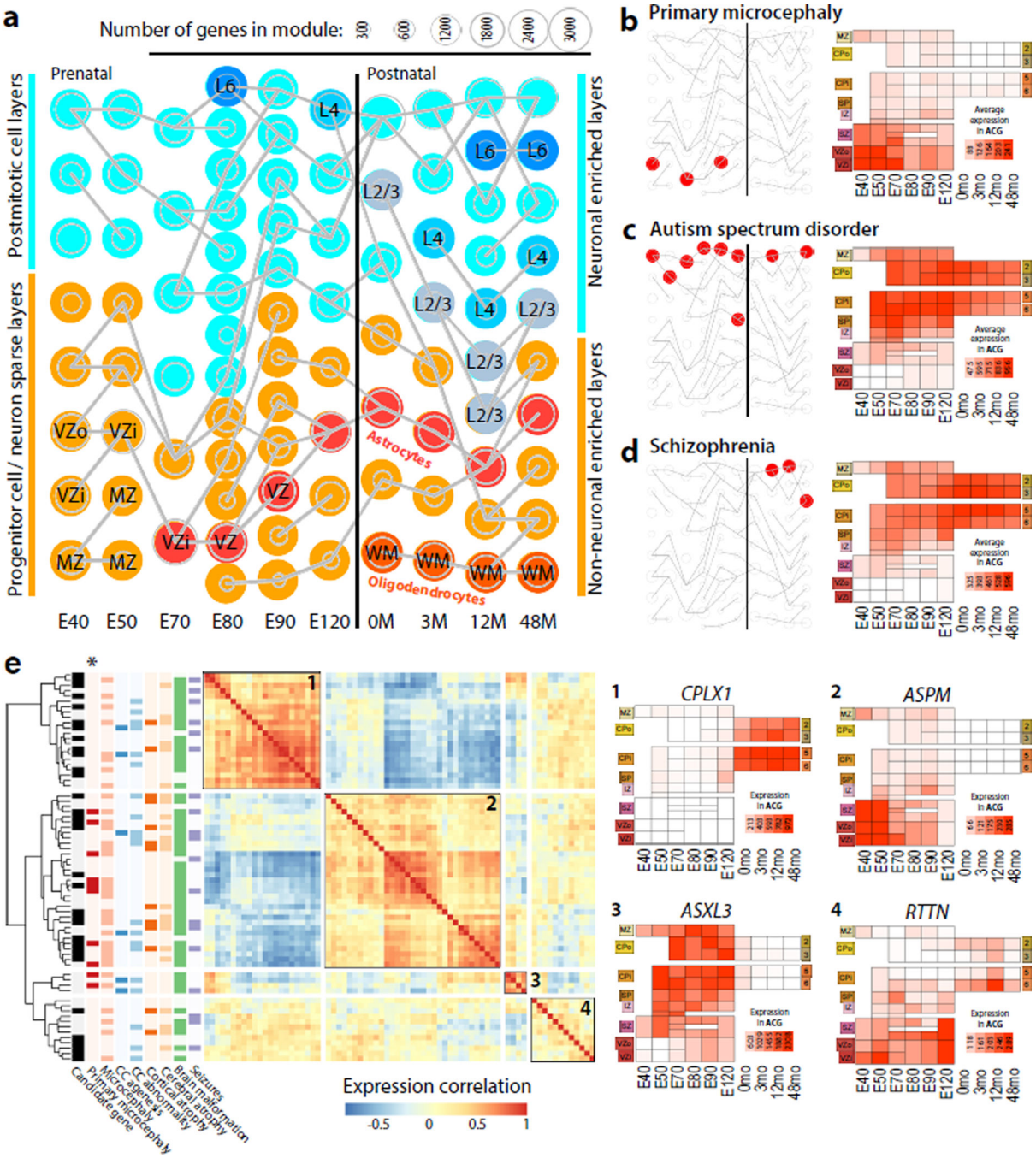


Figure 4. Protracted maturation of neocortex through young adulthood

a, Expression patterns of genes best marking each cortical layer progressively change with time. Each dot represents the average across-layer expression correlation of the 10% of genes best distinguishing layers at one age (source list) with their expression levels in another age (comparison). Values are plotted against a scaled difference in age. **b**, Layer-enriched gene expression shows progressive and dramatic change with age. Binary heatmap showing expression of layer-enriched genes (rows) across ages (columns). Genes are ordered and divided based on the median age of laminar expression into four chronological groups:

early (median age < E120), middle (E120 – 3M), late (> 3M) and persistent. **c**, Developmental expression of representative genes from each chronological group in layer 6. **d**, Number of genes differentially expressed between rostral (ACG) and caudal (V1) cortex in superficial (L2/3), deep (L5/6) and proliferative (SZ, VZ) layers across development. Most differences are found in postmitotic layers during early prenatal and late postnatal development. **e-h**, Genes marking specific cell classes show regional enrichment consistent with earlier neuro- and gliogenesis in ACG compared to V1. Left column: enrichment of regional genes (from **d**) with markers for **e**, cell cycle (progenitors); **f**, GABAergic and **g**, glutamatergic neurons in V1; and **h**, astrocytes. Size of box indicates significance of cell type enrichment in a given layer, region, and age. Right column: expression levels (mean +/− SE) of canonical marker genes for each cell type in V1 (blue) and ACG (red).

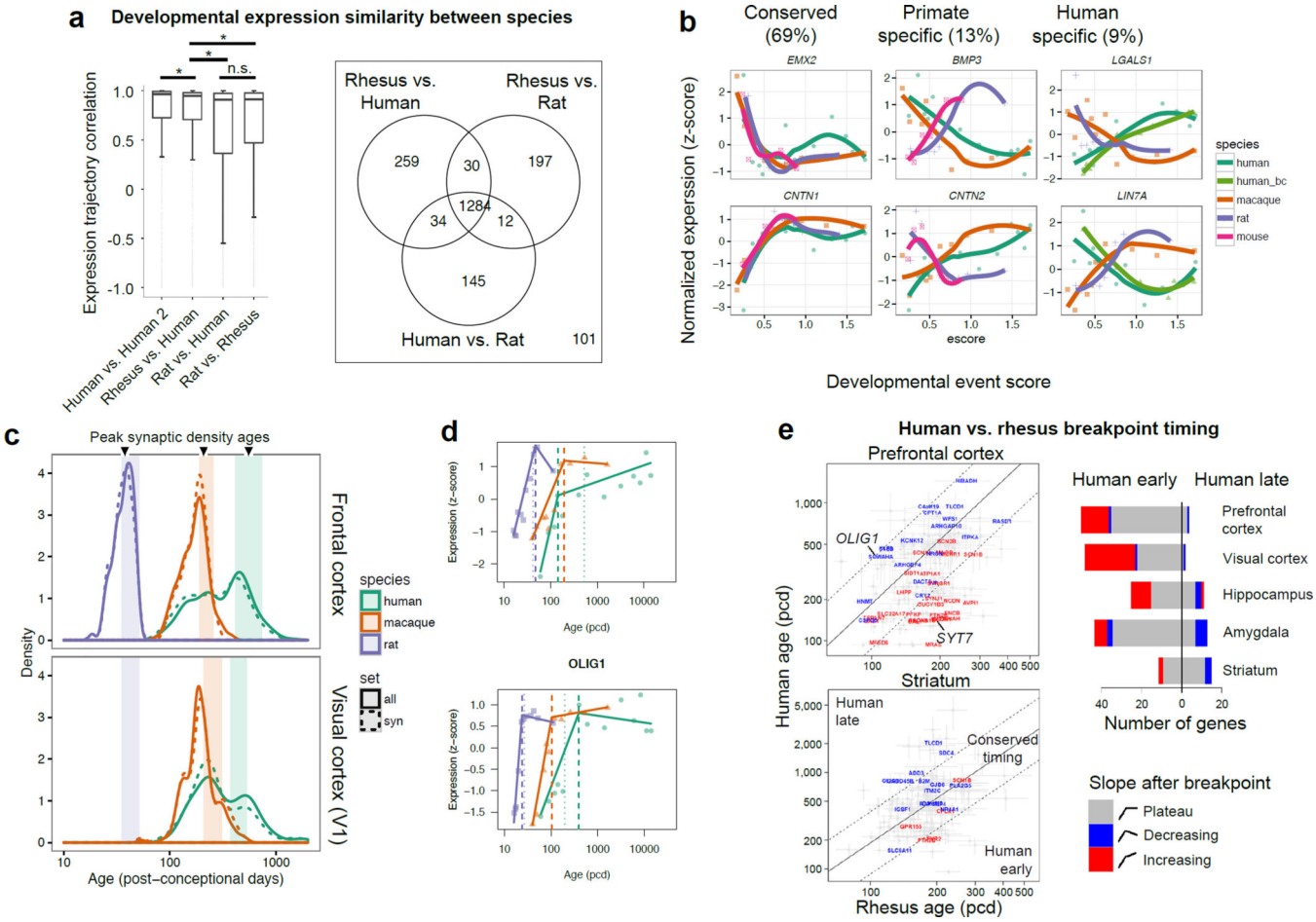


Figure 5. Spatiotemporal localization of disease-specific associations in developing cortex

a, WGCNA gene co-expression networks (modules; Supplementary Table 9) based on pooled V1 and ACG samples analyzed independently at each age (columns). Modules were tested for significant gene set enrichment and overlap with hypergeometric tests. Modules significantly enriched for markers of glial cell classes ($P < 10^{-15}$) or cortical layers ($P < 10^{-30}$) are color-coded and annotated. Remaining modules are colored or labeled based on maximal expression in postmitotic (neuron-enriched; cyan) layers or progenitor or largely non-neuronal (WM, Layer 1; orange) layers. Modules from adjacent ages with the most highly significant gene overlap ($P < 10^{-50}$) are connected by grey lines. **b-d**, Left: Modules significantly enriched for risk genes associated with neurodevelopmental disorders (empirically corrected $P < 0.1$; red discs). Right: Average expression pattern of genes found in at least two enriched modules. **b**, Genes related to primary autosomal recessive microcephaly are enriched in early non-neuronal modules, and show maximal expression in early prenatal VZ. **c**, Genes related to autism spectrum disorder are enriched in modules associated with cortical neurons, and show highest expression in cortical plate across development. **d**, Genes related to schizophrenia show similar neuronal layer enrichment to autism genes, but restricted to postnatal ages. **e**, Left: Genes related to intellectual disability are associated with different brain malformations and/or seizures and show four major expression patterns over cortical development. Right: Expression heatmaps of known and

candidate disease genes representing developmental profiles of each cluster. ID genes also associated with primary microcephaly are significantly enriched in cortical progenitor-enriched pattern 2 (* $P < 0.001$; multivariate distance matrix regression permutation test). CC, corpus callosum.

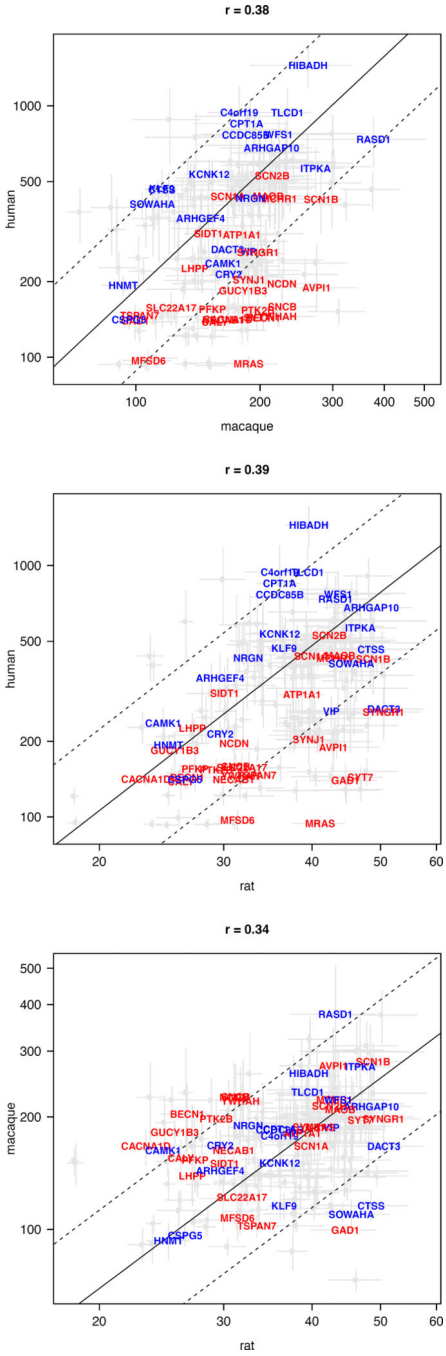


Figure 6. Conserved and human-specific gene expression trajectories in frontal cortex

a, Left: Boxplots of pairwise correlations between developmental expression trajectories (median ± 25th and 75th percentiles, whiskers at 1.5 times the inter-quartile range) of orthologous genes profiled in rhesus monkey, rat and two human data sets. Kruskal-Wallis rank sum test, post-hoc Wilcoxon signed rank paired tests: * P < 0.001 (Bonferroni-corrected), n.s. not significant. Right: Venn diagram showing the number of conserved (R > 0.5) genes between each pair of species. **b**, Examples of conserved and species-specific gene trajectories. Colors and symbols are consistent in **b-d. c**, Distribution of breakpoint ages for

179 orthologs (solid lines) of which 81 were synapse related (dashed lines) in frontal cortex and V1. Shaded bars indicate periods of peak synaptic density (95% confidence intervals) in each species (Extended Data Extended Data Fig. 7b). **d**, Genes with unexpectedly early and late breakpoints in expression trajectories in human. Observed breakpoints (dashed lines) compared to expected breakpoints (dotted lines) based on timing in rhesus monkey. **e**, Many genes in human cortex have early breakpoints followed by prolonged increase in expression. Left: Comparison of breakpoint timing (points with 95% confidence intervals) between human and rhesus shows a biased population of early breakpoint genes in human cortex but not striatum. Many of these early breakpoint genes continue to increase expression through young adulthood (red) in contrast to genes with relatively conserved timing that tend to plateau (grey) or decrease (blue). Ages corresponding to equivalent (solid lines) or nearby (dashed lines) developmental stages (event scores within ± 0.2; see **Methods**) in human and rhesus monkey are shown. Right: Bar plot summarizing the number of genes that have an early or late breakpoint in human in different brain regions.