

# Transcriptional landscape of the prenatal human brain

Jeremy A. Miller<sup>1\*</sup>, Song-Lin Ding<sup>1\*</sup>, Susan M. Sunkin<sup>1</sup>, Kimberly A. Smith<sup>1</sup>, Lydia Ng<sup>1</sup>, Aaron Szafer<sup>1</sup>, Amanda Ebbert<sup>1</sup>, Zackery L. Riley<sup>1</sup>, Joshua J. Royall<sup>1</sup>, Kaylynn Aiona<sup>1</sup>, James M. Arnold<sup>1</sup>, Crissa Bennet<sup>1</sup>, Darren Bertagnolli<sup>1</sup>, Krissy Brouner<sup>1</sup>, Stephanie Butler<sup>1</sup>, Shiella Caldejon<sup>1</sup>, Anita Carey<sup>1</sup>, Christine Cuhacian<sup>1</sup>, Rachel A. Dalley<sup>1</sup>, Nick Dee<sup>1</sup>, Tim A. Dolbear<sup>1</sup>, Benjamin A. C. Facer<sup>1</sup>, David Feng<sup>1</sup>, Tim P. Fliss<sup>1</sup>, Garrett Gee<sup>1</sup>, Jeff Goldy<sup>1</sup>, Lindsey Gourley<sup>1</sup>, Benjamin W. Gregor<sup>1</sup>, Guangyu Gu<sup>1</sup>, Robert E. Howard<sup>1</sup>, Jayson M. Jochim<sup>1</sup>, Chihchau L. Kuan<sup>1</sup>, Christopher Lau<sup>1</sup>, Chang-Kyu Lee<sup>1</sup>, Felix Lee<sup>1</sup>, Tracy A. Lemon<sup>1</sup>, Phil Lesnar<sup>1</sup>, Bergen McMurray<sup>1</sup>, Naveed Mastan<sup>1</sup>, Nerick Mosqueda<sup>1</sup>, Theresa Nalwai-Cecchini<sup>2</sup>, Nhan-Kiet Ngo<sup>1</sup>, Julie Nyhus<sup>1</sup>, Aaron Oldre<sup>1</sup>, Eric Olson<sup>1</sup>, Jody Parente<sup>1</sup>, Patrick D. Parker<sup>1</sup>, Sheana E. Parry<sup>1</sup>, Allison Stevens<sup>3,4</sup>, Mihovil Pletikos<sup>5</sup>, Melissa Reding<sup>1</sup>, Kate Roll<sup>1</sup>, David Sandman<sup>1</sup>, Melaine Sarreal<sup>1</sup>, Sheila Shapouri<sup>1</sup>, Nadiya V. Shapovalova<sup>1</sup>, Elaine H. Shen<sup>1</sup>, Nathan Sjoquist<sup>1</sup>, Clifford R. Slaughterbeck<sup>1</sup>, Michael Smith<sup>1</sup>, Andy J. Sodt<sup>1</sup>, Derric Williams<sup>1</sup>, Lilla Zöllei<sup>3</sup>, Bruce Fischl<sup>3,4</sup>, Mark B. Gerstein<sup>6,7</sup>, Daniel H. Geschwind<sup>8</sup>, Ian A. Glass<sup>2</sup>, Michael J. Hawrylycz<sup>1</sup>, Robert F. Hevner<sup>9,10</sup>, Hao Huang<sup>11</sup>, Allan R. Jones<sup>1</sup>, James A. Knowles<sup>12</sup>, Pat Levitt<sup>13,14</sup>, John W. Phillips<sup>1</sup>, Nenad Šestan<sup>5</sup>, Paul Wohnoutka<sup>1</sup>, Chinh Dang<sup>1</sup>, Amy Bernard<sup>1</sup>, John G. Hohmann<sup>1</sup> & Ed S. Lein<sup>1</sup>

**The anatomical and functional architecture of the human brain is mainly determined by prenatal transcriptional processes. We describe an anatomically comprehensive atlas of the mid-gestational human brain, including *de novo* reference atlases, *in situ* hybridization, ultra-high resolution magnetic resonance imaging (MRI) and microarray analysis on highly discrete laser-microdissected brain regions. In developing cerebral cortex, transcriptional differences are found between different proliferative and post-mitotic layers, wherein laminar signatures reflect cellular composition and developmental processes. Cytoarchitectural differences between human and mouse have molecular correlates, including species differences in gene expression in subplate, although surprisingly we find minimal differences between the inner and outer subventricular zones even though the outer zone is expanded in humans. Both germinal and post-mitotic cortical layers exhibit fronto-temporal gradients, with particular enrichment in the frontal lobe. Finally, many neurodevelopmental disorder and human-evolution-related genes show patterned expression, potentially underlying unique features of human cortical formation. These data provide a rich, freely-accessible resource for understanding human brain development.**

The human brain develops following a complex, highly stereotyped series of histogenic events that depend on regulated differential gene expression, and acquired or inherited disruption can lead to devastating consequences. Mainly due to limitations in access to human prenatal tissue, most developmental studies are performed in mouse or non-human primates (however, see refs 1–5). Yet significant species differences exist, necessitating the study of human brain. For example, the human neocortex has undergone massive evolutionary expansion, particularly in superficial layers, probably due to differences in rates of progenitor pool expansion during neurogenesis compared to other species<sup>6</sup>. A secondary progenitor zone, the subventricular zone (SZ) is present in all mammals, but is split into an outer and inner region in primates<sup>7</sup>. The transient subplate zone (SP) is greatly expanded in human<sup>8</sup>, as is the subpial granular zone (SG), a transient compartment at the pial surface composed primarily of tangentially migrating neurons<sup>9</sup>. Furthermore, there is evidence for species differences in the developmental origin of cortical GABAergic interneurons. In mouse, nearly all

originate from the striatal ganglionic eminences (GEs) of the ventral telencephalon<sup>10</sup>; however, the origin of human cortical interneurons remains controversial<sup>7,11–14</sup>. Finally, understanding the emergence of cortical specialization for language can only be studied in humans.

Recent studies have begun to analyse the developing brain and neocortical transcriptome. Profiling of mid-gestational human brain<sup>3,5</sup> identified many genes differentially expressed between major regions, including genes associated with human-accelerated conserved noncoding sequences (haCNSs)<sup>15</sup>. Gene expression also varies between cell populations, and more detailed analysis of layers of fetal mouse neocortex found >2,500 genes differentially expressed between ventricular zone (VZ), subventricular zone, intermediate zone (IZ), and cortical plate (CP)<sup>16</sup>. Species differences in distinct fetal transient zones, including the subventricular zone<sup>2</sup>, subplate zone<sup>17,18</sup>, cortical plate<sup>6</sup>, and subpial granular zone<sup>9</sup>, have also been described.

The goal of the current project was to create resources for studying prenatal human brain development and the early roots of neurodevelopmental

<sup>1</sup>Allen Institute for Brain Science, Seattle, Washington 98103, USA. <sup>2</sup>Division of Genetic Medicine, Department of Pediatrics, University of Washington, 1959 North East Pacific Street, Box 356320, Seattle, Washington 98195, USA. <sup>3</sup>Department of Radiology, Harvard Medical School, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. <sup>4</sup>Computer Science and AI Lab, MIT, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Department of Neurobiology and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, Connecticut 06510, USA. <sup>6</sup>Program in Computational Biology and Bioinformatics, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. <sup>7</sup>Department of Computer Science, Yale University, New Haven, Connecticut 06520, USA. <sup>8</sup>Program in Neurogenetics, Department of Neurology and Semel Institute David Geffen School of Medicine, UCLA, Los Angeles, California 90095, USA. <sup>9</sup>Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, Washington 98101, USA. <sup>10</sup>Department of Neurological Surgery, University of Washington School of Medicine, Seattle, Washington 98105, USA. <sup>11</sup>Advanced Imaging Research Center, UT Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>12</sup>Zilkha Neurogenetic Institute, and Department of Psychiatry, University of Southern California, Los Angeles, California 90033, USA. <sup>13</sup>Department of Pediatrics, Children's Hospital, Los Angeles, California 90027, USA. <sup>14</sup>Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA.

\*These authors contributed equally to this work.

and psychiatric disorders<sup>19</sup>. These include anatomical reference atlases similar to those for model organisms<sup>20–22</sup>, and an anatomically comprehensive, detailed transcriptional profiling of normal mid-gestational brain modelled on atlases of adult mouse and human brain<sup>23,24</sup> and using methods for selective analysis of discrete structural nuclei and layers<sup>25</sup>. These data are freely accessible as part of the BrainSpan Atlas of the Developing Human Brain (<http://brainspan.org>) via the Allen Brain Atlas data portal (<http://www.brain-map.org>).

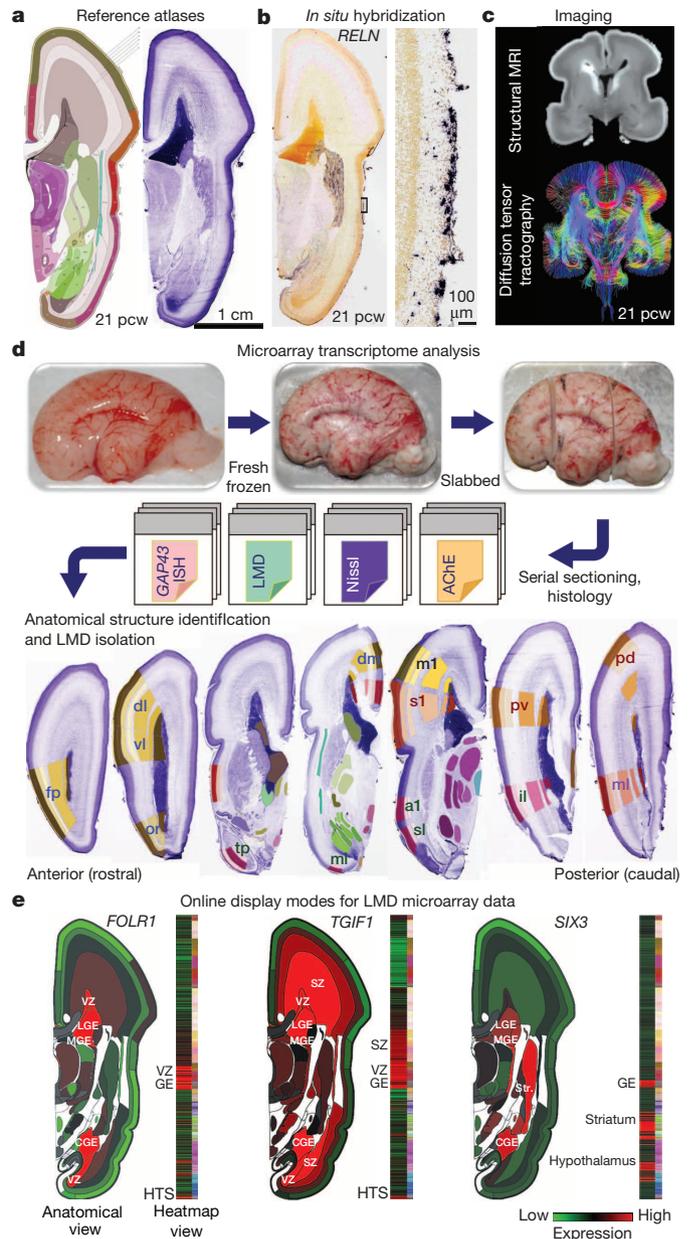
### Transcriptome analysis of prenatal brain

Four intact high-quality mid-gestational brains, two from 15 and 16 post-conceptual weeks (pcw) and two from 21 pcw (Supplementary Table 1), were used to create detailed *de novo* reference atlases and transcriptome data sets (Fig. 1). The entire left hemisphere of each specimen was coronally, serially cryosectioned onto polyethylene naphthalate (PEN) membrane slides for laser microdissection (LMD), with interleaved slides for histological staining (Nissl, acetylcholinesterase (AChE), and *in situ* hybridization for *GAP43*) for detailed structure identification. Approximately 300 anatomical regions per specimen were isolated for RNA isolation, amplification and microarray analysis on custom 64K Agilent microarrays<sup>23</sup> (Fig. 1d, Supplementary Table 2 and Extended Data Fig. 1). For one 15 pcw and one 21 pcw specimen, the right hemisphere was processed similarly but used for *in situ* hybridization (Fig. 1b) and Nissl staining. These data were anatomically delineated to make digital reference atlases (Fig. 1a), which allow the representation of transcriptome data in native anatomical coordinates. Figure 1e illustrates the specificity of anatomical sampling using this representation. For example, the folate receptor *FOLR1* is selectively expressed in the ventricular zone and ganglionic eminences. Sufficient folate intake is essential for normal neuronal development<sup>26</sup>, and mutations in *FOLR1* cause severe neurological sequelae due to cerebral folate transport deficiency. Similarly, two genes associated with abnormal cortical development in holoprosencephaly, *TGIF1* and *SIX3* (ref. 27), are also enriched in cortical germinal zones. Finally, structural magnetic resonance imaging (MRI) and diffusion weighted MRI (DWI) data from 3 approximately age-matched brains (Fig. 1c and Extended Data Fig. 2), as well as reconstruction of fetal white matter tracts for three additional brains (Extended Data Fig. 3 and see ref. 28), are included for anatomical reference in the online resource.

### Laminar transcriptional patterning

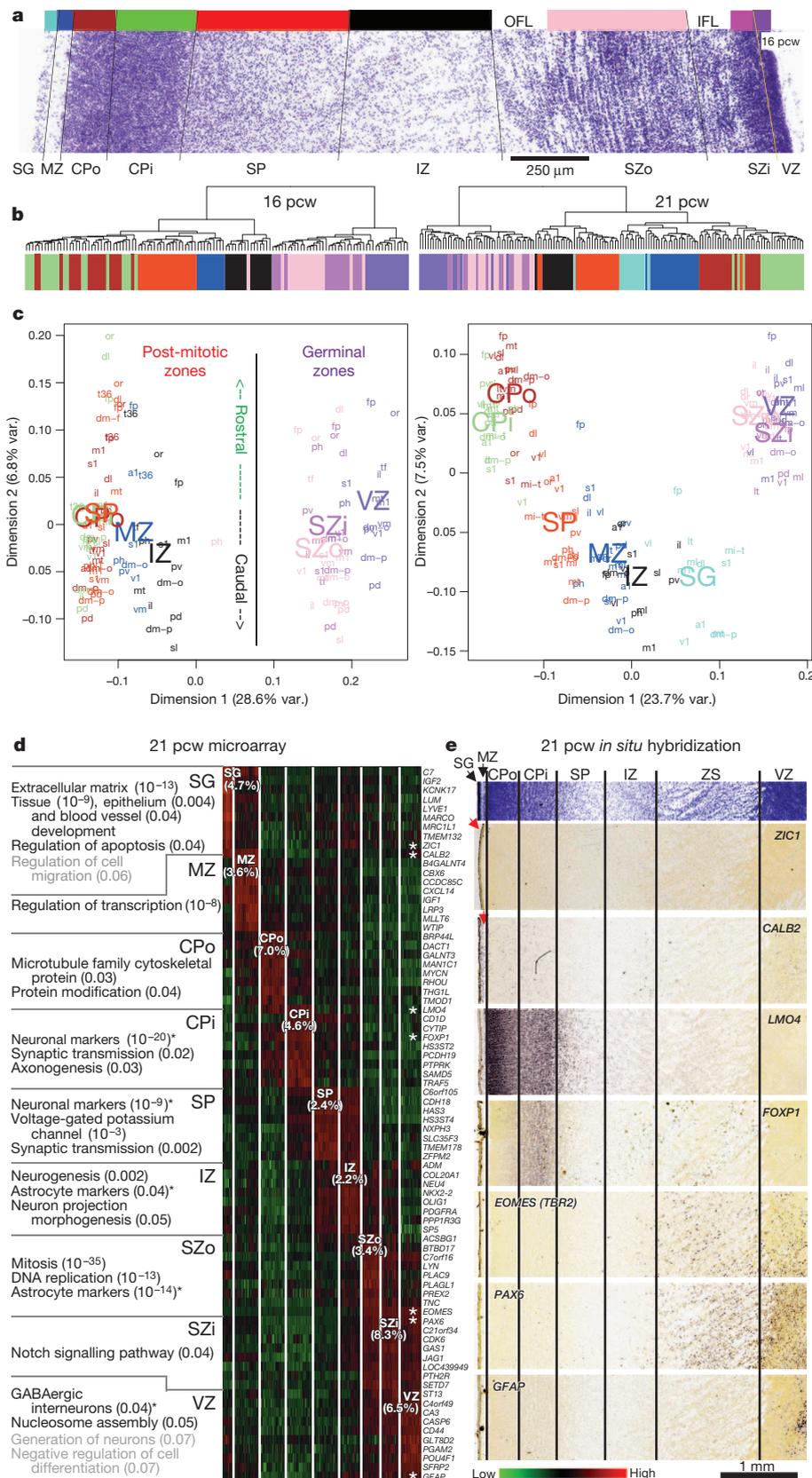
We assayed ~25 areas of the developing neocortex, delineating nine layers per area (here referring to fetal mitotic and post-mitotic zones and not layers 1–6 of mature neocortex)<sup>29</sup>: Subpial granular zone, marginal zone (MZ), outer and inner cortical plate (CPO; CPI), subplate zone, intermediate zone (or inner subplate), outer and inner subventricular zone (SZo; SZi), and ventricular zone (Fig. 2a). Approximately 95% of RefSeq genes are expressed in developing neocortex, compared with 84% identified using these arrays in adult<sup>23</sup>. Different layers show robust and unique molecular signatures, and samples group by layer using hierarchical clustering (using differential genes from ANOVA,  $P < 10^{-29}$ ) at both time points (Fig. 2b). Samples also cluster with multi-dimensional scaling, where dimension one separates samples by layer, with germinal zones (VZ, SZi, SZo) distinct from layers containing primarily post-mitotic cells, and dimension two roughly reflects rostro-caudal position (Fig. 2c).

To identify laminar signatures at 21 pcw, we correlated each gene with a binary vector (1 in tested layer versus 0 elsewhere), which identified ~2,000 layer-enriched genes with  $R > 0.5$  in both brains (Supplementary Table 3). Each layer included genes with high laminar specificity (Fig. 2d), although the SZi profiles tended to overlap with the neighbouring SZo and ventricular zone. *In situ* hybridization validated the specificity of layer enriched genes (Fig. 2e). For example, the Cajal-Retzius cell marker *CALB2* (ref. 30) showed enrichment in the marginal zone and subpial granular zone as expected. The cortical progenitor markers *TBR2* and *PAX6* are enriched in germinal layers as in mouse, although



**Figure 1 | Prenatal human brain atlas components.** **a**, Nissl stained (right) and corresponding annotated reference atlas (left) plate, colour coded by structure. **b**, *In situ* hybridization for *RELN* showing expression in Cajal-Retzius cells at low (left) and high (right) magnification in marginal zone. **c**, High resolution MRI and tract DWI of fixed *ex cranio* brain. **d**, Experimental strategy for systematic histology, anatomical delineation and LMD-based isolation of discrete anatomical structures for microarray analysis. Nissl, acetylcholinesterase (AChE) and *GAP43* *in situ* hybridization were used to identify structures. **e**, Quantitative representation of microarray data for *FOLR1*, *TGIF1* and *SIX3*. GE, ganglionic eminence. See Supplementary Table 2 for other anatomical abbreviations.

*Pax6* in mouse is restricted to the ventricular zone, whereas it is also highly expressed in the human subventricular zone<sup>31</sup>. Ventricular-zone-restricted *GFAP* expression probably marks radial glia (RG)<sup>32</sup>. Finally, expression of *ZIC1*, associated with Dandy-Walker congenital brain malformation<sup>33</sup>, was restricted to the pia mater overlying the cortex, therefore indicating that subpial granular zone samples captured pial cells in addition to granule cells. However, whereas mouse *Zic1* is expressed by virtually all Cajal-Retzius neurons<sup>34</sup>, our results indicate that this is not true in human as *ZIC1* and *CALB2* expression do not overlap in the marginal zone (Figs 1b and 2e).



**Figure 2 | Laminar gene expression mirrors developmental processes in prenatal neocortex.** **a**, Nissl section from 16 pcw cortex showing layers dissected for analysis (represented by the colour bar). The outer and inner fibre layers (OFL, IFL) were omitted from the dissection. **b**, Sample clustering based on the 100 most significant genes differentiating layers by ANOVA ( $P < 10^{-29}$ ) groups samples by layer at both time points. **c**, Multidimensional scaling using all genes demonstrates clustering of post-mitotic versus germinal zones (dimension 1) and to a lesser extent by rostrocaudal position (dimension 2). **d**, Layer-enriched gene expression, based on correlation to binary templates at 21 pcw (Methods). Enriched genes in each layer relate to cellular makeup and developmental maturity of those cells. Asterisks indicate genes investigated further (see panel e).  $P$  values are included in parentheses. Additional categories just missing significance at  $P < 0.05$  are shown in grey. **e**, Validation of laminar enrichment by *in situ* hybridization for genes with asterisks in d. See Supplementary Table 2 for anatomical abbreviations.

These laminar expression patterns mirror cellular composition and developmental processes, shown by enrichment analysis (Fig. 2d, Supplementary Table 4 and Methods). SZo-enriched categories primarily related to cell division and contained many astrocytic markers probably

expressed in outer radial glia (ORG)<sup>7</sup>. Functional ontology of post-mitotic layers reflected developmental maturity. Subplate, which contains the earliest-generated neurons, showed enrichment for mature neuronal markers and synaptic transmission, reflecting early thalamic

afferent input by mid-gestation<sup>35</sup>. The next oldest neurons in CPi are additionally enriched for genes involved in forming connections, whereas the youngest neurons in CPo are primarily enriched for terms related to metabolism rather than mature neuronal function.

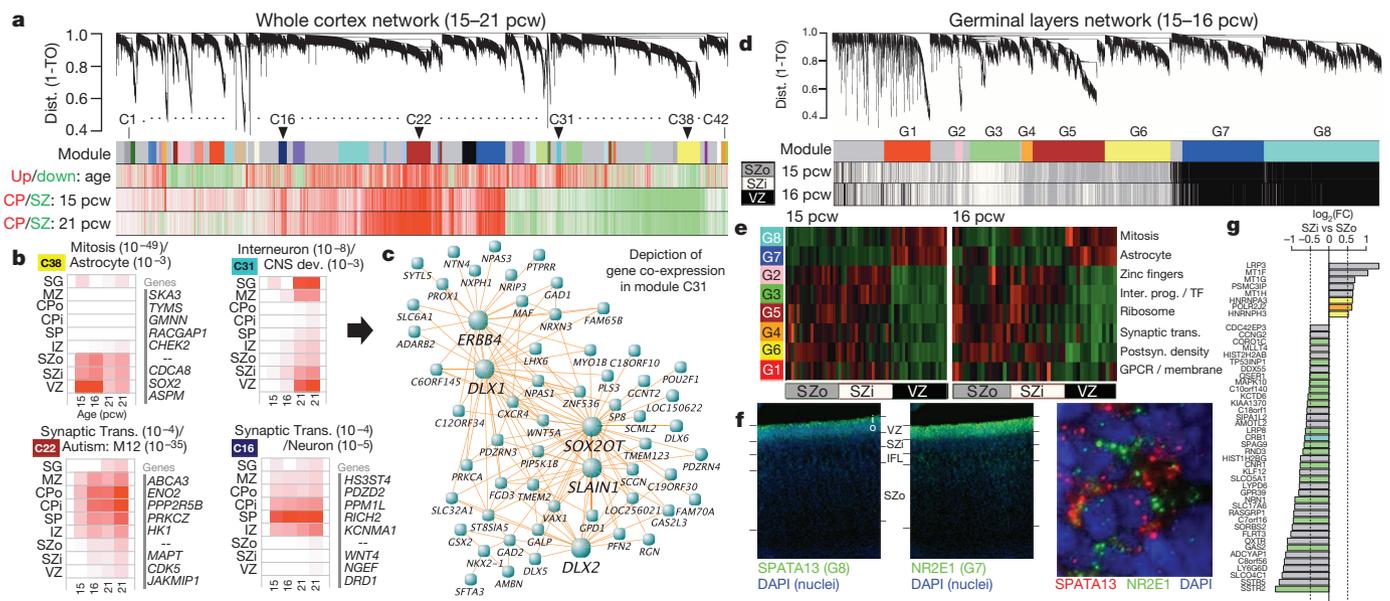
### Gene networks discriminate fetal cell types

To identify principal features of the developing cortical transcriptome, we performed weighted gene co-expression network analysis (WGCNA)<sup>36</sup> on all 526 neocortical samples, and identified 42 modules of co-expressed genes (Fig. 3a, Supplementary Tables 5 and 6 and Supplementary Methods). WGCNA clusters genes with similar expression patterns in an unbiased manner, allowing a biological interpretation of transcriptional patterns (layer, cell type, biological process, disease and so on)<sup>23,36–38</sup>. Here, most gene clusters ('modules') corresponded to layers and/or changes with age, (Fig. 3a, b and Extended Data Fig. 4) whereas areal patterning appeared to be a more subtle transcriptional feature. For example, module C16 is enriched in subplate (Fig. 3b, lower right), and shows hallmarks of mature neuronal function. Module C38 contains genes enriched in germinal layers, and also decreased expression with age (Fig. 3b, upper left). This module has a large signature of glia and cell division, suggesting that these genes reflect decreasing progenitor cell division. Conversely, module C22 is enriched in newly generated post-mitotic neurons of the cortical plate, and increases with age (Fig. 3b, lower left). Importantly, given the small sample size, this temporal patterning in C38 and C22 is corroborated by RNA-seq data from a larger time series of cortical development contained in the BrainSpan resource (Extended Data Fig. 5). Genes in module C22 significantly overlap with genes showing altered expression in postnatal human brains of patients affected by autism<sup>38</sup>. This suggests involvement of autism risk factors in early development of excitatory cortical neurons, consistent with other recent studies<sup>39,40</sup>.

Finally, we identified a module (C31) with particular enrichment in subplate granular zone and ventricular zone (Fig. 3b, upper right), containing many interneuron-associated genes (Fig. 3c). DLX1 and DLX2, homeobox transcription factors essential for interneuron migration and

survival<sup>10</sup>, were central ('hub') genes therein. There is controversy regarding the origin of cortical interneurons in primates. The argument has been made that a substantial proportion of interneurons are generated locally in the ventricular zone<sup>12,13</sup>. However, several recent studies have shown strong evidence that, as with rodents, primate interneurons are generated extracortically in the ganglionic eminences<sup>11,14</sup>. To address this issue, we generated a new network using neocortical as well as ganglionic eminences samples, and examined the distribution of C31 genes therein. Most genes were assigned to a module showing common enrichment in both the ganglionic eminences and the ventricular zone (Extended Data Fig. 6). Although this finding does not resolve the origin of cortical interneurons, it shows that transcriptional programs associated with these cells in both structures are highly similar.

Germinal layers contain various cortical progenitors including radial glia in the ventricular zone, intermediate progenitors (IP) in the SZi, and ORG in the SZo (reviewed in ref. 41), and these radial glia may be quite diverse<sup>42</sup>. To search for coherent expression profiles marking putative progenitor populations, we created a consensus co-expression network using samples from VZ, SZi and SZo in the 15 and 16 pcw brains, only including genes differentially expressed between these layers (Fig. 3d; Supplementary Table 7 and Supplementary Methods). This network will only identify intergenic relationships common to both brains and should therefore produce robust co-expression relationships that exclude specimen-specific features or changes with age. We found eight co-expression modules with selective enrichment in either the subventricular zone or the ventricular zone (Fig. 3d, e), highly conserved between brains ( $P < 10^{-50}$ ), and highly distinct. Both ventricular-zone-enriched modules (G7, G8) contained cell cycle genes and many astrocyte markers, indicating that these modules may represent different radial glia populations. Fluorescent *in situ* hybridization (FISH) confirmed that representative genes in both G7 (NR2E1) and G8 (SPATA13) are enriched in the ventricular zone versus subventricular zone (Fig. 3f and Extended Data Fig. 7). Surprisingly, rather than labelling distinct cell populations in the ventricular zone, these genes labelled mutually exclusive



**Figure 3 | Co-expression analyses of prenatal cortex.** **a**, WGCNA cluster dendrogram on all 526 neocortical samples groups genes into 42 distinct modules (row 1). Rows 2–4 show strong differential expression relationships between module genes and age or cortical layer. TO, topological overlap. **b**, Module eigengene (ME) expression of four notable modules in **a**, averaged across brain and layer. Modules are biologically characterized with significant category enrichments and representative genes. White, low expression; red, high expression. **c**, Many top gene–gene connections for module C31 are

shown, including several known GABAergic interneuron genes. **d**, Cluster dendrogram for consensus network focused on germinal layers identifies modules (row 1) enriched in each layer (rows 2 and 3). **e**, ME heatmap shows differential VZ/SZ expression for 8 modules, along with enriched gene sets. **f**, FISH on 15 pcw frontal cortex shows enrichment of SPATA13 and NR2E1 in mutually exclusive subcellular localization in VZ. **g**, Genes enriched in SZi and differentially expressed between SZo and SZi. Genes colour-coded by module (grey = unassigned).

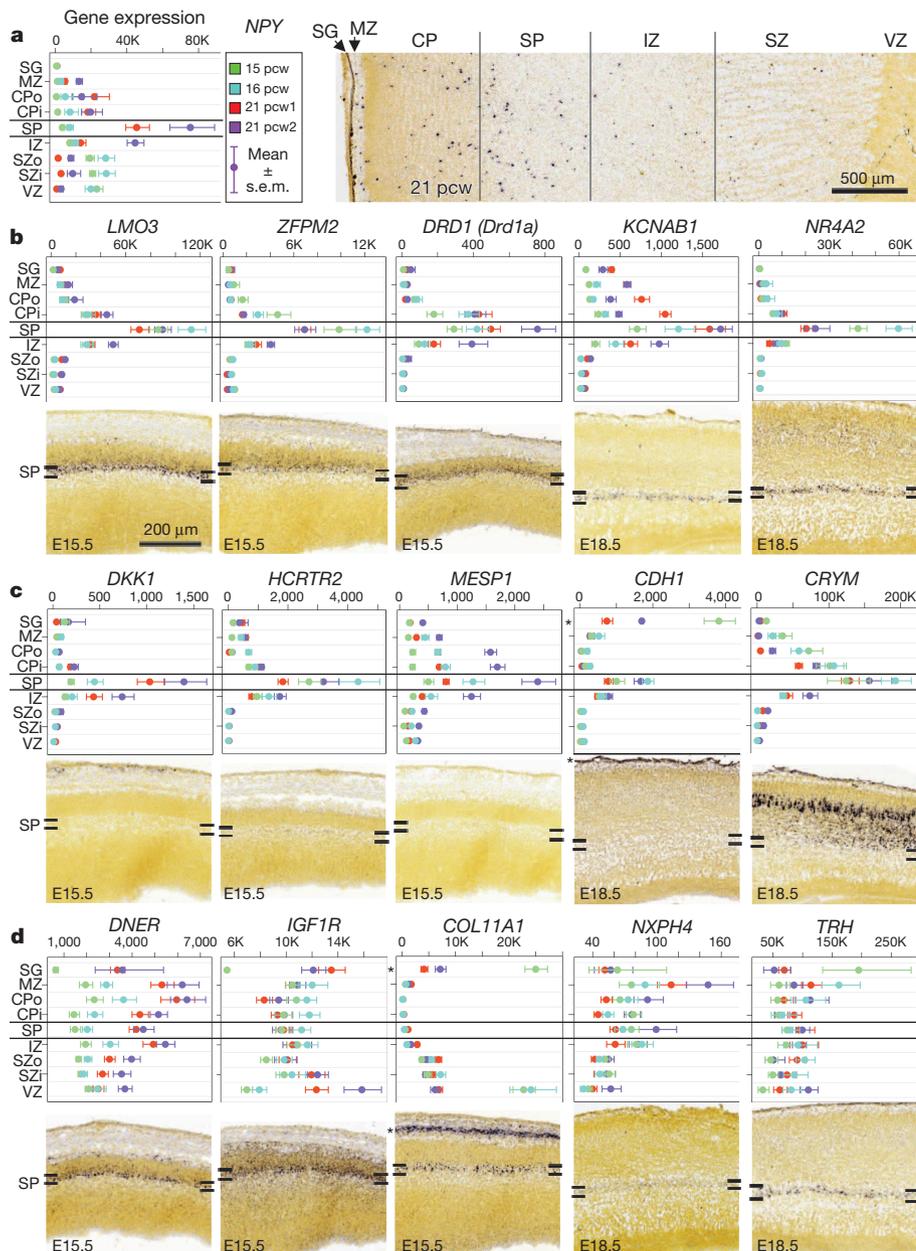
subcellular locations within the same ventricular zone cells (note non-overlapping cytoplasmic localization of SPATA13 and NR2E1), indicating that these modules probably represent differentially regulated biological processes within radial glia cells.

We were particularly interested in identifying differences between the inner and outer subventricular zone (SZi and SZo), which is absent in mice. Modules G2 and G3 are enriched in the SZi (Fig. 3d, white bars in bottom two rows), and G3 is enriched for genes marking *Svet1* + intermediate progenitors (IPs) in E14 mouse subventricular zone<sup>43</sup>, including *ELAVL4*, *LRP8*, *NEUROD1*, *NRN1*, *SLC17A6*, *SSTR2* and *TP53INP1*. Modules G4–G6 are enriched in the SZo (Fig. 3d, grey bars); however, these modules are enriched for neuron-associated categories probably expressed in post-mitotic cells during their radial migration. To identify expression specific to SZo or SZi using a more targeted approach, we searched for genes both maximally expressed in the subventricular zone and differentially expressed between SZo and SZi (Fig. 3g;  $t > 4$ ,  $P < 0.01$ ,  $\log_2(\text{FC}) > 0.5$ ). Remarkably, few genes met both criteria at 15 or 16 pcw: 39 genes were enriched in SZi (including several in module G3), whereas only eight were specifically enriched in SZo.

These results are consistent with a previous study of laminar enrichment in 13–16 pcw human prenatal cortex, which also found few genes specifically enriched in SZo (55) or SZi (61)<sup>2</sup>.

### Species differences in subplate

The subplate is a largely transient zone beneath the cortical plate that plays an important role in establishment of thalamocortical connectivity (see ref. 35 for review). Subplate generation is protracted in primates<sup>8</sup>, and its thickness particularly expanded in human<sup>8</sup>. In mouse and other species<sup>18</sup> this layer is molecularly distinct, and our laminar profiling also identified many subplate-enriched genes in human (Fig. 2). For example, *NPY* is enriched in subplate at 21 pcw (but not 15 or 16 pcw) as shown both by microarray and *in situ* hybridization (Fig. 4a). To facilitate a comparative analysis, we identified a high confidence set of 150 subplate-enriched genes in human or mouse (Supplementary Table 8). Many genes showed similar enrichment in the developing mouse and human subplate, including the known subplate markers *KCNAB1* and *NR4A2* (refs 17, 18 and 44) (Fig. 4b). Several genes showed enrichment in developing human but not mouse subplate, including the hypocretin (orexin)



**Figure 4 | Common and distinct subplate markers in human and mouse.** **a**, *NPY* is enriched in subplate at 21 pcw but not 15 or 16 pcw based on microarray (left) and *in situ* hybridization (right). Microarray data are plotted as the average  $\pm$  standard error of the mean (s.e.m.) for each layer in each brain. **b**, Genes with subplate enrichment in both species, based on microarray data in human (top) and *in situ* hybridization data at embryonic (E) day E15.5 or E18.5 mouse (bottom). **c**, Genes with subplate enrichment in human but either no expression (*DKK1*, *HCRTR2* and *MESP1*) or no subplate enrichment (*CDH1* and *CRYM*) in mouse. **d**, Genes with subplate enrichment in developing mouse, but not human subplate. Asterisks indicate common expression between human and mouse in other layers (that is, subpial granular zone and pia mater). Mouse *in situ* hybridization images taken from the Allen Developing Mouse Brain Atlas.

receptor *HCRT2* (Fig. 4c and Extended Data Fig. 8a), which is thought to regulate sleep-wakefulness and is highly expressed in mouse hypothalamus<sup>45</sup>. Conversely, *Trh* and *Nxph4* show enriched expression in mouse but not human subplate (Fig. 4d and Extended Data Fig. 8b). *Trh* is also not expressed in the rat subplate<sup>18</sup>, suggesting this pattern is specific to mouse. These results indicate that the evolutionary elaboration of subplate in primates is associated with altered gene expression.

### Developmental gradients in neocortex

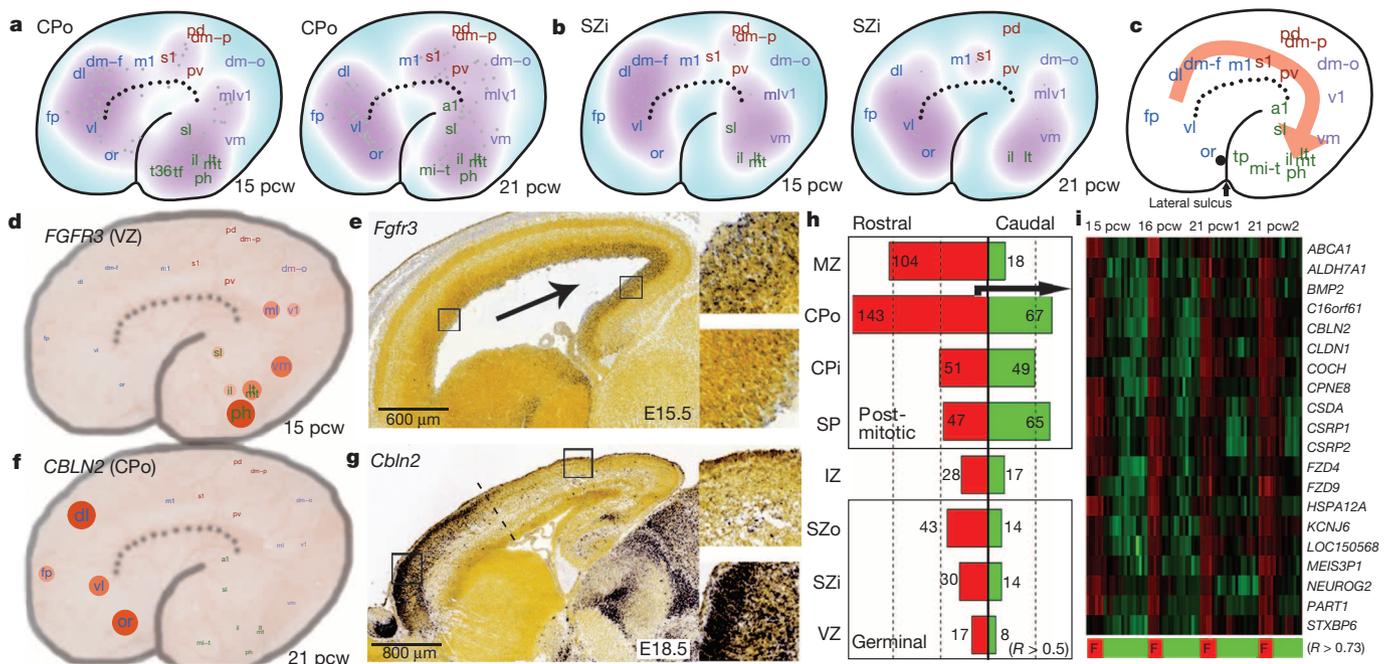
Cortical patterning is likely a result of intrinsic signalling, controlled in part by graded expression of transcription factors during early cortical development, followed by extrinsic signalling from thalamic afferents after the start of corticogenesis<sup>46–49</sup>. We sought to identify putative patterning centres, defined as regions where many genes show peak expression tapering off with distance, for each layer of the human prenatal cortex using an unbiased approach. To do so, we first assigned three-dimensional (3D) coordinates to each cortical sample, and then identified the location of maximum expression of the most graded genes in all four brains. In several layers, including CPo and SZi (Fig. 5a, b), most of these genes peaked in the frontal or temporal lobes. Rather than peaks in presumptive functional areas, this suggests a generally rostro-caudal organization axis that is better characterized as fronto-temporal, following the contour of the developing cortex. To identify such fronto-temporal patterning genes directly, we correlated gene expression in each cortical layer against the angular position of each neocortical region, as illustrated schematically in Fig. 5c. All layers contained gradient genes conserved across all four brains ( $P \sim 0$ , permutation analysis; FDR < 2.4%; Supplementary Methods), and each layer contained distinct sets of gradient genes, particularly when comparing germinal with post-mitotic layers (Supplementary Table 9). Gradient genes in the ventricular zone probably reflect intrinsic areal specification as the ventricular zone does not receive thalamic innervation. Our results indicate that rostral and caudal genes that were identified in grossly dissected cortex<sup>3</sup> mainly represent gradients in post-mitotic cells, as they show significant overlap

with gradient genes in marginal zone, cortical plate and intermediate zone (Extended Data Fig. 9).

Some features of areal patterning appear to be preserved between mouse and human. For example, *FGFR3*, which is known to cause defects in human temporal cortex when mutated<sup>50</sup>, shows significant caudal enrichment in all germinal layers in both species (Fig. 5d, e). Conversely, *CBLN2* shows rostral enrichment in human CPo (Fig. 5f) and in mouse (Fig. 5g), with an abrupt expression cutoff, implicating *CBLN2* in early rostral patterning. Comparing these data to a microarray analysis of rostral versus caudal cortex in E14 mouse<sup>51</sup> identified 20 additional genes with consistent rostrocaudal gradation between species (Supplementary Table 9 and Supplementary Methods). We find more rostrally than caudally enriched genes in nearly every cortical layer (Fig. 5h), whereas two studies of gradient expression in prenatal mouse brain identified more<sup>49</sup> or comparable<sup>51</sup> caudally enriched genes, indicating potential species differences related to areal patterning. This human frontal asymmetry is most apparent in the outer post-mitotic layers, the marginal zone and CPo (Fig. 5i) and in SZo, which generates most superficial cortical plate neurons in primate<sup>7,8</sup>, indicating that these genes may play a part in the expansion and reorganization of human prefrontal cortex<sup>8</sup>. Alternatively, this asymmetry could reflect temporal differences, as peak generation of excitatory neurons in visual cortex is delayed relative to frontal cortex<sup>52</sup>.

### Patterned expression of genes near haCNSs

Conserved non-coding sequences (CNSs) are genomic regions with exceptionally high similarity across divergent organisms, and therefore thought to be important for organism viability. CNSs are typically located by genes important for developmental regulation, and many show significant enhancer activity in brain<sup>5</sup>. Genes near CNSs with significantly accelerated rates of substitution in the human lineage (haCNSs)<sup>15</sup> are particularly likely to show differential expression between regions of developing human neocortex<sup>3,5</sup>, indicating transcriptional regulation by haCNSs may be important in human-specific neurodevelopment.



**Figure 5 | Areal patterning in the developing neocortex.** **a, b**, Density plot showing the location of highest expression (colour coded purple) for genes with gradient-like expression in CPo (**a**) or SZi (**b**) in each brain. **c**, Schematic illustrating the predominant direction of gene gradients, which follow a fronto-temporal trajectory. **d**, *FGFR3* shows caudal enrichment in germinal zones of the developing human cortex. Samples are plotted on a schematic of the prenatal cortex, with expression level indicated by circle size and colour.

**e**, Similar enrichment is seen in the developing mouse cortex. **f, g**, *CBLN2* shows rostral enrichment in cortical plate of both human (**f**) and mouse (**g**). **h**, Barplot showing common rostrally (red) and caudally (green) enriched genes across brains for each layer in humans. **i**, Heatmap representation of the top 20 rostrally enriched genes in the CPo shows selective enrichment in frontal (F) lobe (F, in bottom colour bar).

**Table 1 | Human-specific neurodevelopmental processes enriched for genes near haCNSs**

Category	Subcategory	O	O/E	Total (%)	P value
Areal	All	46	3.54	9.9	$2.56 \times 10^{-14}$
	Rostral (R)	34	4.24	11.8	$2.60 \times 10^{-13}$
	Caudal (C)	14	2.77	7.7	$1.78 \times 10^{-4}$
	Post-mitotic (P)	35	3.37	9.4	$1.01 \times 10^{-14}$
	IZ	11	8.80	24.4	$2.34 \times 10^{-9}$
	Germinal (G)	7	2.68	7.4	$4.65 \times 10^{-3}$
Network	C31 (interneuron)	18	5.31	14.8	$1.15 \times 10^{-9}$
	C16 (subplate)	25	3.98	11.1	$9.80 \times 10^{-10}$

O, observed; E, expected. Areal genes include all genes in Fig. 5h. No other layers (Fig. 2) or modules (Fig. 3) were significantly enriched for genes near haCNSs.

Our results confirm and extend these findings (Table 1 and Supplementary Table 9). Rostrally enriched genes include significantly more haCNSs than caudally enriched genes, consistent with the expanded frontal cortex in primates and the developmental role of haCNSs. We also find more haCNSs with areal expression patterns in post-mitotic than germinal layers, even after accounting for the larger number of areal genes in post-mitotic layers. Surprisingly, nearly 25% of regional genes in the intermediate zone (11/45) were haCNSs. This result, which cannot be explained by over-representation of neural adhesion genes in the intermediate zone, suggests that the intermediate zone is of particular importance in areal cortical identity during human development<sup>53</sup>.

Finally, we assessed the distribution of haCNSs in modules from our whole cortex network (Fig. 3a–c). Only two of the 42 tested modules were enriched for haCNSs: the interneuron-related (C31) and subplate-enriched (C16) modules, which both mark processes with features potentially distinct to the human lineage. Interestingly, *FOXP2*, implicated in the specialization of language areas<sup>54</sup>, is included in module C16, and shows enrichment in parietal and temporal lobes including presumptive language areas (Extended Data Fig. 10). No layers other than subplate were significantly enriched for haCNSs (data not shown). Together these results support the hypothesis that transcriptional networks underlying the evolution of human neocortex can at least partly be traced to haCNSs.

## Discussion

Studies of the developing human brain are essential for elucidating the details of human brain formation, function and evolutionary differences, and for understanding developmental mechanisms underlying neurodevelopmental disorders such as autism<sup>39,40</sup> and schizophrenia<sup>19</sup>. The atlas of the mid-gestational human brain described here, part of the BrainSpan Atlas of the Developing Human Brain, builds on digital molecular brain atlasing efforts in mouse<sup>20,22,24</sup> and adult human<sup>23</sup> by providing transcriptome resources on prenatal specimens typically inaccessible for research. Several recent studies have assayed a limited set of brain structures<sup>1,3–5</sup> and layers<sup>2</sup> from prenatal human brain. In contrast, the current project aimed for anatomical comprehensiveness at a fine nuclear/laminar level, albeit with a small number of specimens. This degree of specificity necessitated using available methods for small sample amplification and DNA microarrays (the same platform recently used for adult human<sup>23</sup>), but newer techniques may soon allow moving to the resolution of single cells using RNA sequencing for complete transcriptome coverage<sup>55</sup>.

Many differences in cortical development between human, non-human primate and rodent have been documented, including an expanded subplate zone<sup>8</sup> and subpial granular zone<sup>9</sup>, expansion of association areas particularly in frontal lobe<sup>8</sup>, expansion of superficial layers that greatly increase the extent of cortico-cortical connections, and the appearance of a secondary proliferative zone, the SZo, that probably allow the massive expansion of human cortex<sup>6,7</sup>. We find transcriptional features related to each of these anatomical features, although we were able to identify only minimal molecular differences between the inner and outer subventricular zones, leaving open the question of what distinguishes this

primate-specific zone of cortical precursors. These data also provide a powerful map to pin an anatomical and developmental locus on genes related to neurodevelopmental disease origins and human-specific brain function and evolution. Although the current analyses only scratch the surface, these data will be extremely useful for generating and testing new hypotheses about molecular substrates for specific features of human brain development and function.

## METHODS SUMMARY

Anatomically comprehensive transcriptional profiling of prenatal human brains used high-throughput tissue processing and data generation pipelines for post-mortem brain anatomical delineation, sample isolation and microarray analysis. Data visualization and mining tools were developed to create a publicly accessible data resource (<http://www.brainspan.org/>).

**Post-mortem tissue acquisition and screening.** Tissue was provided by the Birth Defects Research Laboratory (BDRL) at the University of Washington and Advanced Bioscience Resources (ABR; Alameda, California). Tissue with no known history of maternal drug or alcohol abuse, potential teratogenic events, or HIV1/2, HepB or HepC infection, was eligible for inclusion in the study. Due to confidentiality standards of tissue providers, it is impossible to rule out potential neuropathologies in this tissue; however, no neuropathological defects were observed in histological data derived from these tissues. Eligible tissue was screened to ensure cytoarchitectural integrity (analysis of Nissl-stained sections) and high RNA quality. RNA integrity numbers (RINs) ranged from 8.2–9. All work was performed following guidelines for the research use of human brain tissue (ABR) or the UAGA and NOTA guidelines for acquisition and distribution of human tissue for bio-medical research purposes (BDRL) and with approval by the Human Investigation Committees and Institutional Ethics Committees of each institute. Appropriate written informed consent was obtained and all available non-identifying information was recorded for each sample. Four specimens were used for microarray profiling (Supplementary Table 1).

**Sample processing.** For each specimen, the frozen left hemisphere was cut into ~2.0-cm-thick slabs. Slabs were sectioned to allow histological staining, as shown in Fig. 1d. Total RNA was isolated and microarray data were generated by Covance Genomics Laboratory (Seattle, Washington) on Agilent 8 × 60K custom design arrays (AMADID no. 024915)<sup>20</sup>.

**Microarray data analysis.** Data were preprocessed by removing outliers and then normalizing. Differential expression was assessed using correlation to a template vector or by measuring fold change. WGCNA was performed as previously described<sup>56</sup>. Lists were characterized using gene ontology and cell type enrichments. Enrichment of haCNSs was quantified using hypergeometric tests.

**In situ hybridization.** *In situ* hybridization used a semi-automated non-isotopic technology platform as previously described<sup>24</sup>, with some modifications such as a reduction in proteinase K concentration. Riboprobes were designed to overlap probe designs for homologous mouse genes in the Allen Developing Mouse Brain Atlas (<http://developingmouse.brain-map.org/>). FISH was performed as previously described<sup>56</sup>.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 March 2013; accepted 26 February 2014.

Published online 2 April 2014.

- Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523 (2011).
- Fietz, S. A. *et al.* Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proc. Natl Acad. Sci. USA* **109**, 11836–11841 (2012).
- Johnson, M. B. *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009).
- Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
- Lambert, N. *et al.* Genes expressed in specific areas of the human fetal cerebral cortex display distinct patterns of evolution. *PLoS ONE* **6**, e17753 (2011).
- Kriegstein, A., Noctor, S. & Martínez-Cerdeño, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nature Rev. Neurosci.* **7**, 883–890 (2006).
- Hansen, D. V., Lui, J., Parker, P. & Kriegstein, A. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
- Smart, I., 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* **12**, 37–53 (2002).

9. Brun, A. The subpial granular layer of the foetal cerebral cortex in man. Its ontogeny and significance in congenital cortical malformations. *Acta Pathologica et Microbiologica Scandinavica* **179** (suppl.), 3–98 (1965).
10. Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. Interneuron migration from basal forebrain to neocortex: dependence on *Dlx* genes. *Science* **278**, 474–476 (1997).
11. Ma, T. *et al.* Subcortical origins of human and monkey neocortical interneurons. *Nature Neurosci.* **16**, 1588–1597 (2013).
12. Yu, X. & Zecevic, N. Dorsal radial glial cells have the potential to generate cortical interneurons in human but not in mouse brain. *J. Neurosci.* **31**, 2413–2420 (2011).
13. Letinic, K., Zoncu, R. & Rakic, P. Origin of GABAergic neurons in the human neocortex. *Nature* **417**, 645–649 (2002).
14. Hansen, D. V. *et al.* Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nature Neurosci.* **16**, 1576–1587 (2013).
15. Prabhakar, S., Noonan, J., Pääbo, S. & Rubin, E. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**, 786 (2006).
16. Ayoub, A. E. *et al.* Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. *Proc. Natl Acad. Sci. USA* **108**, 14950–14955 (2011).
17. Montiel, J. *et al.* Hypothesis on the dual origin of the mammalian subplate. *Front. Neuroanat.* **5**, 25 (2011).
18. Wang, W. Z. *et al.* Comparative aspects of subplate zone studied with gene expression in sauropsids and mammals. *Cereb. Cortex* **21**, 2187–2203 (2011).
19. Insel, T. R. Rethinking schizophrenia. *Nature* **468**, 187–193 (2010).
20. Paxinos, G. & Franklin, K. *Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates* 4th edn (Academic Press, 2012).
21. Saleem, K. & Logothetis, N. *A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates* 2nd edition (Academic Press, 2012).
22. Dong, H. W. *Allen Reference Atlas: a Digital Color Brain Atlas of the C57BL/6J Male Mouse* (Wiley, 2008).
23. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
24. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
25. Bernard, A. *et al.* Transcriptional architecture of the primate neocortex. *Neuron* **73**, 1083–1099 (2012).
26. Steinfeld, R. *et al.* Folate receptor alpha defect causes cerebral folate transport deficiency: a treatable neurodegenerative disorder associated with disturbed myelin metabolism. *Am. J. Hum. Genet.* **85**, 354–363 (2009).
27. Dubourg, C. *et al.* Molecular screening of *SHH*, *ZIC2*, *SIX3*, and *TGIF* genes in patients with features of holoprosencephaly spectrum: mutation review and genotype–phenotype correlations. *Hum. Mutat.* **24**, 43–51 (2004).
28. Huang, H. *et al.* Anatomical characterization of human fetal brain development with diffusion tensor magnetic resonance imaging. *J. Neurosci.* **29**, 4263–4273 (2009).
29. Bystron, I., Blakemore, C. & Rakic, P. Development of the human cerebral cortex: Boulder Committee revisited. *Nature Rev. Neurosci.* **9**, 110–122 (2008).
30. Vogt Weisenhorn, D. M., Prieto, E. W. & Celio, M. R. Localization of calretinin in cells of layer I (Cajal–Retzius cells) of the developing cortex of the rat. *Brain Res. Dev. Brain Res.* **82**, 293–297 (1994).
31. Englund, C. *et al.* Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J. Neurosci.* **25**, 247–251 (2005).
32. Levitt, P. & Rakic, P. Immunoperoxidase localization of glial fibrillary acidic protein in radial glial cells and astrocytes of the developing rhesus monkey brain. *J. Comp. Neurol.* **193**, 815–840 (1980).
33. Grinberg, I. *et al.* Heterozygous deletion of the linked genes *ZIC1* and *ZIC4* is involved in Dandy–Walker malformation. *Nature Genet.* **36**, 1053–1055 (2004).
34. Inoue, T., Ogawa, M., Mikoshiba, K. & Aruga, J. *Zic* deficiency in the cortical marginal zone and meninges results in cortical lamination defects resembling those in type II lissencephaly. *J. Neurosci.* **28**, 4712–4725 (2008).
35. Kostović, I. & Judas, M. The development of the subplate and thalamocortical connections in the human foetal brain. *Acta Paediatr.* **99**, 1119–1127 (2010).
36. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).
37. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature Neurosci.* **11**, 1271–1282 (2008).
38. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
39. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
40. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
41. Gaspard, N. & Vanderhaeghen, P. Laminar fate specification in the cerebral cortex. *F1000 Biol. Rep.* **3**, 6 (2011).
42. Betizeau, M. *et al.* Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442–457 (2013).
43. Kawaguchi, A. *et al.* Single-cell gene profiling defines differential progenitor subclasses in mammalian neurogenesis. *Development* **135**, 3113–3124 (2008).
44. Butler, D. M., Ono, J. K., Chang, T., McCaman, R. E. & Barish, M. E. Mouse brain potassium channel  $\beta 1$  subunit mRNA: cloning and distribution during development. *J. Neurobiol.* **34**, 135–150 (1998).
45. Gerashchenko, D. *et al.* Hypocretin-2-saporin lesions of the lateral hypothalamus produce narcoleptic-like sleep behavior in the rat. *J. Neurosci.* **21**, 7273–7283 (2001).
46. Mallamaci, A. & Stoykova, A. Gene networks controlling early cerebral cortex arealization. *Eur. J. Neurosci.* **23**, 847–856 (2006).
47. O'Leary, D. D. Do cortical areas emerge from a protocortex? *Trends Neurosci.* **12**, 400–406 (1989).
48. Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).
49. Sansom, S. N. & Livesey, F. Gradients in the brain: the control of the development of form and function in the cerebral cortex. *Cold Spring Harb. Perspect. Biol.* **1**, a002519 (2009).
50. Hevner, R. F. The cerebral cortex malformation in thanatophoric dysplasia: neuropathology and pathogenesis. *Acta Neuropathol.* **110**, 208–221 (2005).
51. Pinto, L. *et al.* AP2 $\gamma$  regulates basal progenitor fate in a region- and layer-specific manner in the developing cortex. *Nature Neurosci.* **12**, 1229–1237 (2009).
52. Rakic, P. Neurogenesis in adult primates. *Prog. Brain Res.* **138**, 3–14 (2002).
53. Allendoerfer, K. L. & Shatz, C. J. The subplate, a transient neocortical structure: its role in the development of connections between thalamus and cortex. *Annu. Rev. Neurosci.* **17**, 185–218 (1994).
54. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
55. Qiu, S. *et al.* Single-neuron RNA-seq: technical feasibility and reproducibility. *Front. Genet.* **3**, 124 (2012).
56. Sorensen, S. A. *et al.* Correlated gene expression and target specificity demonstrate excitatory projection neuron diversity. *Cereb. Cortex* <http://dx.doi.org/10.1093/cercor/bht243> (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We wish to thank the Allen Institute founders, P. G. Allen and J. Allen, for their vision, encouragement and support. We express our gratitude to past and present Allen Institute staff members R. Adams, A. Alpisa, A. Boe, E. Byrnes, M. Chapin, J. Chen, C. Copeland, N. Dotson, K. Fotheringham, E. Fulfs, M. Gasparini, T. Gilbert, Z. Haradon, N. Hejazinia, N. Ivanov, J. Kinnunen, A. Kriedberg, J. Laoenkue, S. Levine, V. Menon, E. Mott, N. Motz, J. Pendergraft, L. Potekhina, J. Redmayne-Titley, D. Rosen, C. Simpson, S. Shi, L. Velasquez, U. Wagley, N. Wong and B. Youngstrom for their technical assistance. We would also like to thank J. Augustinack, T. Benner, A. Mayaram, M. Roy, A. van der Kouwe and L. Wald from the Fischl laboratory. Also, we wish to acknowledge Covance Genomics Laboratory (Seattle, Washington) for microarray probe generation, hybridization and scanning. In addition, we express our gratitude to Advanced Bioscience Resources, for providing tissue used for expression profiling and reference atlas generation as well as to the Laboratory of Developmental Biology, University of Washington, for providing tissue used for expression profiling and reference atlas generation. The Laboratory of Developmental Biology work was supported by the National Institutes of Health (NIH) Award Number 5R24HD0008836 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development. The BrainSpan project was supported by Award Number RC2MH089921 (Pis: E. Lein and M. Hawrylycz, Allen Institute for Brain Science) from the National Institute of Mental Health. The content is solely the responsibility of the respective authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

**Author Contributions** E.S.L., S.-L.D., K.A.S. and S.M.S. contributed significantly to the overall project design. S.M.S., K.A.S., A.E., A.B., and P.W. managed the tissue and sample processing in the laboratory. K.A., J.M.A., C.B., D.B., K.B., S.B., S.C., A.C., C.C., R.A.D., G.G., J.G., L.G., B.W.G., R.E.H., T.A.L., Na.M., N.F.M., N.-K.N., A.O., E.O., J.Pa., P.D.P., S.E.P., M.P., Me.R., J.J.R., K.R., D.S., Me.S., S.S., N.V.S. and Mi.S. contributed to tissue and sample processing. E.H.S., Z.L.R., T.N.-C., and I.A.G. contributed to establishing the tissue acquisition pipeline. N.D., J.N. and A.B. contributed to protocol development. A.S.P., L.Z., B.F., and H.H. contributed to MR and DWI imaging and analysis. J.M.J., C.R.S., and D.W. provided engineering support. S.-L.D., R.A.D., P.D.P., D.S. and J.G.H. contributed to the neuroanatomical design and implementation. S.-L.D., B.A.C.F., Ph.L., B.M., J.J.R., R.F.H., N.Se. and J.G.H. contributed to the reference atlas design, quality control and implementation. L.N., A.S. and C.D. managed the creation of the data pipeline, visualization and mining tools. L.N., A.S., T.A.D., D.F., T.P.F., G.Gu, C.L.K., C.La., F.L., N.Sj. and A.J.S. contributed to the creation of the data pipeline, visualization and mining tools. J.A.M., S.-L.D., R.F.H., C.-K.L., M.J.H., S.M.S. and E.S.L. contributed to data analysis and interpretation. M.B.G., D.H.G., J.A.K., Pa.L., J.W.P., N.Se. and A.R.J. contributed to overall project design and consortium management. E.S.L. and M.J.H. conceived the project, and the manuscript was written by J.A.M. and E.S.L. with input from all other authors.

**Author Information** These data are freely accessible as part of the BrainSpan Atlas of the Developing Human Brain (<http://brainspan.org>), also available via the Allen Brain Atlas data portal (<http://www.brain-map.org>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.L. ([edl@alleninstitute.org](mailto:edl@alleninstitute.org)).

## METHODS

**Post-mortem tissue acquisition and screening.** Tissue was provided by the Birth Defects Research Laboratory (BDRL) at the University of Washington and Advanced Bioscience Resources (ABR; Alameda, California), who obtained appropriate written informed consent and provided available non-identifying information for each sample. Tissue with no known history of maternal drug or alcohol abuse, potential teratogenic events, or HIV1/2, HepB or HepC infection, was eligible for inclusion in the study. Due to confidentiality standards of tissue providers, it is impossible to rule out potential neuropathologies in this tissue; however, no neuropathological defects were observed in histological data derived from these tissues. Eligible tissue was screened to ensure cytoarchitectural integrity (analysis of Nissl-stained sections) and high RNA quality. RNA integrity numbers (RINs) ranged from 8.2–9. All work was performed according to guidelines for the research use of human brain tissue (ABR) or the UAGA and NOTA guidelines for the acquisition and distribution of human tissue for bio-medical research purposes (BDRL) and with approval by the Human Investigation Committees and Institutional Ethics Committees of each institute from which samples were obtained. Specimens for microarray profiling consisted of two 21 pcw females, one 15 pcw male and one 16 pcw female (Supplementary Table 1).

**Laser microdissection and RNA isolation.** Slabs from the frozen brains were serially cryosectioned at 14  $\mu\text{m}$  onto PEN slides for LMD (Leica Microsystems) and a 1:10 Nissl series was generated for neuroanatomical reference. After drying for 30 min at room temperature, PEN slides were frozen at  $-80^\circ\text{C}$ . Slides were later rapidly fixed in ice-cold 70% ethanol, lightly stained with cresyl violet to allow cytoarchitectural visualization, dehydrated and frozen at  $-80^\circ\text{C}$ . LMD was performed on a Leica LMD6000 (Leica Microsystems) using the cresyl violet stain to identify target brain regions. Samples captured include cortical and subcortical regions and are listed for each brain in the ontological sample map (Supplementary Table 2).

Microdissected tissue was collected directly into RLT buffer from the RNeasy Micro kit (Qiagen) supplemented with  $\beta$ -mercaptoethanol. Samples were volume adjusted with RLT Buffer to 75  $\mu\text{l}$ , vortexed, centrifuged and frozen at  $-80^\circ\text{C}$ . RNA was isolated for each brain region following the manufacturer's instructions. RNA samples were eluted in 14  $\mu\text{l}$ , and 1  $\mu\text{l}$  was run on the Agilent 2100 Bioanalyzer (Agilent Technologies) using the Pico 6000 assay kit. Samples were quantitated using the Bioanalyzer concentration output. The average RNA Integrity Number (RIN) of all 1,202 passed experimental samples was 6.3.

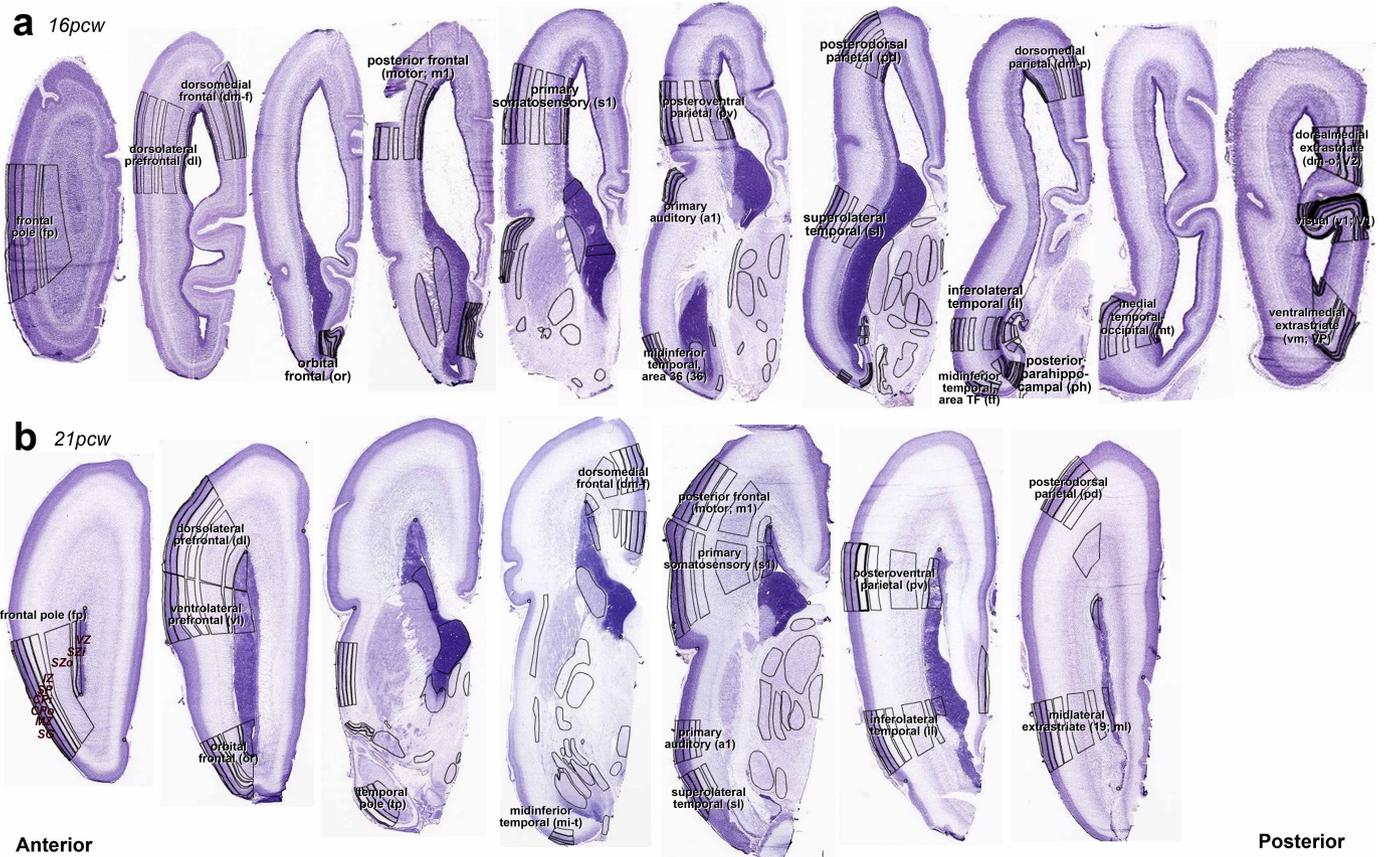
**mRNA profiling.** Sample amplification, labelling, and microarray processing were performed by the Covance Genomics Laboratory (Seattle). Briefly, samples were amplified using a custom two-cycle RT-IVT amplification protocol. For each sample, 5 ng of total RNA was mixed with 250 ng of pBR322 (Life Technologies) to act as a carrier. The MessageAmp II aRNA Amplification Kit was used for the first round of amplification and the Amino Allyl MessageAmp II aRNA Amplification Kit for the second round of amplification (Life Technologies). Following amplification, 5  $\mu\text{g}$  of cRNA was labelled with Cy3 mono-Reactive Dye (GE Healthcare). Each labelled aRNA was resolved using a Bioanalyzer with RNA 6000 Nano kit reagents (Agilent Technologies) before hybridization. Samples were evaluated for yield and size distribution, then normalized to 600 ng input, fragmented, and hybridized to Agilent human  $8 \times 60\text{K}$  arrays. Gene expression data quality was assessed using standard Agilent quality control metrics. To control for batch effects, common RNA pool control samples were amplified and hybridized in each batch. A total of 1,225 samples passed sample quality control (QC), including 1,202 experimental samples and 23 control samples. The data discussed in this publication are accessible through the Allen Brain Atlas data portal (<http://www.brain-map.org>) or directly at (<http://www.brainspan.org>).

**Microarray data analysis.** All microarray data was subjected to QC and ERCC spike-in assessments, and any failing samples were omitted from the analysis.

Biological outliers were identified by comparing samples from related structures using hierarchical clustering and inter-array correlation measures. Data for samples passing QC were normalized in three steps: (1) 'within-batch' normalization to the 75th percentile expression values; (2) 'cross-batch' bias reduction using ComBat<sup>57</sup>; and (3) 'cross-brain' normalization as in step 1. Differential expression assessments were done using template vector correlation, where 1 = 'in group' and 0 = 'not in group', or by measuring the fold change, defined as mean expression in category divided by mean expression elsewhere. False discovery rates were estimated using permutation tests (Supplementary Methods). WGCNA was performed on all neo-cortical samples using the standard method<sup>56,58</sup>, and on germinal layers by defining a consensus module in the 15 and 16 pcw brains<sup>59</sup>, only including genes differentially expressed across these layers (5,494 genes; ANOVA  $P < 0.01$ , Benjamini-Hochberg adjusted). Gene list characterizations were made using a combination of module eigengene/representative gene expression, gene ontology enrichment using DAVID<sup>60</sup>, and enrichment for known brain-related categories<sup>61,62</sup> using the R function userListEnrichment<sup>63</sup>. Module C31 is depicted using VisANT<sup>64</sup>: the top 250 gene-gene connections based on topological overlap are shown, with histone genes removed for clarity. Rostral-caudal areal gradient genes were identified as follows: first, the centre of each neocortical region was identified at 21 pcw in Euclidean coordinates; second, the rostral/caudal region position was estimated as an angle along the lateral face of the brain centred at the temporal/frontal lobe juncture (ordering lobes roughly as frontal, parietal, occipital, temporal; Fig. 5c); third, for each brain gene expression in each layer was (Pearson) correlated with this region position; and finally, genes with  $R > 0.5$  in all four brains were identified. A similar strategy was used to identify unbiased areal gradient genes (Supplementary Methods). Enrichment of haCNSs was determined using hypergeometric tests. Samples in all plots are ordered in an anatomically relevant manner. Unless otherwise noted, all  $P$  values are Bonferroni corrected for multiple comparisons.

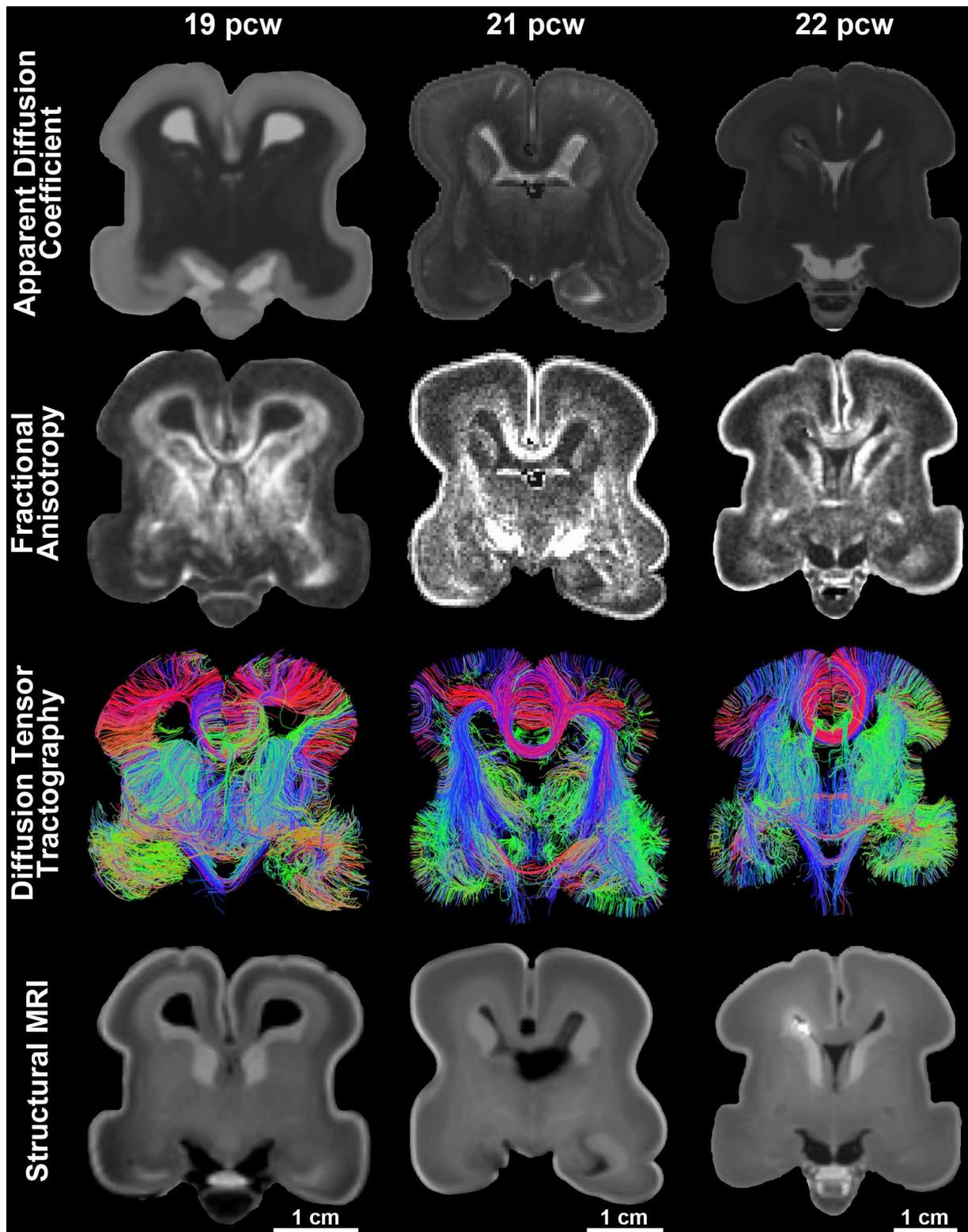
**In situ hybridization.** Nonisotopic colorimetric *in situ* hybridization was performed as described previously<sup>24</sup> with some modifications such as a reduction in proteinase K concentration. Briefly, following cryosectioning of fresh-frozen samples at 20  $\mu\text{m}$ , tissue sections were fixed, acetylated and subsequently dehydrated. Digoxigenin-based riboprobe labelling coupled with TSA amplification and alkaline-phosphatase-based colorimetric detection was used to label target mRNAs in expressing cells. Riboprobes were designed to overlap probe designs for homologous genes in mouse in the Allen Developing Mouse Brain Atlas (<http://developingmouse.brain-map.org/>). FISH was run as previously described<sup>56</sup> except that high resolution images were captured on an Olympus Fluoview 1000 Confocal Microscope at  $\times 60$  magnification.

57. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
58. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
59. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1**, 54 (2007).
60. Huang, D. W., Sherman, B. & Lempicki, R. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
61. Cahoy, J. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
62. Sugino, K. *et al.* Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature Neurosci.* **9**, 99–107 (2006).
63. Miller, J. A. *et al.* Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics* **12**, 322 (2011).
64. Hu, Z., Mellor, J., Wu, J. & DeLisi, C. VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics* **5**, 17 (2004).



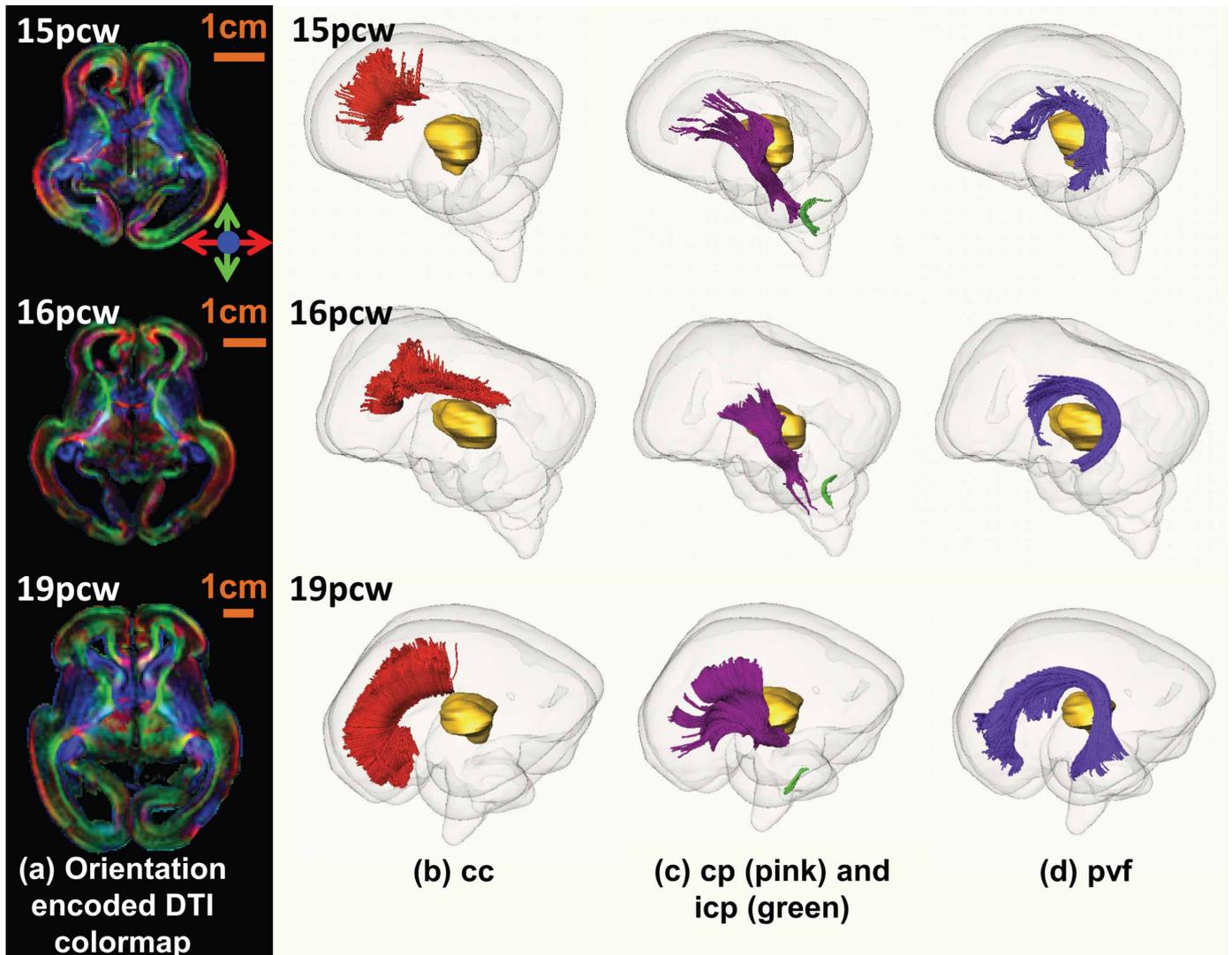
**Extended Data Figure 1 | Representative Nissl sections for laser microdissection (LMD) of 16 pcw and 21 pcw brains. a, b,** Nissl-stained sections were annotated and used to determine LMD region boundaries for 16 pcw (a) and 21 pcw (b) brains. Regions from adjacent sections on PEN

membrane slides were captured using these annotations as guidelines. Labels show full name and abbreviation for representative planes of section through presumptive neocortical regions. Panel b is a higher resolution modified version of the bottom row in Fig. 1d of the main manuscript.



**Extended Data Figure 2 | Overview of magnetic resonance imaging data acquired from the post-mortem, formalin-fixed human fetal brain samples.** Diffusion-weighted MRI were acquired for each sample using a steady state free precession sequence ( $b = 730$  s per  $\text{mm}^2$ ; 44 directions), yielding maps of apparent diffusion coefficient (ADC) (first row) and of fractional anisotropy

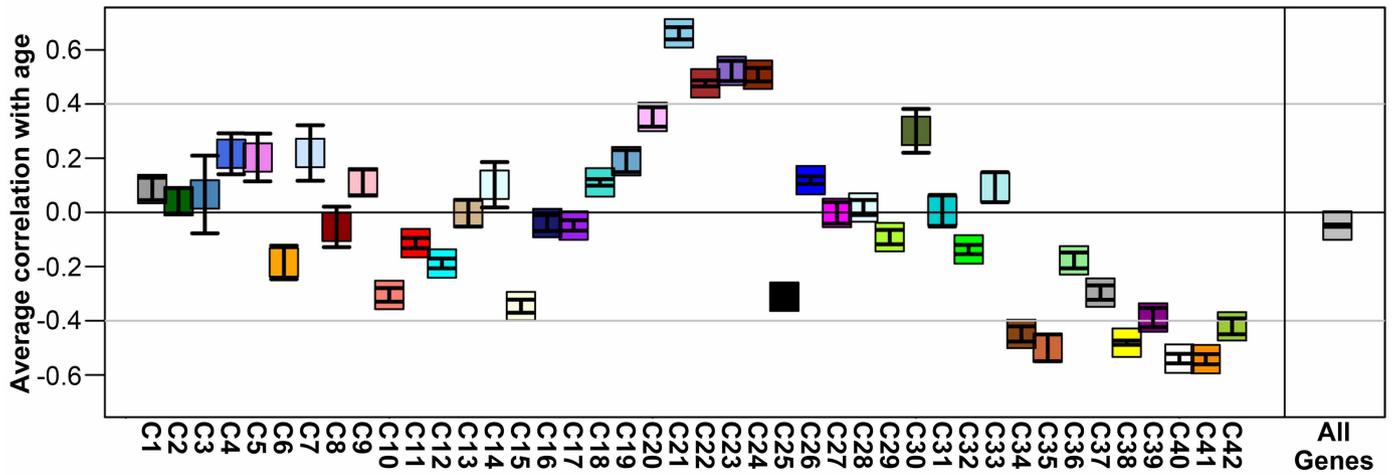
(FA) (second row). Whole-brain deterministic tractography results (third row) represent visualization of tractography output data filtered by a coronal slice filter. Structural data were acquired for each sample using a multi-echo flash sequence with images acquired at  $\alpha = 40$  providing optimal contrast to identify cortical and subcortical structures of interest (fourth row).



**Extended Data Figure 3 | White matter fibre tracts in fetal human brain.** a–d, Orientation-encoded diffusion tensor imaging (DTI) colormaps in the left panel (a) and the three-dimensionally reconstructed fetal white matter fibres in the right panels (b–d) for fetal brains at 15 pcw (upper row), 16 pcw (middle row) and 19 pcw (lower row). The orientation-encoded DTI colormaps are in axial planes at the anterior commissure level. The red, pink, green and purple fibres in the right panels are cc in (b), cp and icp in (c) and pvf in (d),

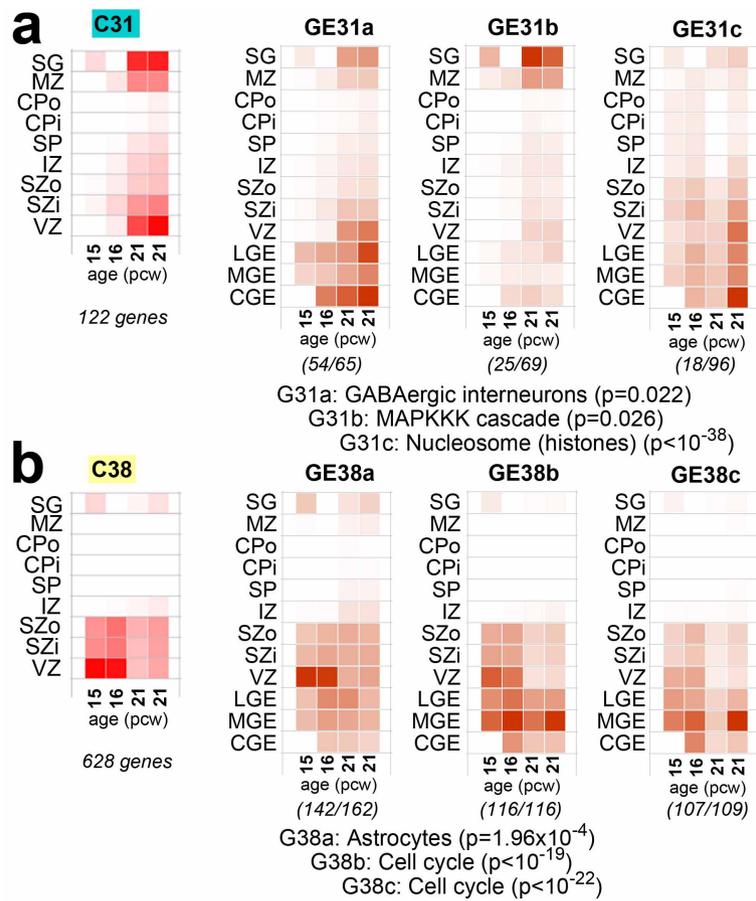
respectively. The transparent whole brain and yellow thalamus are also shown as anatomical guidance in (b–d). The scale bars are shown in the left panel (a). The fibre name abbreviations are: cc, corpus callosum; cp, cerebral peduncle; icp, inferior cerebellar peduncle; pvf, periventricular fibres (transient fibres coursing around the germinal matrix and only existing in the prenatal brain).





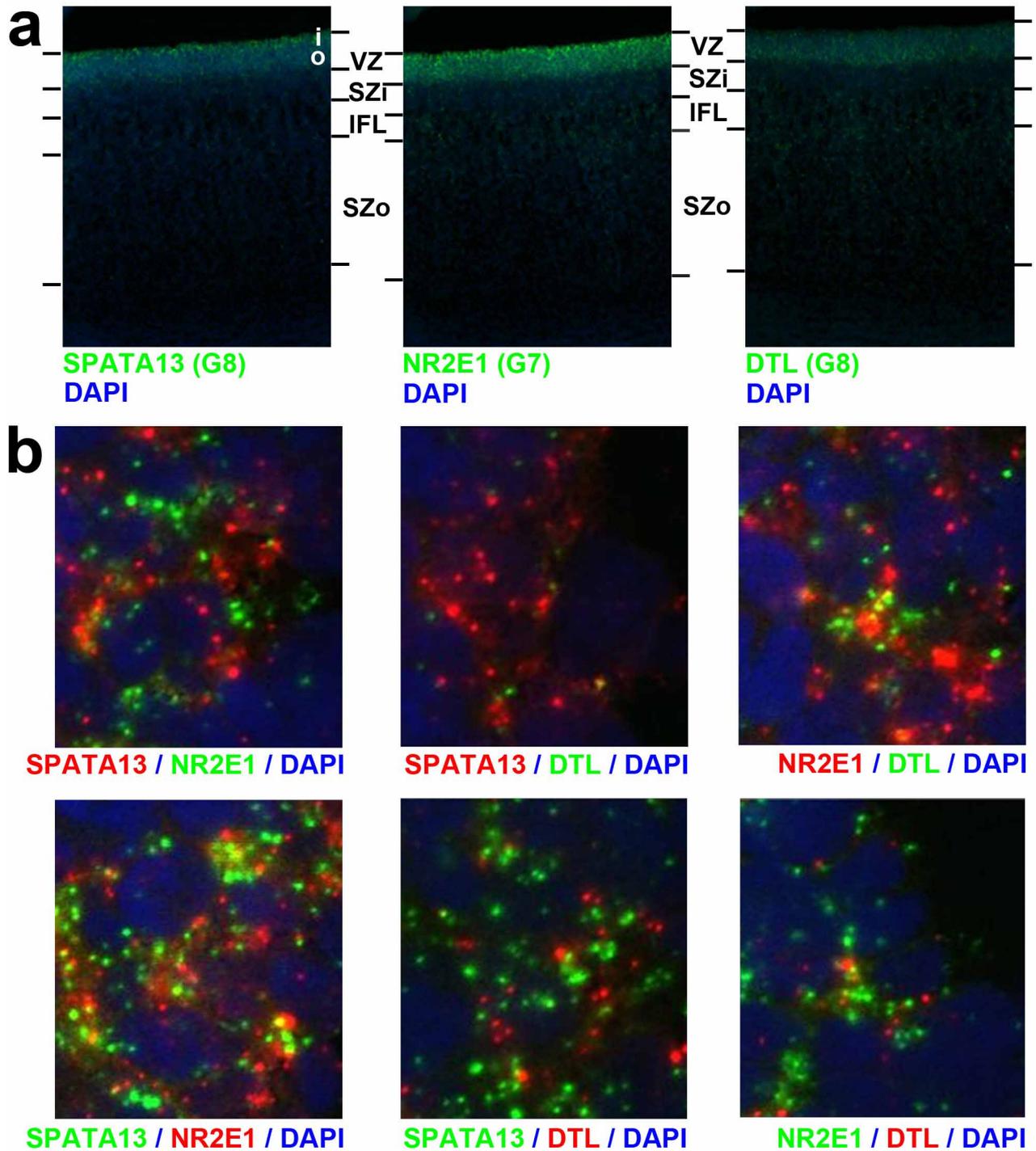
**Extended Data Figure 5 | Temporal patterning of whole cortex WGCNA modules across early to mid-gestational periods in BrainSpan RNA-seq cortical data.** RNA-seq RPKM values for 8–22 pcw specimens in the BrainSpan database for genes assigned to WGCNA modules (Fig. 3 in main manuscript) were correlated with age. For each module (Fig. 3a–c; x axis), the average correlation ( $\pm$  standard error of the mean) between expression of genes

in that module and age (y axis) is plotted. Many of the modules show increases (positive correlation) or decreases (negative correlation) with age. In particular, modules C38 (decreasing with age) and C22 (increasing with age) presented in the main manuscript (see Fig. 3b, left column) show consistent trends with age in both data sets.



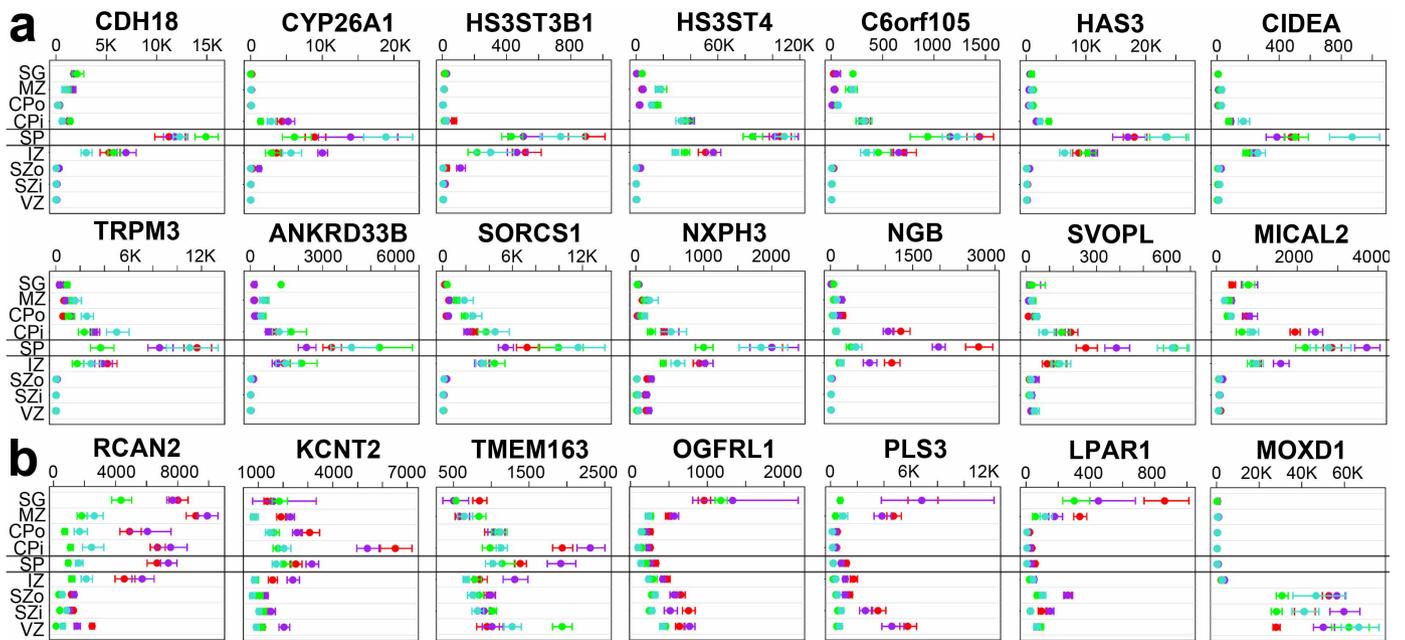
**Extended Data Figure 6 | Gene sets corresponding to GABAergic interneurons and proliferating layers also are highly expressed in the ganglionic eminences.** To examine the relationship between genes enriched in the cortical ventricular zone, including gene modules associated with GABAergic interneurons and mitotically active proliferative cells, WGCNA was performed on the combined cortical and GE samples (referred to as the 'GE network'). **a**, Genes from module C31 in the whole cortex WGCNA (GABAergic interneurons) are assigned primarily to three modules in the GE network. GE31a has a similar pattern in cortex as C31, is highly expressed in GE and is enriched in genes associated with GABAergic interneurons. Other genes from C31 were assigned to modules with other cortical patterns and

functional ontological associations (GE31b, GE31c). **b**, Genes from module C38 in the whole cortex WGCNA also divide primarily into three GE modules that are enriched in both the cortical germinal layers and the GE. These modules are enriched for genes expressed in astrocytes, potentially reflecting expression in radial glia, and are associated with cell cycle. For all plots, module eigengene (ME) expression is averaged across brain and layer (as in Fig. 3b), also including LGE, MGE and CGE (referred to here collectively as GE). Numbers in parentheses below each plot show the number of genes from module C31 in **a**, or C38 in **b**, out of the total number of module genes in the newly generated network. One representative enrichment category for each module is shown with enrichment  $P$  value.



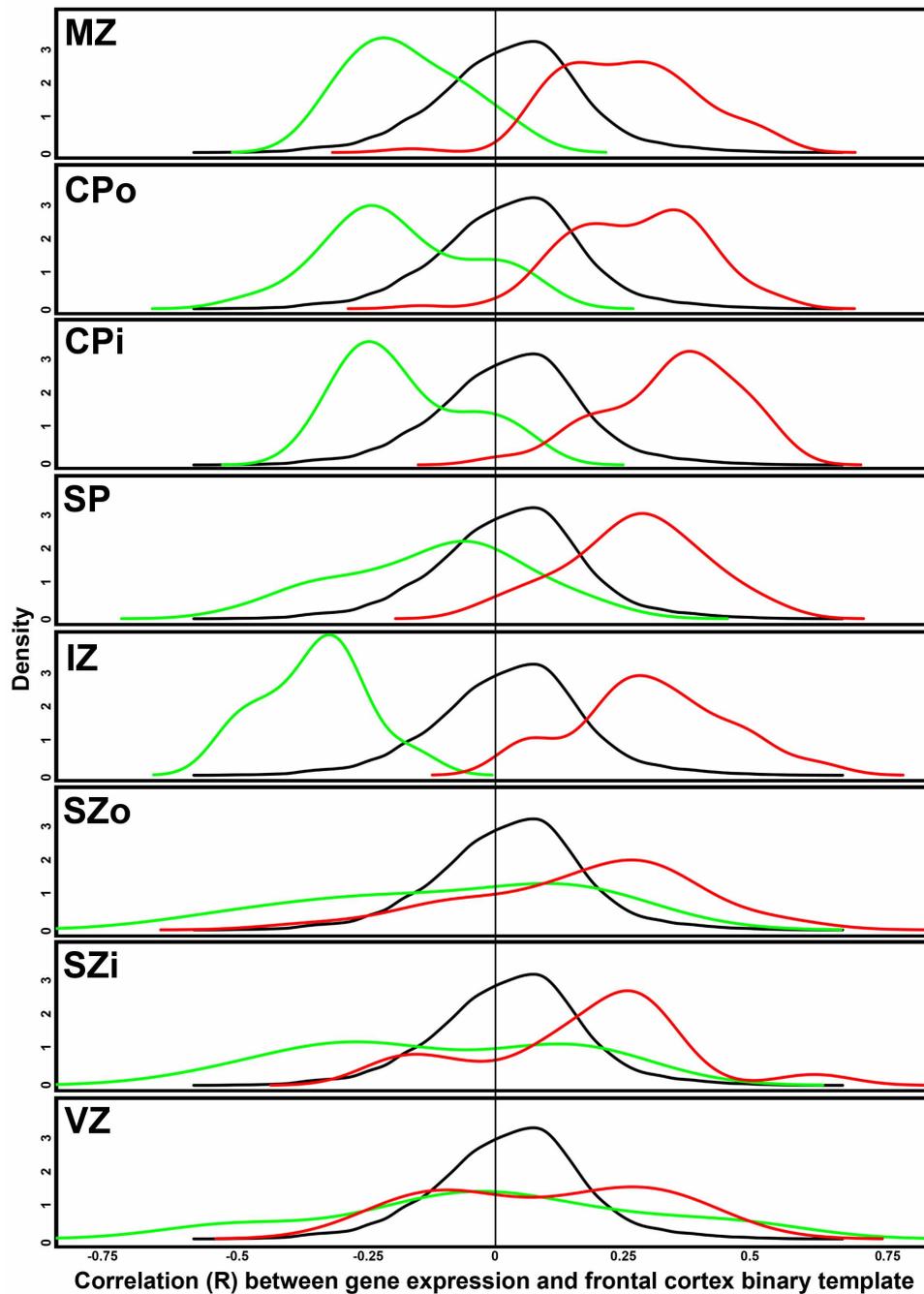
**Extended Data Figure 7 | FISH of hub genes in ventricular-zone-enriched modules shows expected laminar enrichment and largely non-overlapping subcellular distributions.** **a**, Fluorescent *in situ* hybridization (FISH) in proliferative layers of 15 pcw human cortex for three genes in modules G7 and G8 in the germinal layers network shown in Fig. 3 of the manuscript (see Fig. 3d–f)—SPATA13, NR2E1 and DTL. All three genes show enrichment in the ventricular zone compared to the subventricular zone as expected based on

microarray data. Nuclei are labelled with DAPI (blue). **b**, High-magnification images in the ventricular zone show double labelling for each pair of genes (with fluor reversal, lower row) and show complex subcellular distributions. SPATA13, NR2E1 and (to a lesser extent) DTL appear to be expressed in most cells in the ventricular zone, but these genes are typically expressed in non-overlapping punctate cytoplasmic locations (excluded from DAPI-stained nuclei in blue). **b** is at  $\times 50$  magnification relative to **a**.



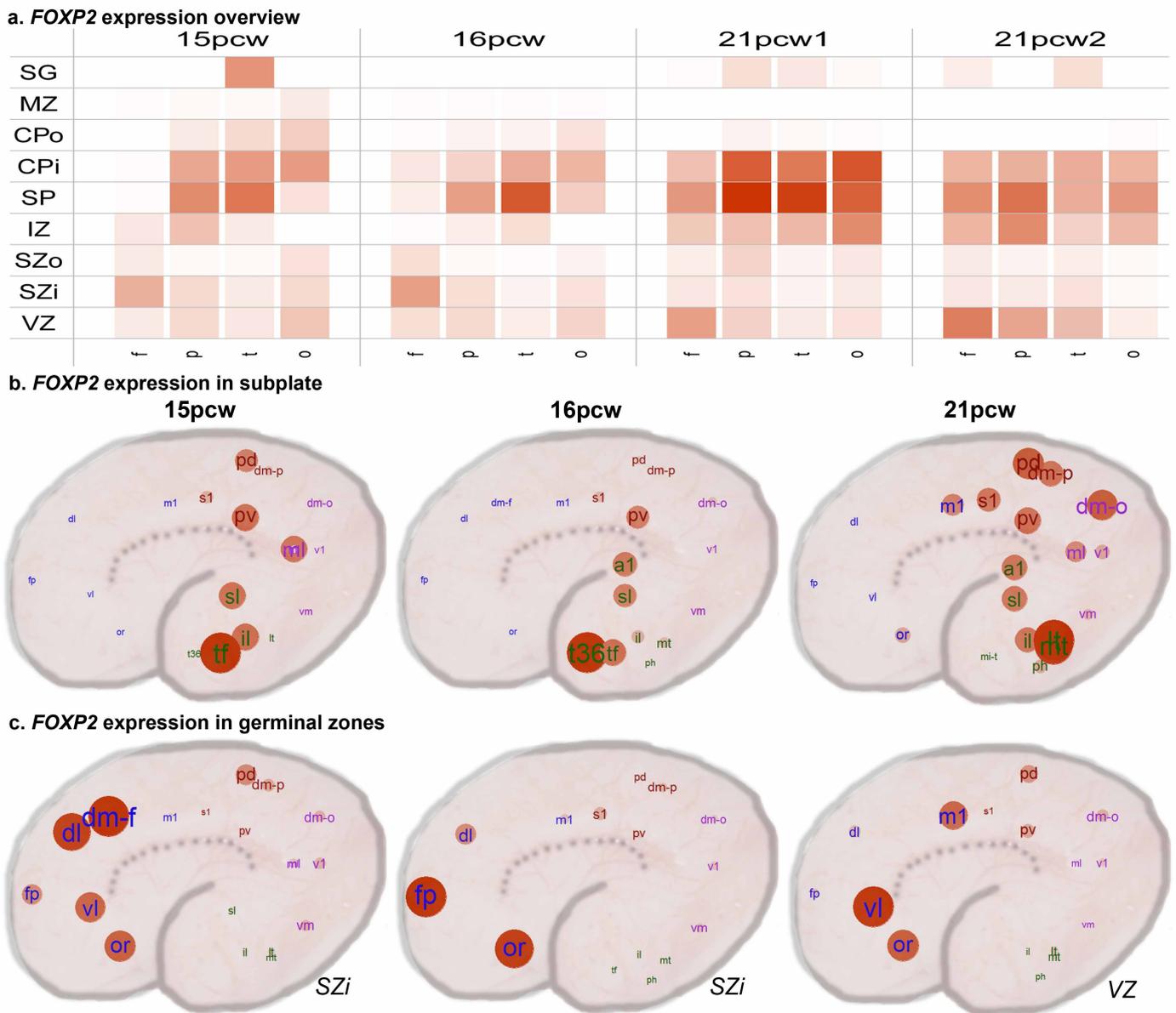
**Extended Data Figure 8 | Laminar gene expression of putative SP markers for human and mouse in prenatal human cortex.** **a**, Novel human subplate-enriched genes showing at least eightfold enrichment in subplate in all four prenatal human brains. *CDH18*, a known subplate marker in mouse, is presented as a positive control. **b**, Genes with differences in subplate expression

between mouse and human. These genes have been reported as subplate-enriched in mouse studies but do not show human subplate enrichment. Labelling as in Fig. 4a of the main manuscript. Microarray data are plotted as the average  $\pm$  s.e.m. for each layer in each of the four brains analysed (colours).



**Extended Data Figure 9 | Areal gradients are consistent with patterns in BrainSpan RNA-seq cortical data, particularly for post-mitotic layers.** RNA-seq RPKM values for 8–22 pcw specimens in the BrainSpan database were used to assess rostral-caudal patterning for all genes in prenatal development. Specifically, gene expression was correlated with a template of frontal cortex samples (1) versus samples from other cortical areas (0), such that positive correlations correspond to rostral enrichment. The same density plot

of the resulting correlations is plotted for each layer in black. For each layer (except the subpial granular zone), density plots for the subset of rostral (red) and caudal (green) genes identified in this study (Fig. 5h) are shown. Note the significant offset of density curves for rostral and caudal genes in MZ, CPo, CPi, and IZ (and other layers to a lesser extent), indicating good agreement in areal gradient genes between studies.



**Extended Data Figure 10 | Areal and laminar expression patterning of FOXP2.** a, Summarized expression levels of FOXP2 across each lobe, layer and brain. b, FOXP2 shows enrichment in parietal and temporal regions

overlapping Wernicke’s area in subplate at all three time points. c, FOXP2 shows enrichment in frontal cortex in germinal zones. Red corresponds to higher expression.

## Supplementary Methods

### *Experimental methods*

Additional information regarding post-mortem tissue acquisition and screening, laser microdissection and RNA isolation, mRNA profiling, microarray preprocessing, and *in situ* hybridization are available from the Allen Brain Atlas data portal ([www.brain-map.org](http://www.brain-map.org)) in white papers, which can be accessed from the "Documentation" tab of the relevant atlas. For the prenatal human laser microdissection (LMD) microarray atlas, this can be found under the "Developing Human Brain" link, or from the BrainSpan website ([www.brainspan.org](http://www.brainspan.org)).

Additional information on single and double fluorescent *in situ* hybridization (FISH) protocols have been previously published<sup>24,56,66</sup>. We have now published a comprehensive analysis showing that the Agilent microarray used in our study provides accurate measurements of relative gene expression values<sup>65</sup>.

### *Differential gene expression analyses and sample clustering and visualization*

Assessments of differential expression between two sample groups were done using correlation to a template vector, by measuring the fold change, or by calculating a T score. Specifically, genes were considered layer-enriched at 21pcw if the Pearson correlation (R) between gene expression and the template vector that layer (where 1="in layer" and 0="not in layer") was greater than 0.5 in both brains from 21pcw (and likewise for the earlier two brains, 15-16pcw; **Fig. 5c**; **Suppl. Table 3**). Fold change between SZo and SZi (**Fig. 3f**) or between SP and the remaining layers (**Suppl. Table 8**) was defined as mean expression in the first layer across all relevant brains divided by mean expression in the other layer or layers, and was typically

reported in log<sub>2</sub> space. Finally T-scores were calculated using a T test comparing samples from 15/16 pcw vs. 21pcw as an estimate for changing expression patterns with age, or comparing samples from CP vs. SZ at 15/16 or 21pcw as an estimate for changing expression patterns between germinal and postmitotic layers (**Fig. 3a**). Identification of rostral-caudal areal gradient genes (**Fig. 5**) was explained in detail in the on-line Methods. Unless otherwise noted, all p-values are Bonferroni corrected for multiple comparisons. Gene lists were characterized as described in the Methods and for modules below.

Laminar structure of the prenatal neocortex is visualized using hierarchical clustering (based on the most differential genes from Bayes ANOVA,  $p < 10^{-29}$ ; **Fig. 2b**), as well as using multidimensional scaling (MDS) across all genes (**Fig. 2c**). In both plots germinal zones (VZ, SZi, SZo) group distinctly from layers containing primarily postmitotic cells. For other visualizations, red represents higher expression and white or green represents lower expression.

### ***Weighted gene co-expression network analysis (WGCNA)***

Two weighted gene co-expression network analyses (WGCNA) were performed. First, WGCNA was run on all neocortical samples using the standard method (**Fig. 3a-c**)<sup>36,58</sup>, including the highest expressed probe for the 16892 most highly expressed genes (based on 99<sup>th</sup> quantile of expression  $> 100$ ), as a secondary filter to remove low expressers. Second, WGCNA was run on germinal layers by defining a consensus network in the 15 and 16pcw brains<sup>59</sup>, only including probes differentially expressed across these layers (**Fig. 3d-e**; 5494 probes = 4401 unique genes; ANOVA  $p < 0.01$ , Benjamini-Hochberg adjusted). Note that this network considered all passing probes, and not just the mostly highly expressed probes for each gene.

For each network we calculated pair-wise Pearson correlations between each gene pair, and then transformed this matrix into a signed adjacency matrix using a power function. The components of this matrix (connection strengths) were then used to calculate "topological overlap" (TO), a robust and biologically meaningful measurement of gene similarity, which quantitatively assesses the similarity of two genes' co-expression networks<sup>36,67</sup>. Genes were hierarchically clustered using the 1-TO as the distance measure, and initial module assignments were determined by using a dynamic tree-cutting algorithm<sup>68</sup>. Differential expression between genes and either age or cortical layer were calculated using a one-sided T-test, and displayed as color bars in **Fig. 3a** to highlight module/phenotype relationships. The consensus network was run by first calculating TO values independently for the 15pcw and 16pcw brain samples, appropriately scaling the TO matrices, and then taking the minimum TO for each pair of genes as the representative value<sup>59</sup>. The effect of this is to emphasize connections that are present in both brains, and omit connections that are only present in a single brain, either due to changes with age or differences between individuals.

Modules in each network were characterized using several strategies. First, modules were annotated based on gene ontology enrichment using Enrichment Analysis Systematic Explorer (EASE)<sup>69</sup>. Second, modules were further annotated by measuring their overlap with previous brain-related lists using the function `userListEnrichment` in the WGCNA library setting `useBrainLists=TRUE`<sup>63</sup>. These lists include modules from previous WGCNA studies of human and mouse brain, and experimentally-derived marker genes for cell-types, and region-enriched and disease-specific genes from previous microarray studies, and others (see the help file of this function for more details and references). Third, modules were characterized based on the

expression of the module eigengene (ME), or the first principle component of the module.

Specifically, plots showing the expression of MEs across the data in both networks were created in order to better visualize expression patterns. Finally, the Pearson correlation between each gene and each ME—referred to as a gene's module membership (kME)<sup>37,70</sup>—was calculated (**Suppl. Table 5**). Module C31 is also depicted using VisANT<sup>64</sup>: the top 250 gene-gene connections based on topological overlap are shown as orange lines, with histone genes removed for clarity. In these plots, each point is a gene (or node) and genes with at least 15 connections (i.e., hub genes) are displayed with larger points and in larger font. The positioning of the nodes is done by first using an automatic layout, then manually adjusting the positions so that each gene name is visible.

### *Permutation analysis to identify expected number of gradient genes*

To identify rostral vs. caudal genes, we performed the following analysis independently for each layer. For each subregion, we defined the rostral-caudal angle as the angular distance from the point where the temporal and frontal cortices meet (since the frontal and temporal lobes are physically adjacent but separated by the lateral sulcus), rather than as a linear distance. For each gene we correlated the expression level of that gene in each subregion with the corresponding rostral-caudal angle independently for each brain. Genes with  $R > 0.5$  in all four brains were considered rostrally-enriched and those with  $R < -0.5$  were considered caudally-enriched. The permutation test was done by repeating the exact same procedure, but with the rostral-caudal angles of the subregions randomized in Euclidean space 100 different ways. This strategy does not make the assumption that genes are independent, does not depend on the specific genes used in the analysis, and ensures that subregions are shuffled identically for each layer in each brain.

Using this strategy we find an average of 0.89 rostral genes (never more than 8), and 0.31 caudal genes (never more than 6) in our permutations, which is much fewer than the 463 rostral and 252 caudal genes identified in our data set. Using these results, a conservative estimate of false discovery rate would be  $6 / 252$  or  $\sim 2.4\%$ . The observed numbers of gradient genes would almost definitely not occur by chance ( $p \sim 0$ ). As an alternative strategy, we also directly identified the number of expected areal genes across brains as a function of the number of areal genes in each brain independently (specifically,  $[\# \text{ in Brain1}] * [\# \text{ in Brain2}] * [\# \text{ in Brain3}] * [\# \text{ in Brain4}] / [\text{total } \# \text{ genes}]^3$ ), and found comparable results—for most layers we find  $<1$  rostral and  $<1$  caudal gene by chance.

### *Unbiased gradient gene analysis*

A similar strategy was used to identify unbiased areal gradient genes as was used to identify gradient genes in **Fig. 5h** (see **Methods**). Specifically, the gradient center for each gene was estimated by finding the center of mass of expression, and then the Euclidean distance between this point and each region was correlated with gene expression. A density plot of the gradient centers for each gene with  $R > 0.65$  is shown in **Fig. 5a-b**.

### *Comparison between human and mouse gradient genes*

To identify genes showing consistent rostral or caudal enrichment in human and mouse, we compared our lists of gradient genes identified in human cortex to gradient genes identified in control C15Bl/6J mice at E14<sup>51</sup>. To calculate rostrocaudal gradients in mouse, we first downloaded the data set generated in this study (GSE12134) from Gene Expression Omnibus (GEO)<sup>71</sup>, which was run using the Affymetrix Mouse Genome 430 (version 2.0) Array. These

data included arrays run on rostral and on caudal cortex from three control mice (as well as three mutant mice, which were omitted from our analysis). Mouse genes were mapped to orthologous human genes, and only the probe with the highest average expression was considered in this analysis, as previously described<sup>63</sup>. For the subset of gradient genes identified in human cortex that were probed on this array, a paired T test was used to quantify the significance of rostral or caudal enrichment for each gene across the three mouse brains. The 20 genes found to show gradient expression in mouse (with paired T statistics of  $t > 4$ ) in the same direction as in human are reported in **Suppl. Table 9**.

#### *Statistical assessments addressing potential issues of low specimen number*

Due to the unique nature of our data set—large number of samples, but a very small number of specimens—we performed additional statistical tests to assess the validity of some of our results. For our laminar genes (**Fig. 2**) and for our rostral-caudal genes (**Fig. 5**), we assessed the likelihood that a specific gene list is valid within a given specimen, and also assessed the extent to which the gene lists identified in multiple specimens are likely to be true positives. For the case of laminar genes, our analysis included the two 21pcw specimens. To address the false positive rate within each specimen, we randomly permuted the layer labels for each sample 100 times, and repeated our analysis to determine the number of laminar genes expected by chance. In our second 21pcw specimen, we rarely found any laminar genes by chance and never found more than 8, compared with 3874 in our actual analyses, suggesting a false positive rate of approximately  $8 / 3874 = 0.002$  or better. The other specimen had a false positive rate closer to 0.0003, using the same strategy. In short, nearly all of our laminar genes within a given brain are true positives. To address the issue of false positives due to small specimen number, we asked

the following question: given the number of genes identified in each layer of each of our two 21pcw specimens independently, how many would we expect to find in both brains by chance? Depending on the layer, the expected number of laminar genes was between 2.2-8.3% of our observed number of laminar genes. Thus, ~95% of the identified laminar markers would likely represent generally reliable laminar markers at 21pcw.

To test for false positives in our rostral and caudal marker gene lists, we performed a permutation test as described above, and found a false positive rate of 2.4% across specimens. When we looked at the number of observed rostral caudal genes within a specimen compared with the number expected by chance, however, we note that some of these rostral caudal gene lists are very noisy, particularly for the older specimens. Specifically, we find that with a single brain we would expect false positive rates of between 10-85%, indicating that for this analysis, it is critical to include as many specimens as possible, and that four specimens is enough to identify a small number of high-confidence areal genes (which we validate using RNA-Seq in **Extended Data Figure 9**).

Finally, to assess the level of transcriptional similarity between our four specimens, we repeated our multidimensional scaling analysis (**Fig. 2c**), this time including samples from all four specimens simultaneously. We then measured the average distance between matching neocortical areas on this plot for each pair of specimens. The resulting values are quantitative distances in arbitrary units (in principal component space), which represent summary values of the transcriptional distance between specimens. We find a linear relationship ( $R=0.88$ ,  $p=0.02$ ) between difference in developmental weeks of a specimen and mean transcriptional distance of a

specimen. Specifically, the two 21pcw specimens have a distance of 0.033, which is much less than the distances between the 21pcw specimens and the 16pcw (0.114 & 0.116) or the 15pcw (0.194 & 0.185) specimen. The distance between the two younger specimens is 0.096. Together these analyses show that, while small specimen number is a valid concern, we are still able to find statistically validated, biologically reasonable results using many samples from four brains.

### Additional References

- 65 Miller, J. A. *et al.* Improving reliability and absolute quantification of human brain microarray data by filtering and scaling probes using RNA-Seq. *BMC Genomics* **15**, (2014).
- 66 Thompson, C. L. *et al.* Genomic anatomy of the hippocampus. *Neuron* **60**, 1010-1021, (2008).
- 67 Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551-1555, (2002).
- 68 Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719-720, (2008).
- 69 Hosack, D., Dennis, G., Sherman, B., Lane, C. & Lempicki, R. Identifying biological themes within lists of genes with EASE. *Genome biology* **4**, R70, (2003).
- 70 Horvath, S. & Dong, J. Geometric Interpretation of Gene Coexpression Network Analysis. *PLoS Comput Biol* **4**, e1000117, (2008).
- 71 Edgar, R., Domrachev, M. & Lash, A. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* **30**, 207-210, (2002).

**Supplementary Tables****Supplementary Table 1: Donor profiles of four prenatal human specimens.**

<b>Specimen ID</b>	<b>Alias</b>	<b>Age (pcw)</b>	<b>Sex</b>	<b>Race</b>	<b>Average RIN</b>	<b>Hemisphere</b>
H376.IV.02	21 pcw1	21 4/7	F	AA	8.5	Left
H376.IV.03	21 pcw2	21	F	Asian	9	Left
H376.IIIA.02	15 pcw	15	M	Caucasian	8.2	Left
H376.IIIB.02	16 pcw	16	F	Asian	8.2	Left

Column A shows the specimen ID. Column B lists the alias specimen ID. Column C, D, and E lists the age, sex, and race, respectively. The average RIN and hemisphere used in the analysis are shown in columns F and G, respectively.

*The remaining supplementary tables (2-9) are included as separate files.*