Cortical morphology at birth reflects spatio-temporal patterns of gene expression in the fetal brain

Ball, G.,[1,2] Seidlitz, J.,[3,4] O'Muircheartaigh, J.,[2] Dimitrova, R.,[2] Fenchel, D.,[2] Makropoulos, A.,[2] Christiaens, D.,[2] Schuh, A.,[5] Passerat-Palmbach, J.,[5] Cordero-Grande, L.,[2] Hughes, E.,[2] Price, A.,[2] Hajnal, J.V.,[2] Rueckert, D.,[5] Robinson, E.C.[2] & Edwards, A.D.[2]

1. Developmental Imaging, Murdoch Children's Research Institute, Melbourne, Australia
2. Centre for the Developing Brain, Department of Perinatal Imaging & Health, King's College London, UK
3. Developmental Neurogenomics Unit, National Institute of Mental Health, Bethesda, USA
4. Department of Psychiatry, University of Cambridge, UK
5. Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK

Corresponding author
Dr Gareth Ball
Developmental Imaging
Murdoch Children's Research Institute
The Royal Children's Hospital
Parkville 3052
Australia

gareth.ball@mcri.edu.au

Abstract

In this study, we test the hypothesis that noninvasive markers of brain development reflect the spatiotemporal patterning of genes underlying corticogenesis during gestation and the developmental staging of the neonatal brain. Additionally, we test the selective vulnerability of molecular processes underlying cortical development to disruption following preterm birth. We find that gene expression in the fetal cortex is mirrored by a principal mode of variation in the neonatal cortex. Specifically, regional variation in cortical morphology and microstructure reflect differences in developmental maturity across cortical areas, indexed by the differential timing of gene expression across multiple cell types in the fetal cortex. Further, the effects of preterm birth are temporally and spatially coincident to developmental processes involving the differentiation and specialisation of cortical oligodendrocyte populations. This work provides an experimental framework to link molecular developmental mechanisms to macroscopic measures of cortical anatomy in early life, demonstrating the relationship between fetal gene expression and neonatal brain development and highlighting the specific impact of early exposure to the extrauterine environment due to preterm birth.

Keywords

brain development; transcriptomics; cortex; premature birth; magnetic resonance imaging

Introduction

The human cortex is composed of functionally distinct regions organised within broadly hierarchical systems.[1–4] While the mechanisms behind the emergence of this complex topography are not yet fully understood, cortical patterning is underwritten by dynamic regulation of gene transcription during gestation.[5–8] The advent of modern transcriptomic technologies has allowed the precise mapping of cortical gene expression in the prenatal period.[9,10] Differential gene transcription across cortical regions is most pronounced during prenatal development and varies along a well defined gradient from lower to higher order areas.[6,8,9,11] Interruption to the precisely timed dynamics of gene transcription during gestation is implicated in the onset of common developmental cognitive and neuropsychiatric disorders.[12–15]

Recently, the *post mortem* transcription of thousands of genes across the adult brain has been compiled to form brain-wide, gene expression atlases.[16] This allows precise comparison between spatial patterns of cortical gene expression and neuroanatomy quantified using Magnetic Resonance Imaging (MRI).[17,18] Neuroimaging studies have found patterns of gene expression in the adult cortex are mirrored by regional variation in cortical morphometry[19,20] and functional organisation,[21] and are associated with neuroimaging markers of developmental disorders.[22,23] Similar databases detailing cerebral gene transcription across the full human lifespan from early embryonic stages to adulthood are now available.[9,12] This has created an unprecedented opportunity to explore the molecular correlates of neuroimaging markers of early brain development.

Advances in neonatal neuroimaging now permit the quantification of developmental neuroanatomy *in vivo* at a higher-resolution than previously possible.[24–28] Imaging studies of the developing human brain shortly after birth have characterised a highly dynamic period of cerebral change defined by significant increases in brain volume,[29,30] cortical thickness and surface area,[31–35] progressive white matter myelination[36–40] and ongoing configuration and consolidation of functional brain networks.[41–43] Further, the truncation of gestation due to preterm birth is associated with widespread alterations in brain development indexed by MRI at the time of normal birth,[44–49] highlighting the sensitivity of noninvasive neuroimaging to detect disruptions in early developmental processes.

The combination of these technologies opens a new window on early brain development, facilitating a comparison between patterns of prenatal cortical gene expression and the development of the brain at around the time of birth, as well as providing a platform to test mechanistic hypotheses about the impact of early disruptions to brain development during gestation. In this study, we explore the association between *in vivo* measures of cortical morphometry at birth and regional patterns of fetal gene transcription to test the hypothesis that noninvasive markers of brain development reflect the spatiotemporal patterning of genes underlying corticogenesis during gestation and the developmental staging of the neonatal brain. Additionally, we test the selective vulnerability of molecular processes underlying cortical development to disruption following preterm birth.

We find that gene expression in the fetal cortex is mirrored by a principal mode of variation in cortical morphometry at birth. Regional variation in cortical development is predicted by the differential timing of gene expression across the cortex and truncation of the intrauterine period by preterm birth is encoded by effects that suggest a selective vulnerability of developmental cell populations to early extrauterine exposure.

Results

A principal organisational gradient in the neonatal cortex

Using high-resolution structural and diffusion MRI data acquired from a large cohort of healthy neonates (n=292, 54% male, median [range] gestational age at scan = 40.86 [37.29-44.71]), we extracted six measures of cortical morphology (cortical thickness) and microstructure (T1w/T2w contrast; Fractional Anisotropy, FA; Mean Diffusivity, MD; Intracellular Volume Fraction, fICVF; and Orientation Dispersion Index, ODI) from eleven cortical regions-of-interest (ROI) with corresponding mRNA-sequencing in a prenatal transcriptomic dataset[12] (Fig 1A, Fig S1).

We used a hierarchical clustering algorithm to characterise inter-regional variation in neonatal cortical morphology and microstructure, grouping regions-of-interest together based on the similarity of their metric profiles (Fig 1B). Regional clustering was repeated with 10,000 bootstrapped samples, selecting *n* subjects with replacement from the full term-born cohort before calculating the group average matrix and clustering. Using the silhouette score[50], a measure of intra-cluster coherence, to select the optimal number of clusters, we identified four regional clusters (mean $\pm$ S.D. silhouette score = 0.67 $\pm$ 0.01, 10,000 bootstraps; Fig S2), the membership of which reflected, in part, a basic functional hierarchy within the cortex. We found the silhouette score of the four-cluster solution was significantly greater than expected by chance after comparison to a null distribution built from permuting assigned regional cluster labels ($p<0.001$, 10,000 permutations). Pairwise distances in the metric profile data were significantly preserved by the four-cluster solution (cophenetic correlation = 0.90), compared to a second null distribution drawn from simulated multivariate data with matched covariance (see Methods; $p<0.001$, 10,000 samples).

Clusters 1 and 2 comprised higher order cortical regions including frontal (dorsolateral prefrontal, DLPFC; ventrolateral prefrontal, VLPFC; orbitofrontal, OFC) cortex, inferior parietal cortex (IPC) and inferior temporal cortex (ITC) in cluster 1, and medial frontal cortex (MFC) and superior temporal cortex (STC) in cluster 2. In contrast, cluster 3 comprised primary sensorimotor regions in the auditory (A1C), motor (M1) and sensory (S1) cortex. Primary visual cortex (V1) formed its own cluster separate to other sensory regions.

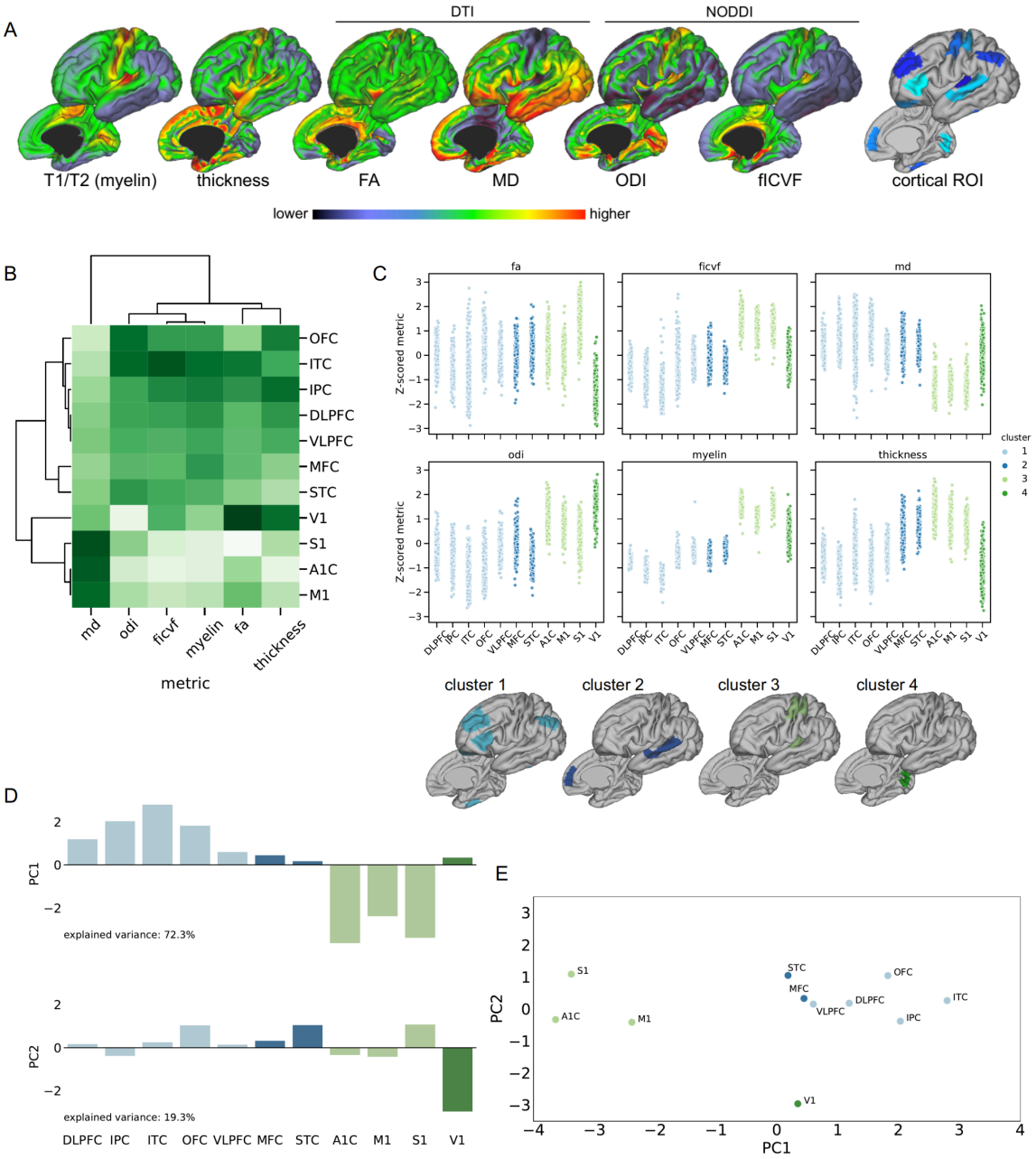

Figure 1: A principal organisational gradient in the neonatal cortex as indexed by multi-modal MRI. A. Average cortical imaging metrics in a cohort of healthy, term-born neonates (n=292). Metrics derived from structural MRI (T1/T2 contrast, cortical thickness) and diffusion MRI model parameters using DTI (FA and MD) and NODDI (fICVF and ODI). Cortical regions-of-interest based on anatomical references with corresponding developmental transcriptomic data (see Methods). B. Clustering of mean regional cortical metrics. Metrics are Z-transformed across regions to facilitate comparison; darker colours indicate relatively lower values. C. Z-scored cortical metrics are shown for each subject grouped within each cortical region-of-interest and coloured by cluster. Bottom row shows cortical regions coloured by cluster membership. D. PCA weights for each metric for the first two principal components, coloured by cluster. E. Position of each cortical region-of-interest based in PCA state space, the position of each region is dictated by its component score for the first two principal components. Regions are labelled and coloured by cluster.

Similar regional profiles are evident across metrics (Fig 1C) with the pattern of inter-regional variation reflecting the full transcortical patterns shown in Fig 1A. Comparable regional patterns were observed in FA and cortical thickness, and between ODI, fICVF and the T1w/T2w contrast (Fig 1B,C), with higher FA, thicker cortex and a higher T1w/T2w contrast in primary somatomotor cortex (Fig 1A, C). MD displayed an opposing trend across regions, lowest in primary somatomotor regions (cluster 3) and highest in fronto-parietal regions (cluster 1). Performing an omnibus test, we found significant differences across clusters in all metrics except for MD (repeated measures ANOVA: FA $F_{3,855}$=11.4, p<0.001; fICVF $F_{3,855}$=3.9, p=0.008; MD $F_{3,855}$=0.5, p=0.63; ODI $F_{3,855}$=7.2, p<0.001; T1w/T2w $F_{3,855}$=8.3, p<0.001; thickness $F_{3,855}$=53.9, p<0.001).

Based on the similarities in cortical patterning across metrics, we hypothesised that regional variation across metrics could be represented by a small number of latent factors. Using Principal Component Analysis (PCA), we sought a small set of factors, each a linear combination of the original imaging metrics, to maximally explain variance in the full set of cortical measures. Using the group-average region × metric matrix (Fig 1B), we found that the first two components explained 72.3% and 19.3% of the total variance, respectively (Fig 1D,E).

The first component (PC1) ordered cortical regions along a principal gradient with clusters 1 (DLPFC, VLPFC, OFC, IPC, ITC) and 3 (A1C, M1, S1) situated at opposite ends. This gradient is apparent in all cortical metrics, most strongly in T1/T2 contrast, fICVF and mean diffusivity (Fig S3). The second principal component (PC2) predominantly captured anatomical and microstructural differences in V1 compared to other primary cortex (Fig 1C). Clear separation between primary and higher-order cortical regions is apparent based on regional PC1 scores (Fig 1E) lending support to the idea that correlated regional variation across multiple MRI metrics at birth reflects a shared perspective of basic cortical organisation.

The principal imaging gradient is associated with regional patterns of gene expression in mid-gestation

Using a developmental transcriptomic dataset of bulk tissue mRNA data sampled from cortical tissue in 16 prenatal human specimens,[12] we compared regional variation in cortical MRI metrics with prenatal gene expression in anatomically-correspondent cortical regions. Through comparison to five independent single-cell RNA studies of the developing fetal cortex,[10,12,51–53] we selected a set of 5287 marker genes shown to be differentially expressed in cortical cells during gestation. We used a nonlinear mixed-effects approach to model developmental changes in gene expression as a smooth function of age, accounting for inter-specimen variability. This nonlinear model provided a better fit of the expression data for all genes compared to a comparable linear model (range AIC difference:-10.7 to -94.6; range BIC difference: -2.1 to -58.9).

Focusing first on the spatial variation in gene expression across the cortex, we calculated specimen- and age-corrected RPKM values for each gene using the residuals of the nonlinear mixed model (Fig

S4). We then tested the association between spatial variation in gene expression during gestation and the principal imaging gradient at birth using non-parametric correlation (Kendall's τ) between age-corrected RPKM values and PC1 score across cortical regions. Of 5287 genes, 120 displayed a significant (positive or negative) correlation with the principal imaging gradient (PC1) after correction for multiple comparisons with False Discovery Rate ($p<0.05$). In total, 71 genes were positively correlated with PC1, with increasing expression in regions with a positive PC1 score (mean $\pm$ S.D. τ = 0.208 $\pm$ 0.023) and 49 genes displayed the opposite relationship, with higher expression in regions with a negative PC1 score (mean $\pm$ S.D. τ = -0.208 $\pm$ 0.022).

We reasoned that genes associated with the patterning of cortical morphometry at birth may underpin important neurodevelopmental functions. To test this, we performed an over-representation analysis (ORA) for ontological terms associated with specific biological processes in both gene lists. Of 71 genes with spatial patterns of expression that were positively associated with PC1 (denoted PC+), 61 (86%) were annotated to specific functional terms. Using all protein-coding genes transcribed in the bulk RNA dataset as the background reference set, we found significant enrichment of several neurodevelopmental terms including: stem cell differentiation (FDR=0.001, enrichment ratio=9.32), neuron migration (FDR=0.03, enrichment=7.94) and forebrain development (FDR=0.004, enrichment=5.65) (Fig 2A; Table S1). Terms relating to stem cell and neuronal differentiation remained significantly enriched when restricting the background reference set to only include genetic markers of fetal cortical cells (n=5287; Table S1). No biological terms were significantly enriched in genes with a spatial pattern of expression negatively correlated to PC1 (denoted PC-).

Performing weighted gene correlation network analysis (WGCNA)[54] on the PC+ gene set, we identified two co-expressed gene modules (Fig 2B). The largest, Module 1, contained 53 genes including a tightly correlated set of developmental genes with roles in regulating cell growth and differentiation including *EOMES*, *NEUROD4*, *SFRP1* and *TFAP2C* and was significantly enriched for GO terms: nervous system development (GO ID: 7399) and system development (GO ID: 48731, both FDR-corrected p=0.047). On average, expression of genes within Module 1 was highest in younger tissue samples, peaking at around 120 post-conceptional days and decreasing thereafter (Fig 2D). The smaller, second module (Module 2) contained 13 genes, with roles including neuronal signalling (*ERBB4*, *CALB2*, *SCGN*) and neuronal differentiation (*ZNF536*, *DLX1*). Top enriched GO terms included synapse maturation (GO ID: 60074) and synapse organisation (GO ID: 50808, both p=0.059). Average modular gene expression accelerated rapidly between 75 and 125 post-conceptional days, increasing slowly across the remaining age range (Fig 2D). In the PC- gene set, 3 small modules of 7 genes each were identified (Modules 1N-3N; Fig S5), including genes with high neuronal expression (Module 1N; *CDKL5*, *ZBTB18*, *SORCS1*), and genes involved in cellular processes including adhesion and signalling (Module 2N: *ACTN2*, *PTPN2*, *SSX2IP*) and metabolic activity (Module 3N: *DUSP7*, *ST3GAL1*) although no biological terms were significantly enriched in any module. All three modules exhibited a later peak in average expression between 175 and 200 postconceptional days (Fig S5).

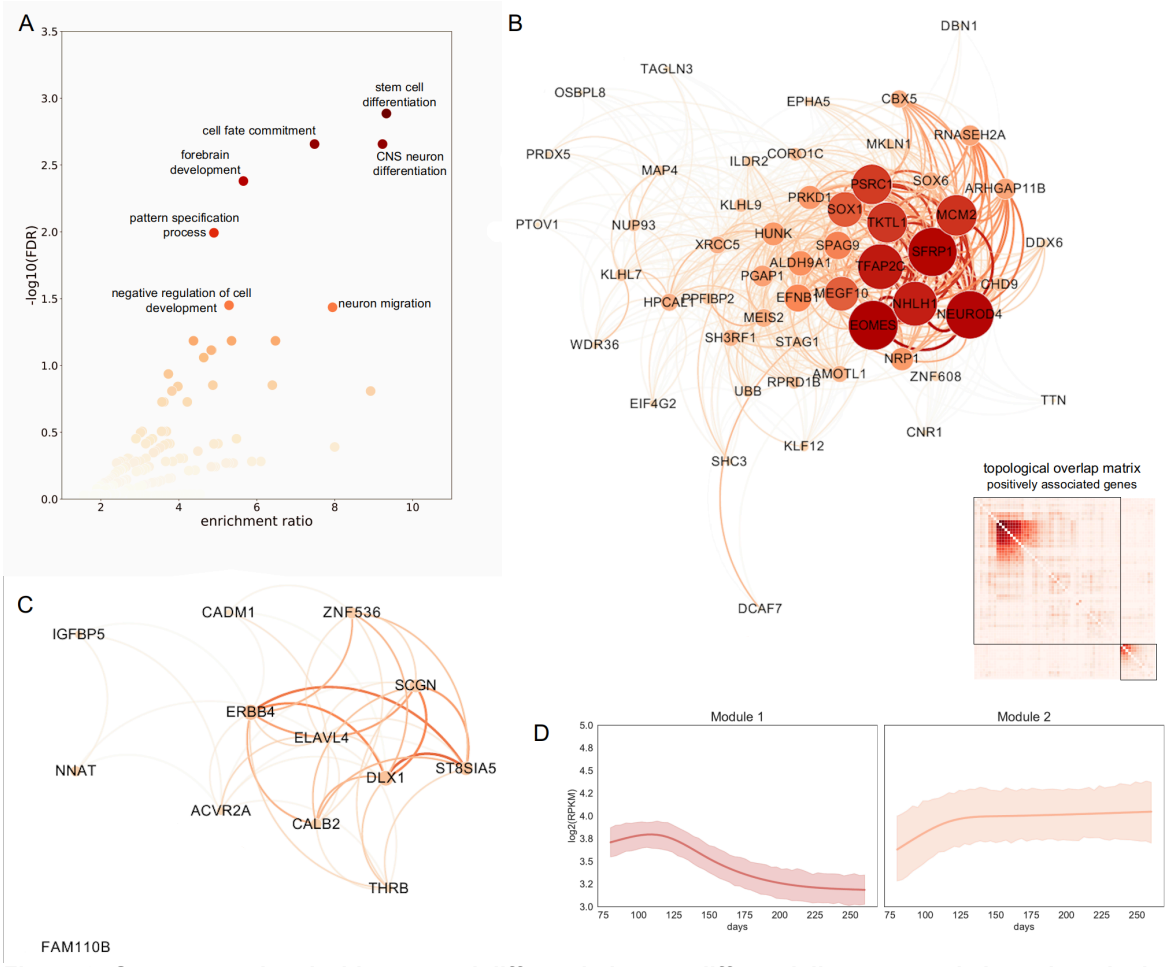

Figure 2: Genes associated with neuronal differentiation are differentially expressed along the principal imaging gradient. A. Volcano plot showing enrichment of GO terms (Biological Processes) in genes with age-corrected expression that is positively associated with PC1. Significantly enriched terms (FDR < 0.05) are labelled. Gene co-expression analysis of all PC+ genes revealed two modules (B: Module 1; C: Module 2; topological overlap matrix showing all inter and intra-modular connections; B, inset). Intra-modular connections are shown with node size and colour indicating strength, and edge thickness and colour indicating weight. D. Module specific developmental trajectories. Average trajectories estimated using nonlinear mixed effects modeling for all genes in Module 1 (right) and Module 2 (left), shaded area indicates 95% C.I.

Imaging-gene associations are enriched for specific cell types in the fetal cortex

To explore these relationships further, we reconstructed cellular gene expression profiles by stratifying the bulk tissue expression data using genetic markers of cell type derived from single-cell RNA studies of the fetal cortex.

Sets of genetic markers for eleven cortical cell classes were initially compiled by combining lists of genes that are differentially expressed in fetal cortical cell populations (Table S2). To verify this grouping, we calculated the average expression trajectories for all genetic markers within each cell type across gestation and used them to calculate a 2D embedding using Uniform Manifold

Approximation and Projection (UMAP; Fig 3).[55] Proximity in the embedded space reflects similarity between average trajectories of gene expression within cell type over time. In the embedded space, cell types clustered by assigned class, and maturational timing (precursor or mature), as well as within cellular subtype (eg: inhibitory and excitatory neurons). Average trajectories for 10 cell classes (excluding a non-specific neuronal class) are shown in Fig S6.

We then tested the enrichment of each cell class within the PC+ and PC- gene sets. We found that PC+ genes were significantly enriched for genes expressed by precursor cells (p=0.0003), specifically, for genes expressed by intermediate progenitor cells (enrichment ratio = 1.63, p=0.0002; Table S3) and inhibitory neurons (enrichment ratio = 3.2, p<0.0001; Fig 3C; Table S3). Posthoc analysis within cell class, revealed specific inhibitory neuron subtypes present in the mid-fetal brain and enriched in the PC+ gene set included migrating cortical interneurons from the caudal ganglionic eminence (In_5,[53] IN-CTX-CGE2,[52] both p<0.0001) and newborn interneurons originating in the medial ganglionic eminence (nIN1;[52] p=0.0017). Finally, we tested whether genes in significantly enriched cell classes were also enriched within specific co-expression modules (Fig 2B). We found that Module 1 was enriched for progenitor genes (enrichment=1.95, p<0.0001) with 34 out of 53 modular genes also expressed by progenitor cells, in addition to 13 genes expressed by inhibitory neurons (enrichment=2.28, p=0.003). Module 2 contained 11 genes (out of 13) expressed by inhibitory neurons (enrichment=7.9, p<0.0001), and 4 expressed by progenitor cells (p>0.05). When only considering genes unique to each cell class, Module 1 was enriched significantly for progenitor genes (shared genes: *MCM2*, *ILDR2*, *EOMES*, *DBN1*, *NEUROD4*, *KLHL9*, *NHLH1*), and Module 2 for inhibitory neuron genes (*CALB2*, *THRB*, *SCGN*, *ERBB4*).

In contrast, PC- genes, with a spatial pattern of expression that was higher in primary somatomotor regions, were enriched for genes expressed by mature cell types (enrichment=1.18, p=0.002). In terms of cell class, genes expressed by oligodendrocytes were enriched within PC-, though not significantly (enrichment=1.75, p=0.056; Fig 3C; Table S3). When considering only marker genes uniquely expressed by each cell class, PC- genes were enriched for unique excitatory neuronal genes (enrichment=2.14, p=0.008; Table S4). Posthoc analysis within this class revealed a single enriched early maturing excitatory neuronal subtype (Ex_4,[53], p<0.0001). In terms of co-expression modules, 6 out of 7 genes in Module 1N were expressed by fetal excitatory neurons (enrichment = 2.07, p=0.022).

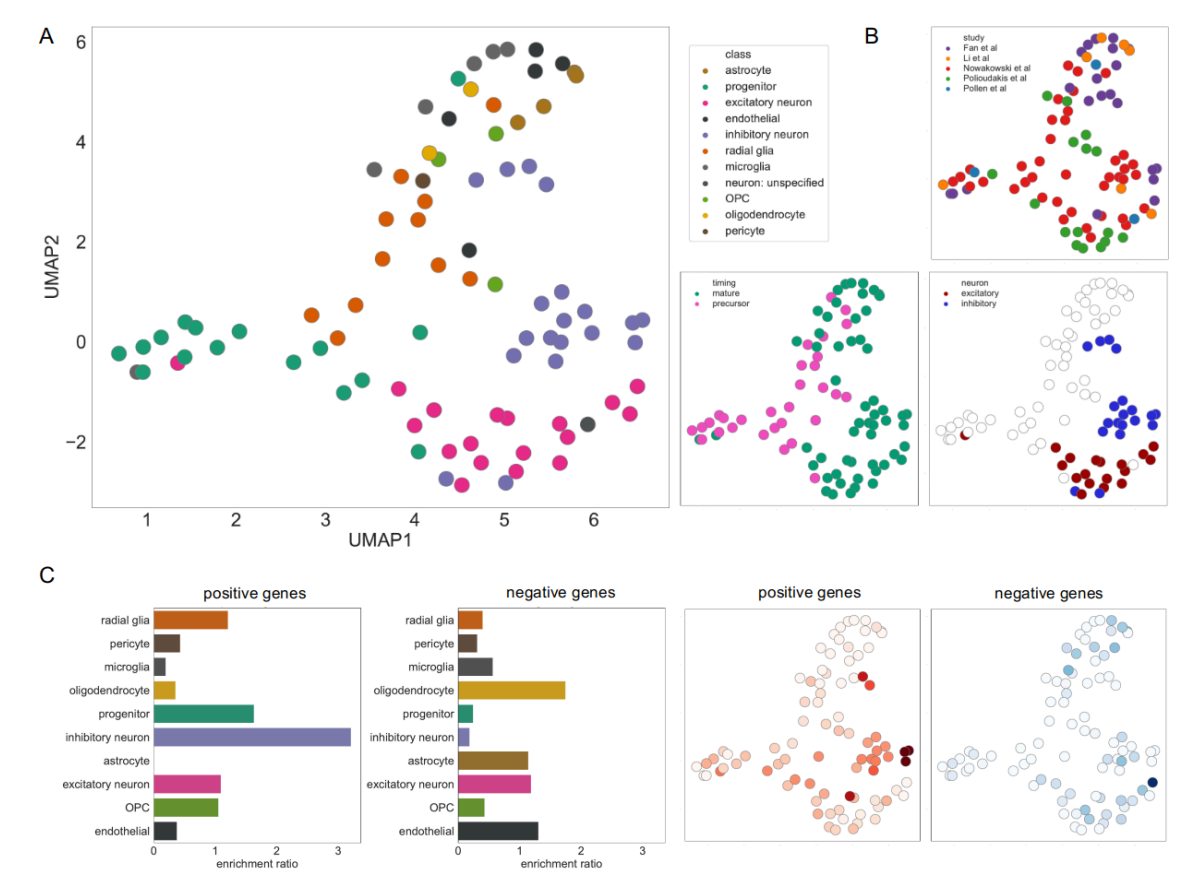

Figure 3: Cell-specific gene expression is associated with cortical morphology at birth. A. UMAP embedding of 86 cell types based on trajectories of relative gene expression over time recovers cell classes, cell timing (B. bottom row, left) and neuronal subtype (B. bottom row, right) but not study (B. top row). C. Enrichment ratio for marker genes expressed by each cell class is shown for PC+ (first) and PC- (second) associated gene sets, with cell-type specific enrichment shown for all 86 cell types (darker colour represent higher enrichment ratio).

Variation in genetic maturation during gestation predicts cortical development at birth

These data suggest that the spatial patterning of gene expression in the developing cortex is mirrored by regional variation in cortical morphology and microstructure at birth. The observed differences in expression across cortical areas additionally suggest a gradient of neurobiological maturation (Fig 4A), potentially represented by the differential timing of gene expression across regions.

To test this hypothesis, we created a model of cortical maturity to capture the relationship between the regional timing of gene expression and tissue maturation. Surmising that tissue maturity would be indexed by a unique pattern of age-dependent gene expression dictated by the developmental maturity of different cell populations, we first trained a regularised kernel regression model to predict the age of each prenatal specimen (n=16) using expression over time of the 120 regionally-variant genes (PC+ and PC-). Using a leave-one-out (LOO) framework, we modelled the association between mean cortical gene expression (averaged over regions) and specimen age using data from 15 out of 16 brains. We then used this model to predict age based on the *regional* gene expression profiles

from each of the left-out specimen's 11 cortical samples. We repeated this process for all 16 specimens. Using this method, we created a regional 'genetic maturation index' (GMI), subtracting specimen age from LOO-predicted age for each cortical sample within a specimen, such that a positive value indicates that genetic maturity within a region is relatively advanced (and therefore predicted to be older) compared to the rest of the cortex and vice versa

We found that genetic maturity varied both across cortical regions and across gestation. We observed an increasingly negative association develop over gestation between a cortical region's genetic maturity based on n=120 significant genes and its position along the principal imaging gradient at birth (r=-0.90, p=0.001, 1000 permutations, Fig 4B,C) such that, in the oldest sample, regional genetic maturity was correlated negatively with PC1 score (r [95% C.I.]=-0.45 [-0.14, -0.74], Table S5) and most advanced in primary somatomotor regions M1 (GMI [95% C.I.]=5.07 [-3.8,13,9]) and S1 (GMI=10.53 [1.7,19,9]). This relationship was also present, though weaker, in a larger model, using expression data from all genes (r=-0.61, p=0.049; Figure S7). We confirmed the specificity of this relationship by confirming that the correlation between PC1 and genetic maturity over time was significantly larger than equivalent models using random selections of n=120 genes from the larger gene set (mean |r|=0.51, p=0.001, 1000 permutations).

Using nonlinear models of gene expression over time, we evaluated regional genetic maturity at several points across gestation (Fig 4D). We found that the relative maturity of regions compared to the rest of the cortex varied over time. Primary somatomotor regions (cluster 3) remained relatively advanced throughout gestation, with an average positive genetic maturity index at all timepoints.In contrast, V1 (cluster 4; Fig 4D) remained relatively delayed across gestation. A divergence in maturity becomes apparent within higher-order regions (cluster 1) by mid-gestation, with some cortical areas (IPC, ITC) falling behind other regions towards the time of birth. These patterns were largely repeated in the larger gene set (Fig S7).

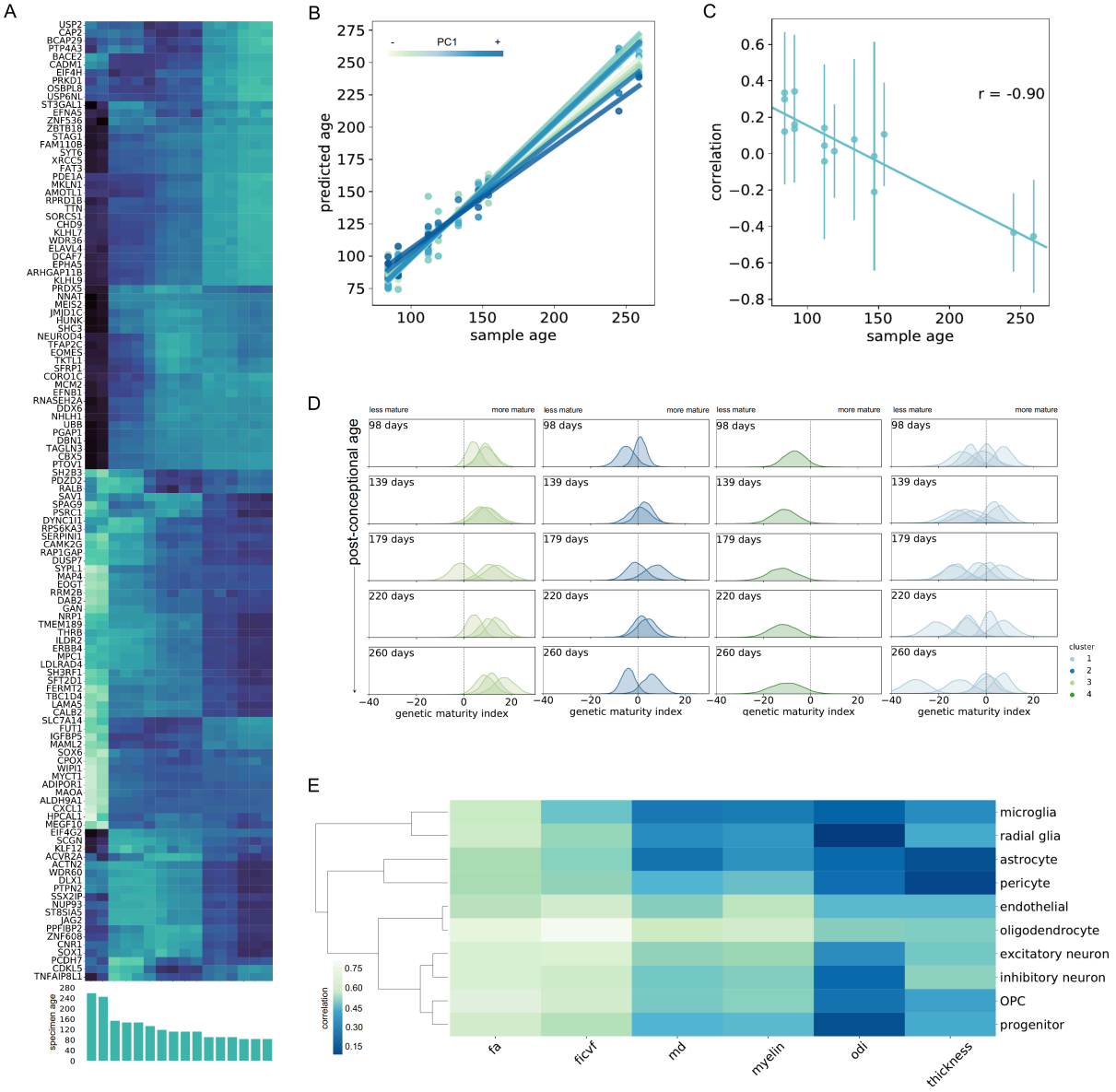

Figure 4: Genetic maturity correlates with regional variation in cortical morphometry at birth. A. Developmental patterns of mean cortical gene expression illustrated in each specimen for all 120 regionally variant genes (PC+ and PC-), ordered by age. B. The relationship between predicted and true sample age for all cortical regions, calculated using leave-one-out cross-validation. Regression lines are plotted separately for each region, and coloured by position along the principal imaging gradient. C. Correlation between regional variation in cortical morphometry (given by position along PC1) and genetic maturity in each specimen, plotted against specimen age. Error bars indicate 95% C.I. D. Developmental trajectories of genetic maturation for regions within each imaging cluster ordered by position along the principal gradient (left-to-right: A1C, S1, M1; MFC, STC; V1; DLPFC, VLPFC, OFC, IPC, ITC). Density plots show the difference between model-predicted age and sample age (5,000 bootstrapped gene samples). Positions to the left of 0 indicate regions that are less mature compared to the mean. Five time windows through gestation are shown, with age in postconceptional days. E. Absolute correlations between differential tissue maturity based on the modelled expression of genes from 10 cell classes at 260 days post-conception and group average metrics of cortical morphometry at birth. Dendrograms shows cell classes with similar differential maturity across cortical regions at around the time of birth.

Finally, we re-calculated regional genetic maturity using only the expression profiles of genes expressed by each cell class. By comparing regional variation in cell-specific genetic maturity at 260 days post-conception with each cortical metric, averaged across the group (Figure 4E), we found that measures of cortical microstructure (fICVF, T1/T2 contrast, MD, FA and ODI) correlated most strongly with the differential timing across regions of gene expression associated with oligodendrocytes (r=0.76, 0.62, -0.58, 0.69 and 0.42 respectively). Where FA and fICVF displayed similar correlations across multiple cell classes, mean T1/T2w contrast and MD showed some specificity, correlating most strongly with oligodendrocytes (r=0.62, -0.58 respectively), excitatory neurons (r=0.47, -0.44) and endothelial cells (r=0.55, -0.42). In contrast, regional variation in cortical thickness correlated most strongly with both differential maturity in genes expressed by inhibitory neurons (r=0.43).

The principal imaging gradient is altered following preterm birth

Based on this evidence, we hypothesised that an interruption to the length of gestation would yield differences in cortical morphology indexed by variation along the principal imaging gradient. To test this, we compared cortical morphology in healthy neonates (n=292) to a cohort of preterm-born infants scanned at term-equivalent age (n=64, 59% male; mean [S.D] gestational age at birth = 32.00 [3.88] weeks).

First, we projected each individual's region × metric matrix onto the first principal imaging component (Fig S8). After correcting for age at scan and sex, regional variation along PC1 explained significantly less variance in preterm individual's imaging data than those born at term (ANCOVA: F=7.9, p=0.005; Fig 5A). Across both groups, the mean variance explained by PC1 increased with age (Fig 5A; F=46.0, p<0.001), with a stronger association in the preterm cohort (interaction: F=6.63, p=0.01) suggesting that the establishment of a principal imaging gradient is ongoing around the time of birth and altered by events surrounding preterm birth.

As the same set of eigenvectors are used to project each individual's data into the principal gradient space, differences in the variance explained by PC1 are dictated by individual differences in cortical metrics. We sought to establish the specific effects of preterm birth on cortical development by investigating group differences across all measures. Using mixed effects linear models including effects of age, birth status and regional PC1 score, we confirmed a significant main effect of birth status on all cortical metrics except for ODI (Table S6-8).

The largest effect was evident in cortical T1w/T2w contrast ($F_{1,354}$=135.53, p<0.0001, Cohen's *d*=1.62; Table S8). On average, cortical T1w/T2w was significantly lower in preterm infants with post hoc analysis confirming this effect was apparent across all image clusters (GLM correcting for age at scan and sex: all p<0.001). In addition, in the preterm cohort, lower gestational age at birth was associated with a lower T1w/T2w contrast in all cortical clusters (GLM: all p<0.01). To a lesser extent, both intracellular volume fraction and FA were, on average, higher in preterm infants (d=0.32, 0.55 respectively), although the direction of this effect was not consistent across cortical regions (Fig S9).

In contrast, average cortical mean diffusivity (*d*=-1.17) and, to a lesser extent, cortical thickness (*d*=-0.65) were higher in preterm infants, with the largest differences in primary visual and sensorimotor regions (clusters 4 and 3, respectively; Fig S9).

The magnitude of regional group differences across all cortical metrics varied as a function of PC1 (Table S7). This effect was again most apparent in T1w/T2w contrast (Fig 5B) where the differences between term and preterm groups formed a strong negative association with PC1 ($r = -0.78$, $p=0.023$ after FDR correction). Similar trends were seen in the other metrics, although none reached significance ($|r| = 0.32$ to 0.68, all $p>0.05$).

We show cortical differences following preterm birth occur in line with the principal gradient and are most apparent in T1/T2w contrast. As such, regional variation in the T1w/T2w contrast acts as a sensitive marker of both tissue maturity in the healthy newborn brain and the adverse impact of preterm birth on cortical development.

Potential vulnerability of cellular processes to the timing of preterm birth

Finally, we investigated the potential that the differences observed in preterm cortex may reflect a selective vulnerability in specific cell populations due to coincidental timing of extrauterine exposure following preterm birth and temporal variations in gene expression. Focusing on the cortical differences observed in T1/T2w contrast, we first estimated gene expression trajectories over the latter stages of gestation (160 to 260 post-conceptional days, approximately 25 to 39 weeks gestational weeks). We then split this preterm period into 10 age windows and within each, we identified genes with expression significantly correlated to the magnitude of group differences in T1/T2w contrast at term-equivalent age (FDR-corrected $p<0.05$, Fig 5C), and tested for the enrichment of gene expression by each of 10 fetal cell types within each windowed gene set. In the early preterm period, we found that mean regional differences in T1w/T2w contrast at term-equivalent ages were significantly associated with genes expressed by both inhibitory and excitatory neurons (windows 1, 2, 3 and 5, hypergeometric statistic: $p<0.05$). In contrast, later in gestation, T1w/T2w differences correlated with the expression of genes enriched for microglia and endothelial cells (windows 8,9 and 10; all $p<0.05$). However, genes enriched for oligodendrocyte expression were significantly associated with T1/T2w differences across the full preterm period (windows 1,2 and 8,9 and 10, all $p<0.05$; Fig 5C, middle). We confirmed the association with oligodendrocyte cell lineage by performing an independent cell-specific enrichment analysis[56] of genes correlated with T1/T2w differences across the preterm period (Fig S10).

We identified genes expressed by oligodendrocytes across multiple age windows (Fig 5D). We found that genes associated with T1w/T2w differences changed across the preterm period, with some exhibiting high expression early (e.g.: *NPC1, AQP6, ANKS1B*) or late (e.g.: *PLLP, MOBP)* and several expressed across the full period (e.g.: *MP1, TOMIL2, OMG*). Using the STRING database,[57] we identified protein-protein interactions between genes expressed across at least 3 age windows (Fig

5E) and found 7 interaction networks, 4 including more than 3 genes (Table S8). We performed a functional enrichment analysis of Reactome pathways[58] to identify specific molecular processes involving genes in each PPI network and identified significantly enriched pathways in 4 networks (each FDR<0.05; Fig 5E, Table S9).

Pathway enrichment analysis revealed significant gene associations across multiple time windows. Genes involved in NMDA signalling in the MAPK/ERK pathway (HSA-438066, HSA-442729, HSA-442982; *DLG1, GRIN2A*) were significantly correlated to T1w/T2w differences across the majority of the preterm period. In contrast, regional expression of genes associated with the MyD88 signalling cascade (HSA-975871; *S100B*, *RPS6KA2*) were most closely correlated to T1w/T2w differences in the latter stages of gestation (windows 5 to 9 and 7 to 10, respectively). Other pathways linked genes expression over multiple time periods. Neurotrophin signalling pathways included genes *OMG* (correlated between windows 1-8) and *ARHGEF10* (windows 2-5), and the Rho-GTPase signalling pathway (HSA-194840) included both *ARHGEF10* and *RHOB* (windows 5-10). Finally, sphingolipid metabolism pathways included genes expressed across both the full window (*ACER3*, windows 2-9) and specifically in later gestation (*SMPD1*, windows 8-10). This highlights multiple metabolic signalling pathways occurring in developmental oligodendrocyte populations during the period most at risk of interruption by preterm birth with a regional specificity correlated to neuroimaging markers of preterm brain injury at birth.

To further explore the potential functional role of genes associated with preterm brain differences, we performed an additional mammalian phenotype enrichment using the top 5% genes identified in each of the preterm age windows (n=1579 total). This yielded significant associations with several morphological and cerebral growth terms based on phenotypes surveyed in genetically-modified rodent models including slow embryonic growth (p=0.0003), decreased brain size (p=0.002) and impaired learning (p=0.043) (Table S10) providing pseudo-experimental evidence for the sensitivity of our approach to identify relevant disruptions to biological processes related to early brain growth.

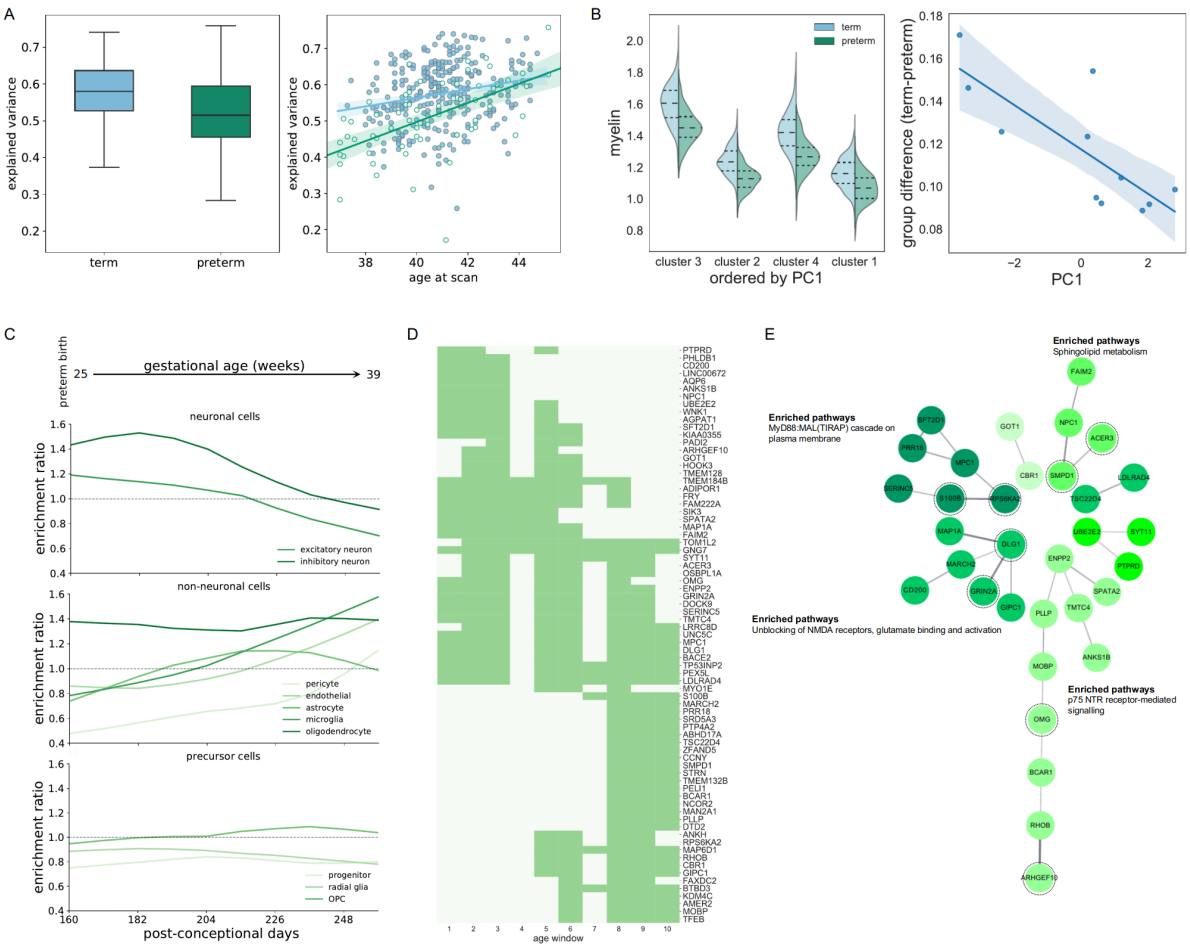

Figure 5: Disruption of the principal imaging gradient in preterm-born infants. A. Group difference in individual variance explained by the principal imaging gradient in term (blue) and preterm (green) infants (left) and the relationship between age at scan and variance explained by the principal gradient across all cortical metrics (right). Regression lines are shown for term (blue) and preterm (green) infants with 95% C.I. B. Group differences in T1w/T2w contrast, averaged across regions within each cluster (left). Clusters are ordered by position along the principal imaging gradient. Correlation between regional group differences in T1w/T2w contrast and PC1 score (left, regression shown with 95% C.I.). C. Enrichment of gene sets from 10 fetal cortical cell classes (top: neuronal; middle: non-neuronal; bottom: precursor) based on genes significantly associated (FDR p<0.05) with group differences in T1w/T2w contrast at 10 timepoints in the preterm period. D. Genes expressed by oligodendrocytes and significantly associated with group differences in T1w/T2w contrasts across at least 3 age windows are shown. Dark green indicates periods where gene expression and T1w/T2w contrast were significantly correlated for each gene (FDR p<0.05) across the preterm period. E. Protein-protein interaction networks derived using STRING with the genes listed in D. Top functional enrichments of molecular pathways are shown where applicable. Genes within each enriched pathway are highlighted.

Discussion

In this study, we aimed to test the hypothesis that noninvasive markers of brain development reflect the spatiotemporal patterning of genes underlying corticogenesis during gestation. We found that gene expression in the fetal cortex is mirrored by a principal mode of variation in the neonatal cortex. Specifically, regional variation in morphometry reflects differences in developmental maturity across cortical areas, indexed by the differential timing of gene expression across multiple cell types in the fetal cortex. Having established this relationship, we found that interruption to gestation through preterm birth resulted in a significant disruptions to MRI-based measures of cortical development by the time of normal birth. Further, the effects of preterm birth are temporally and spatially coincident to developmental processes involving the differentiation and specialisation of cortical oligodendrocyte populations. This work provides an experimental framework to link molecular developmental mechanisms to macroscopic measures of cortical anatomy in early life, demonstrating the relationship between fetal gene expression and neonatal brain development and highlighting the specific impact of early exposure to the extrauterine environment due to preterm birth.

Using advanced MRI acquired close to the time of birth in a large, healthy neonatal population, we mapped multiple measures of cortical morphometry to a single mode of variation, or principal gradient. This gradient represented a broadly hierarchical organisation in the neonatal brain, with lower order sensory and motor regions situated opposite to higher-order regions including parietal, frontal and superior temporal cortex. Cortical hierarchies are a common organisational feature of the mammalian brain,[59–65] represented by regional variations in cell populations,[62] gene expression,[16,66] and connectivity[59] as well as MRI-based measures of functional topography[67] and cortical morphometry in both adults[68] and infants.[69] The optimal mapping of cortical properties onto one or two lower dimensions remains an area of active research,[2] however, several studies have demonstrated that variation along one hierarchy is largely reflected by differences in another,[63,70,71] suggesting that lower-order representations of cortical organisation largely capture shared views of latent neurobiological variation. An important benefit of this approach is the reduction of multiple metrics into a single measure per subject. In our case, this takes advantage of the inherent redundancy across multiple MRI measures of the same cortical regions, producing a latent representation of cortical morphometry across scales. Here, we applied a simple linear mapping, arranging cortical regions along a single dimension using PCA. This was sufficient to explain a significant proportion of variation in MRI-based metrics, with regions with similar cortical profiles clustering together along the principal gradient. Organisation along this gradient was correlated with spatial gene expression gradients and measures of cortical maturity based on developmental variations in gene expression across multiple cell types. Comparison to cell-specific gene expression profiles in late gestation suggested that MRI-based markers of cortical morphology at birth correlated with genes expression by cell populations involving oligodendrocytes, maturing neurons and endothelial cells. This correlation potentially reflects a spatial variation in the developmental timing of processes associated with myelination, neuronal

arborisation and the continued maturation of the brain's vascular networks at around the time of birth.[72–78]

The advent of modern transcriptomic technologies has enabled detailed analyses of the foundational molecular mechanisms underpinning corticogenesis in the human fetal brain.[9,10,12,52,53] Resolved to the level of individual cells, recent studies have performed systematic explorations of gene expression dynamics across cell-cycle progression, migration and differentiation of several major cell types in the fetal brain.[10,53] Combined with regional expression levels of bulk tissue mRNA measured across multiple cortical areas, this allows the spatio-temporal mapping of cell-specific gene expression profiles in the developing brain.[52] Here, we used a development atlas of gene expression, measured across 11 cortical regions from 12 to 37 post-conceptional weeks in 16 separate brain specimens.[12] This data resource provides unparalleled access to the developmental mechanisms ongoing in the cortex during gestation. We found that cortical gene expression during gestation varies along a spatial gradient described by MRI-based markers of cortical morphometry that appears to reflect the differential intrinsic timing of developmental processes across regions.[6,9] We found that a number of genes vary across cortical areas in line with the imaging gradient. In particular, we found that genes with relatively higher expression in higher order regions during gestation were associated with developmentally earlier processes including neuronal differentiation and migration and were predominantly expressed by intermediate precursor cells and early-maturing inhibitory neurons. Using an alternative approach in four mid-gestation brain samples (aged 16-21 pcw), Miller et al. identified a generally rostro-caudal gradient of gene expression progressing along the contours of the developing brain and anchored in frontal and temporal cortex.[9] While some overlap was evident, 60/73 (82%) frontally-enriched genes included in both studies were also positively correlated with the imaging gradient, this may indicate the presence of multiple overlapping intrinsic hierarchies or cellular gradients underlying cortical development.[62,66] Using a machine-learning approach designed to accommodate the large number of genes assayed, we established that the maturation of a given tissue sample could be accurately determined based on temporally-evolving profiles of gene expression. Using the relative advancement or delay in tissue maturity across regions, we observed a correlation between emerging differences in regional genetic maturity and the MRI imaging gradient at term. We identified an interaction between the relative rate of development across regions and length of gestation. This was most notable in the protracted developmental trajectory of the visual cortex in midgestation, as noted elsewhere.[9] Overall, our results lend evidential support to the presence of heterogeneous corticogenetic timing over gestation.[79–83]

We hypothesised that interruption to gestation would lead to cortical disruptions aligned to the principal imaging gradient, and therefore reflecting a deleterious interaction with genetically-determined developmental programs ongoing in the cortex in the latter stages of gestation. To test this, we compared cortical development in healthy newborns to a cohort of preterm-born infants scanned at the time of normal birth. In line with previous observations,[24,35,46,84–86] we found significant differences across most cortical metrics of macro- and microstructure in the preterm brain. The

magnitude of differences between cohorts aligned with the principal imaging gradient suggesting a differential impact of perinatal adversity on cortical development that is potentially encoded by a selective vulnerability across regions due to differential rates of maturation. Adverse intrauterine environments can result in altered patterns of cortical gene expression [87–89] however, we remain cautious on speculating about the causal mechanisms that may underlie the relationships observed in this study without further empirical evidence. The largest effect was observed in the myelin-sensitive T1w/T2w contrast. In adults, regional variation in cortical T1w/T2w contrast is high correlated with quantitative MRI-measures of intracortical myelin and histological maps of cytoarchitecture.[90–92] Myelination in the neonatal cortex is minimal, however, T1w and T2w signal vary as a function of position in the neonatal cortex and the transcortical pattern of T1w/T2w ratio observed in this study mirrors closely that reported in older cohorts, with high values predominant in primary sensory regions.[90,93] In addition, we find that genes with expression correlated to mean group differences in T1w/T2w are enriched for oligodendrocyte expression across the second half of gestation. This mirrors earlier reports, based on microarray data, of correlations between neonatal imaging phenotypes and glial gene expression during gestation.[94] Using a time-resolved analysis, we found several molecular pathways involving genes with spatial and temporal correlation to the potential timing of preterm birth. These included neurotrophic and Rho-GTPase signalling pathways, associated with oligodendrocyte maturation and myelination;[95,96] the MAPK/ERK signaling pathway, associated with oligodendrocyte proliferation,[97] as well as sphingolipid metabolic pathways. We have previously identified risk alleles in preterm born infants in genes involved in lipid metabolism in the developing brain and associated with altered patterns of brain development by term-equivalent age.[98,99] This highlights a potential shared mechanistic pathway by which preterm birth can lead to altered development due to coincidental timing with important myelogenic processes in the developing cortex.

In conclusion, we show that noninvasive imaging of the neonatal brain is sensitive to the differential timing of fetal gene expression across cortical hierarchies. In addition, we find that disruption to this developmental programming by preterm birth results in significant cortical alterations that appear to reflect the selective vulnerability of developing oligodendrocytes in the mid-fetal cortex.

Materials and Methods

Subjects

Infants were recruited and imaged at the Evelina Newborn Imaging Centre, St Thomas' Hospital, London, UK for the Developing Human Connectome Project (dHCP). The study was approved by the UK Health Research Authority (Research Ethics Committee reference number: 14/LO/1169) and written parental consent was obtained for all participants. Neuroimaging and basic demographic data

from the dHCP are available to download from: http://www.developingconnectome.org/second-data-release/

In total, 442 healthy, term-born infants (gestational age at birth > 37 weeks) scanned between February 2015 and November 2018 as part of the dHCP were included. From this cohort, n=362 were successfully processed via the dHCP structural processing pipeline[24] and included after quality control (see below). Of these, diffusion data from n=296 was successfully processed using both DTI and NODDI pipelines.[100,101] A further four subjects were excluded following a final visual inspection due to cropped anatomical images.

Of 107 preterm infants (gestational age at birth < 37 weeks) scanned at term-equivalent age during the same period, one was excluded due to incomplete demographic data, n=84 completed structural MRI processing and n=67 passed diffusion processing after quality control. A further n=3 were removed after final visual inspection.

The final cohort comprised n=292 healthy term-born infants (54% male, mean [S.D] postmenstrual age at birth=39.96 [1.10] weeks, mean [S.D.] age at imaging=40.94 [1.56] weeks) and n=64 preterm infants scanned at term-equivalent age (59% male; born 32.00 [3.88] weeks and imaged at 40.57 [2.25] weeks).

Magnetic Resonance Imaging

MRI was performed on a 3T Philips Achieva (Philips, Netherlands) using a dedicated neonatal imaging system including a neonatal 32 channel phased array head coil.[27] Infants were imaged without sedation. T1- and T2-weighted anatomical images were acquired alongside diffusion MRI and resting state functional MRI (total acquisition time: 63 minutes). Inversion-recovery T1-weighted and T2-weighted images were acquired in sagittal and axial orientations (in-plane resolution: $0.8 \times 0.8mm^2$, slice thickness: 1.6mm with 0.8mm overlap) with TR=4795ms; TI=1740ms; TE=8.7ms; SENSE: 2.27 (axial) and 2.66 (sagittal) for T1-weighted images and TR=12000ms, TE=156ms; SENSE: 2.11 (axial), 2.60 (sagittal) for T2-weighted. Diffusion MRI was acquired with a spherically-optimised set of directions over 4 b-shells ($b=0s/mm^2$: 20 directions; b=400: 64 directions; b=1000: 88 directions; b=2600: 128 directions) with a multiband factor acceleration of 4, TR=3800ms; TE=90ms; SENSE: 1.2 and acquired resolution of 1.5mm × 1.5mm with 3mm slices (1.5mm overlap).[26]

T1- and T2-weighted image stacks were motion corrected and reconstructed using the multi-slice aligned sensitivity encoding method with integration into a 3D volume using a super-resolution scheme.[28,102] Multislice dMRI volumes were reconstructed using an extended SENSE technique.[103]

Image processing

T1- and T2-weighted images were processing using the dHCP structural pipeline (https://github.com/BioMedIA/dhcp-structural-pipeline).[24] Briefly, T2-weighted images were bias corrected (N4),[104], brain-extracted (BET)[105] and segmented into grey matter, white matter and cerebrospinal fluid using DRAW-EM.[106] Cortical surfaces of the right and left hemisphere were then extracted[107] and aligned to a population-specific cortical template[93] using spherical inflation and multimodal surface matching (MSM) with higher order constraints (https://github.com/ecr05/MSM_HOCR).[108,109] This method ensures that all surfaces across participants have one-to-one vertex correspondence with the dHCP neonatal template. For each subject, we extracted the following metrics: cortical thickness (corrected for cortical curvature) and T1w/T2w contrast (calculated using rigidly aligned T1-weighted images).

Diffusion-weighted images were preprocessed by first denoising[110] and removing Gibbs ringing artefacts,[111] followed by a slice-to-volume motion and distortion correction with a slice-level outlier rejection using a multi-shell spherical harmonic signal representation (SHARD).[112] Visual inspection of the 4D images ensured motion correction and outlier rejection was successful and that images of poor quality were excluded from further analysis.

We fit each subject's diffusion data with both a diffusion tensor model, fitted to the b=1000s^2/mm shell and implemented in MRtrix,[113] and the NODDI (Neurite Orientation Dispersion and Density Imaging) model,[101] fit to all shells. NODDI was implemented with the NODDI MATLAB toolbox using the *invivopreterm* tissue type options with the default parameters of intrinsic diffusivity fixed to 1.7 x 10^{-3} mm^2/s and the starting point for values considered as the fraction of intra-neurite space lowered to 0-0.3 (instead of 0-1 in the adult brain) to better fit higher water content in the newborn compared to the mature adult brain.[114]

From these models, we derived parametric maps of fractional anisotropy (FA) and mean diffusivity (MD) from DTI, as well as maps of orientation dispersion index (ODI), quantifying the angular variation of neurite orientation within a voxel and intra-cellular volume fraction (fICVF), indexing the tissue volume fraction restricted within neurites. Cortical diffusion maps were projected to the cortical surface after co-registration with the corresponding anatomical data.

Images were visually inspected after acquisition and after reconstruction, and following each processing pipeline. Any images that failed to successfully complete the processing pipelines or failed visual inspection at any stage were removed from further analysis.

Bulk tissue gene expression data

Preprocessed, bulk tissue cortical gene expression data were made available as part of the PsychENCODE project (available to download at: http://development.psychencode.org/).[12] Tissue

was collected after obtaining parental or next of kin consent and with approval by the institutional review boards at the Yale University School of Medicine, the National Institutes of Health, and at each institution from which tissue specimens were obtained.

In brief, mRNA data were available for post-mortem human brain tissue collected from n=41 specimens aged between 8 post-conceptional weeks (pcw) and 40 postnatal years. For each specimen, regional dissection of up to 16 cerebral regions was performed, including 11 neocortical regions (dorsolateral frontal cortex, DLPFC; ventrolateral frontal cortex, VLPFC; orbitofrontal cortex, OFC; medial frontal cortex, MFC; primary motor cortex, M1; primary sensory cortex, S1; inferior parietal cortex, IPC; primary auditory cortex, A1C; superior temporal cortex, STC; inferior temporal cortex, ITC; primary visual cortex, V1), and five sub-cortical regions (hippocampus, amygdala, striatum, thalamus and cerebellar cortex). Detailed anatomical boundaries for each cortical region at each stage of development are provided elsewhere.[11,12]

Regional tissue samples were subject to mRNA-sequencing using an Illumina Genome Analyzer IIx (Illumina, San Diego, CA) and mRNA-seq data processed using RSEQtools (v0.5).[115] For each sample, reads were aligned to the human genome assembly hg38/GRCh38 and filtered to only include only uniquely mapped reads, and to exclude mitochondrial reads. Gene expression was measured as reads per kilobase of transcript per million mapped reads (RPKM). Finally, conditional quantile normalisation was performed to remove GC-content bias and ComBat used to remove technical variance due to processing site (Yale or USC).[12,116,117]

In this study, we included RPKM data from neocortical samples of prenatal specimens aged 12 post-conceptional weeks onwards (n=16, age range = 12-37 pcw, mean [S.D.] age = 18.4 [7.7] pcw, 50% male, mean [S.D.] number of cortical regions sampled = 9.75 [1.6], mean [S.D.] post-mortem interval = 7.1 [12.6] hours, mean [S.D.] RNA integrity number[RIN][118] = 9.26 [0.73]). Prenatal specimens from the earliest developmental window (8-9 postconceptional weeks) were excluded as some cortical regions (e.g.: M1 and S1) were combined together to account for immature cortical anatomy.[11,12]

The prenatal gene expression data was initially filtered to only include protein-coding genes (NCBI GRCh38.p12, n=18,766 out of a possible 20720). In order to restrict our analysis to focus on genes expressed in the developing cortex, we further filtered this list to only contain genes expressed by cells in the fetal cortex based on the composite list of prenatal cell markers from five independent single-cell RNA studies of the developing fetal cortex (see *'Genetic markers of cell type'* below). This resulted in expression data from a final set of 5287 genes.

Cortical regions-of-interest

To facilitate comparison between developmental RNA and MRI data, we created a set of cortical regions-of-interest (ROI) labels corresponding to the anatomical dissections used for mRNA analysis and aligned to the dHCP imaging data.

To achieve this, we used a reference post-mortem MRI dataset acquired as part of the Allen Institute BrainSpan Atlas of the Developing Human Brain. Details of tissue processing and MRI acquisition are available at: https://help.brain-map.org/download/attachments/3506181/BrainSpan_MR_DW_DT_Imaging.pdf. In brief, MRI was acquired at 3T and 7T (Siemens, Germany) in three post-mortem, whole-brain specimens aged 19, 21 and 22 pcw. In addition, anatomical annotations corresponding to the regional dissections in Miller et al.,[9] Kang et al.[11] and Li et al.[12] were provided on a reconstructed cortical surface from a 19pcw prenatal specimen.[119] Cortical ROI data were available to download in VTK file format, separately for left and right cortical hemispheres (Figure S1).

To generate a set of dHCP-compatible cortical labels, we reconstructed the cortical surface of the 3T 22pcw reference image. First, manually creating a brain mask to remove non-brain tissue, then smoothing using a mean filter of 3mm width. We performed automated tissue segmentation on the smoothed image using the dHCP structural pipeline, manually correcting tissue segmentations on a slice-by-slice basis for accuracy prior to cortical surface reconstruction. Using dHCP tools, the fetal cortical surface was extracted and cortical labels manually transferred onto it based on the reference labels provided by Huang et al.[119] and anatomical descriptions in Li et al.[12] Finally, the fetal surface was inflated to a sphere and co-registered to the earliest timepoint (36 weeks gestational age) of the dHCP cortical surface atlas using multimodal surface matching (MSM).[93,109]

This resulted in a set of 11 cortical ROI, each associated with regional bulk tissue mRNA data sampled across gestation and co-registered with dHCP neuroimaging data to allow correspondent sampling of cortical imaging metrics in the neonatal brain (Fig S1).

Cortical imaging metric analysis

For every subject, mean values of each imaging metric (thickness, T1w/T2w contrast, FA, MD, fICVF, ODI) were calculated within each cortical ROI. Metric values were averaged across hemispheres and outlier values identified and removed using a median absolute deviation (MAD) of > 3.5.

For all healthy term-born infants, regional metrics were Z-transformed and averaged across subjects to produce a group average region × metric matrix representing the relative variation of each imaging metric across cortical regions. We performed hierarchical cluster analysis of regions using average linkage based on the cosine distance between imaging metrics. The optimal number of clusters (from between 2 and 7) was chosen based on the maximum silhouette score.[50] Regional clustering was repeated with 10,000 bootstrapped samples, selecting *n* subjects with replacement before calculating the average matrix and clustering.

We compared the clustering solution to two null models. To test if intra-cluster cohesion of the four-cluster solution was significantly greater than expected by chance, we built a null distribution by first permuting the regional cluster assignments 10,000 times before calculating the silhouette score of the resulting clusters. To test the null hypothesis that all regional metrics were drawn from the same distribution rather than from separate clusters, we drew eleven samples (one per cortical ROI) from a single multivariate normal distribution with mean 0 and covariance specified by the group average region × metric matrix. We then repeated the clustering and calculated how well pairwise distances between the original metric profiles are preserved by the null clustering solution via the cophenetic correlation. We repeated sampling 10,000 times to build a null cophenetic correlation distribution.

We projected the group average data to two dimensions using Principal Component Analysis (PCA) via eigendecomposition of the data covariance matrix. This results in a set of L eigenvectors, W_L, that map the original $n \times p$ data matrix, X onto a set of orthogonal axes as: $T_L = XW_L$. As generally $L < p$, the truncated $n \times L$ matrix, T_L, forms a low-dimensional representation of the original data. We can then project each subject's region × metric matrix, X_s, onto a common set of axes as $T_L^s = X_s W_L$, where T_L^s represents the L component scores for each subject, s.

All analysis was performed in Python (3.7.3) using Scipy (1.3.0)[120] and Scikit-Learn (0.21.2).[121]

Modelling gene expression trajectories

For each gene, we modelled the relationship between gene expression and specimen age using mixed-effects models. Using RPKM data described above, each gene's expression data were first Winsorised to set very small or large outlying values to the 5th and 95th centile values, respectively, to stabilise against extreme values before log2-transformation.

We initially compared two models, modelling regional gene expression as either a linear or nonlinear function of age with fixed effects of sex and RNA integrity number. We accounted for sample-specific variation by including in the model a random intercept for each specimen, such that:

$$y \sim f(v) + X\beta + Zb$$

Where $f(\cdot)$ is a nonlinear function of predictor v, X is an m-observation $\times p$ design matrix modelling p linear, fixed effects and Z is an $m \times (n \cdot r)$ design matrix modelling r random effects across n specimens. In this case, age was included as either a nonlinear predictor, $f(v)$, or as a fixed linear effect alongside sex and RIN. We specified a relatively smooth nonlinear function of age using a natural cubic spline with four knots evenly spaced across the age span. To estimate region-specific trajectories, we calculated a second nonlinear model, additionally including separate smooth functions for each cortical region. Models were compared using AIC and BIC.

We calculated age-corrected RPKM values for each gene in all cortical samples using the residuals of the best-fit nonlinear mixed model (Fig S4) to test the spatial association between gene expression and the principal imaging gradient using non-parametric correlation (Kendall's τ).

Modelling was performed in R (3.6.1) using *nlme*[122] and *mgcv*[123] packages.

Genetic markers of cell type

Genetic markers of cortical cell types were collated from five independent single-cell RNA studies of the fetal cortex.[10,12,51–53] Using single-cell RNA-seq, each study identified sets of genes differentially expressed across cell clusters or types. Cell types were independently defined in each study and a list of all cell types included in this study (n=87) are shown in Table S2. Where applicable, for a given cell type, differentially-expressed genes were included as cell type markers if they were found to be expressed in at least 50% all cells surveyed.[12,52,53] Across all five studies, each cell type was manually assigned to one of 11 cell classes based on text descriptions from each study (astrocyte, endothelial cell, microglia, neuron:excitatory, neuron:inhibitory, neuron:unclassified, oligodendrocyte, oligodendrocyte precursor cell [OPC], pericyte, intermediate progenitor cell, radial glia) and classified as either a precursor or mature cell type (Table S2). For each cell class, omnibus gene lists were created by collating identified gene markers for all cell types within a class. Unique gene lists were created by excluding any genes identified as a marker of more than one cell class.

Cell type embedding

Using the region-specific, nonlinear model specified above, expression trajectories for every gene were estimated for each region at 50 evenly spaced points across the full observation window (12pcw - 37pcw). For each cell type identified in the fetal cortex (see above), expression trajectories for all cell-type gene markers were normalised to unit length, concatenated over regions and averaged to capture both temporal and spatial variation in average gene expression across cell types. Similarity between cell-type gene expression trajectories were then visualised by embedding into a two-dimensional space using Uniform Manifold and Approximation Projection (UMAP) based on Euclidean distance.[55]

Enrichment analyses

We performed over-representation analysis (ORA) of each list of gene markers for each of 10 cell classes (excluding neuron:unclassified), calculating the hypergeometric statistic:

$$p = 1 - \sum_{i=0}^{x} \frac{\binom{K}{i}\binom{M-K}{N-i}}{\binom{M}{N}}$$

Where p is the probability of finding x or more genes from a cell-class-specific gene list K in a set of randomly selected genes, N drawn from a background set, M. We calculated enrichment ratios as the

proportion of cell-class-specific genes in the gene list of interest, compared to the proportion in the full background set. The background gene set was defined as the full list of protein-coding genes included in the analysis (n=5287). We corrected for multiple comparisons across cell classes using False Discovery Rate (FDR).

We additionally performed ORA for Gene Ontology terms using WebGestalt.[124]

Weighted Gene Correlation Network Analysis

We used WGCNA[54] to identify co-expression modules within PC+ and PC- gene sets. We performed topology analysis using a gene × gene adjacency matrix constructed from the residualised log2-transformed RPKM data, after accounting for variance due to age, sex and sample effects (see Modelling gene expression trajectories, above). A soft threshold was chosen to approximate scale free topology in the adjacency matrix (PC+: power=5, r^2=0.77; PC-: power=10, r^2=0.78),[125] before transformation into a topological overlap matrix. Hierarchical clustering was used to assign genes to modules based on the dynamic tree-cutting method.[126] Analysis was performed in R (3.6.1) with the *WGCNA* package.[54]

Predicting genetic maturity

We used gene expression over time to construct a predictive model of genetic maturity using a kernel-based, regularised regression. In machine learning, kernels provide a method to compute the product of two (possibly high-dimensional) vectors as represented in some (unknown) feature space. This allows linear algorithms to learn non-linear functions, and is particularly useful in settings where $n<<p$, as the method can avoid an explicit mapping to the high-dimensional feature space.

Using the n=120 regionally-varying genes (PC+ and PC-), we first calculated regional gene expression profiles, corrected for variance due to sex, RIN and specimen ID while retaining variance due to age, using previously estimated nonlinear models. We then averaged gene expression across cortical regions in each specimen to create a specimen × gene (16 × 120) mean gene expression matrix, where each row represents the normalised log2(RPKM) of each gene for a given specimen, averaged across cortical regions.

To calculate regional variation in genetic maturity, we implemented a leave-one-out (LOO) model using kernel ridge regression (Scikit-Learn; default regularisation parameter, α=1.0), modelling the association between specimen age (in post-conceptional days) and mean cortical gene expression data in 15 out of 16 specimens. We then used this model to predict age using the *regional* gene expression profiles of the remaining, left-out specimen, resulting in eleven age predictions, one per cortical region. We repeated this process, leaving out a different specimen each time.

We calculated a genetic maturity index by subtracting predicted age from each of the eleven predicted cortical sample ages. Thus, for a given age, developmentally more mature regions were expected to express a gene profile more similar to older specimen's mean cortical profile, and thus will have an 'older' predicted age compared to developmentally less mature regions, resulting in a positive genetic maturity index. In order to estimate a stable genetic maturity index, we repeated the modelling using a bootstrapped selection of genes, repeating gene sampling with replacement 5000 times. We also repeated the model using all 5287 genes. We calculated the correlation between regional genetic maturity and PC1 score for each specimen and tested the significance of this relationship by permuting mean gene expression profiles with respect to specimen age 1000 times during model training. For further analysis, we used mean expression from gene sets associated with each cell type to calculate regional genetic maturity.

Group comparison of cortical morphology

We compared regional cortical metrics in term and preterm cohorts using a linear mixed effects modelling approach. For each of six metrics, we modelled metric value as a combination of age, sex, regional PC1 score and birth group status (term or preterm). We included an interaction term for PC1 and birth status to test the hypothesis that preterm birth incurs differential effects across cortical regions in line with the principal gradient. We also included subject ID as a random effect to account for correlated within-subject observations across regions. We fit nested models by Maximum Likelihood, comparing model fits with and without the inclusion of birth status using AIC and BIC (Table S5).

For each metric, we performed post hoc analysis within each imaging cluster, modelling average cluster value as a function of age, sex and birth status.

Developmental gene enrichment

In order to test cell class enrichment over time, we split the preterm period (approximately 160 to 260 post-conceptional days) into 10 age windows. Using nonlinear gene expression trajectories, calculated across cortical regions (see 'Modelling gene expression trajectories' above), we averaged modelled gene expression within each window for every cortical region. Then, in each window, we calculated the non-parametric association (Kendall's τ) between gene expression and the mean difference between term and preterm groups in T1w/T2w contrast in each cortical region and recorded significantly associated genes (FDR-corrected at $p<0.05$). Finally, we performed cell-class enrichment (see 'Enrichment Analyses' above), in each of the 10, time-resolved gene sets.

Acknowledgements

Neuroimaging data were provided by the developing Human Connectome Project, KCL-Imperial-Oxford Consortium funded by the European Research Council under the European Union Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. [319456]. We are grateful to the families who generously supported this trial. RNA-seq data were made available via the PsychENCODE consortium supported

by the NIMH. This research was conducted within the Developmental Imaging research group, Murdoch Childrens Research Institute and the Children's MRI Centre, Royal Children's Hospital, Melbourne, Victoria. It was supported by the Murdoch Childrens Research Institute, the Royal Children's Hospital, Department of Paediatrics, The University of Melbourne and the Victorian Government's Operational Infrastructure Support Program. The project was generously supported by RCH1000, a unique arm of The Royal Children's Hospital Foundation devoted to raising funds for research at The Royal Children's Hospital.

References

1. Badre, D. & D'Esposito, M. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat. Rev. Neurosci.* **10**, 659–669 (2009).
2. Huntenburg, J. M., Bazin, P.-L. & Margulies, D. S. Large-Scale Gradients in Human Cortical Organization. *Trends in Cognitive Sciences* **22**, 21–31 (2018).
3. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
4. Mesulam, M. M. From sensation to cognition. *Brain* **121 (Pt 6)**, 1013–1052 (1998).
5. Lein, E. S., Belgard, T. G., Hawrylycz, M. & Molnár, Z. Transcriptomic Perspectives on Neocortical Structure, Development, Evolution, and Disease. *Annu. Rev. Neurosci.* **40**, 629–652 (2017).
6. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* **89**, 248–268 (2016).
7. Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and Evolution of the Human Neocortex. *Cell* **146**, 18–36 (2011).
8. Charvet, C. J. Distinct developmental growth patterns account for the disproportionate expansion of the rostral and caudal isocortex in evolution. *Front. Hum. Neurosci.* **8**, (2014).
9. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
10. Polioudakis, D. *et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785-801.e8 (2019).
11. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
12. Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, (2018).

13. Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).
14. Velmeshev, D. *et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science* **364**, 685–689 (2019).
15. Lombardo, M. V. *et al.* Maternal immune activation dysregulation of the fetal brain transcriptome and relevance to the pathophysiology of autism spectrum disorder. *Molecular Psychiatry* **23**, 1001–1013 (2018).
16. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
17. Fornito, A., Arnatkevičiūtė, A. & Fulcher, B. D. Bridging the Gap between Connectome and Transcriptome. *Trends Cogn. Sci. (Regul. Ed.)* **23**, 34–50 (2019).
18. Arnatkeviciūtė, A., Fulcher, B. D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353–367 (2019).
19. Shin, J. *et al.* Cell-Specific Gene-Expression Profiles and Cortical Thickness in the Human Brain. *Cereb. Cortex* **28**, 3267–3277 (2018).
20. Reardon, P. K. *et al.* Normative brain size variation and brain shape diversity in humans. *Science* **360**, 1222–1227 (2018).
21. Richiardi, J. *et al.* Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244 (2015).
22. Romero-Garcia, R., Warrier, V., Bullmore, E. T., Baron-Cohen, S. & Bethlehem, R. A. I. Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism. *Molecular Psychiatry* 1 (2018) doi:10.1038/s41380-018-0023-7.
23. Morgan, S. E. *et al.* Cortical patterning of abnormal morphometric similarity in psychosis is associated with brain expression of schizophrenia-related genes. *PNAS* **116**, 9604–9609 (2019).
24. Makropoulos, A. *et al.* The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *NeuroImage* **173**, 88–112 (2018).
25. Bastiani, M. *et al.* Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *Neuroimage* **184**, 801–812 (2019).

26. Hutter, J. *et al.* Time-efficient and flexible design of optimized multishell HARDI diffusion. *Magn Reson Med* **79**, 1276–1292 (2018).
27. Hughes, E. J. *et al.* A dedicated neonatal brain imaging system. *Magnetic Resonance in Medicine* **78**, 794–804 (2017).
28. Cordero-Grande, L., Hughes, E. J., Hutter, J., Price, A. N. & Hajnal, J. V. Three-dimensional motion corrected sensitivity encoding reconstruction for multi-shot multi-slice MRI: Application to neonatal brain imaging. *Magn Reson Med* **79**, 1365–1376 (2018).
29. Knickmeyer, R. C. *et al.* A Structural MRI Study of Human Brain Development from Birth to 2 Years. *J Neurosci* **28**, 12176–12182 (2008).
30. Holland, D. *et al.* Structural Growth Trajectories and Rates of Change in the First 3 Months of Infant Brain Development. *JAMA Neurol* **71**, 1266–1274 (2014).
31. Li, G. *et al.* Mapping Longitudinal Hemispheric Structural Asymmetries of the Human Cerebral Cortex From Birth to 2 Years of Age. *Cereb. Cortex* **24**, 1289–1300 (2014).
32. Garcia, K. E. *et al.* Dynamic patterns of cortical expansion during folding of the preterm human brain. *PNAS* **115**, 3156–3161 (2018).
33. Wang, F. *et al.* Developmental topography of cortical thickness during infancy. *PNAS* **116**, 15855–15860 (2019).
34. Jha, S. C. *et al.* Environmental Influences on Infant Cortical Thickness and Surface Area. *Cereb Cortex* **29**, 1139–1149 (2019).
35. Kapellou, O. *et al.* Abnormal cortical development after premature birth shown by altered allometric scaling of brain growth. *PLoS Med.* **3**, e265 (2006).
36. Wang, S. *et al.* Quantitative assessment of myelination patterns in preterm neonates using T2-weighted MRI. *Scientific Reports* **9**, 1–12 (2019).
37. Deoni, S. C. L., Dean, D. C., O'Muircheartaigh, J., Dirks, H. & Jerskey, B. A. Investigating white matter development in infancy and early childhood using myelin water faction and relaxation time mapping. *Neuroimage* **63**, 1038–1053 (2012).
38. van der Knaap, M. S. & Valk, J. MR imaging of the various stages of normal myelination during the first year of life. *Neuroradiology* **31**, 459–470 (1990).
39. Counsell, S. J. *et al.* MR Imaging Assessment of Myelination in the Very Preterm Brain. *AJNR Am J Neuroradiol* **23**, 872–881 (2002).

40. Dean, D. C. *et al.* Mapping an index of the myelin g-ratio in infants using magnetic resonance imaging. *NeuroImage* **132**, 225–237 (2016).

41. Doria, V. *et al.* Emergence of resting state networks in the preterm human brain. *Proc. Natl. Acad. Sci. U. S. A* **107**, 20015–20020 (2010).

42. Fransson, P., Åden, U., Blennow, M. & Lagercrantz, H. The Functional Architecture of the Infant Brain as Revealed by Resting-State fMRI. *Cereb. Cortex* **21**, 145–154 (2011).

43. Cao, M. *et al.* Early Development of Functional Network Segregation Revealed by Connectomic Analysis of the Preterm Human Brain. *Cereb Cortex* **27**, 1949–1963 (2017).

44. Ball, G. *et al.* The effect of preterm birth on thalamic and cortical development. *Cereb. Cortex* **22**, 1016–1024 (2012).

45. Ball, G. *et al.* Development of cortical microstructure in the preterm human brain. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9541–9546 (2013).

46. Ball, G. *et al.* Multimodal image analysis of clinical influences on preterm brain development. *Ann. Neurol.* **82**, 233–246 (2017).

47. Boardman, J. P. *et al.* Abnormal deep grey matter development following preterm birth detected using deformation-based morphometry. *Neuroimage* **32**, 70–78 (2006).

48. Counsell, S. J. *et al.* Axial and radial diffusivity in preterm infants who have diffuse white matter changes on magnetic resonance imaging at term-equivalent age. *Pediatrics* **117**, 376–386 (2006).

49. Ajayi-Obe, M., Saeed, N., Cowan, F. M., Rutherford, M. A. & Edwards, A. D. Reduced development of cerebral cortex in extremely preterm infants. *Lancet* **356**, 1162–1163 (2000).

50. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53–65 (1987).

51. Pollen, A. A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053–1058 (2014).

52. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

53. Fan, X. *et al.* Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res* **28**, 730–745 (2018).

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

55. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]* (2018).

56. Dougherty, J. D., Schmidt, E. F., Nakajima, M. & Heintz, N. Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res.* **38**, 4218–4230 (2010).

57. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

58. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).

59. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb.Cortex* **1**, 1–47 (1991).

60. Barone, P., Batardiere, A., Knoblauch, K. & Kennedy, H. Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J. Neurosci.* **20**, 3263–3281 (2000).

61. Amir, Y., Harel, M. & Malach, R. Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *Journal of Comparative Neurology* **334**, 19–46 (1993).

62. Cahalane, D. J., Charvet, C. J. & Finlay, B. L. Systematic, balancing gradients in neuron density and number across the primate isocortex. *Front Neuroanat* **6**, (2012).

63. Oligschläger, S. *et al.* Gradients of connectivity distance in the cerebral cortex of the macaque monkey. *Brain Struct Funct* **224**, 925–935 (2019).

64. Fulcher, B. D., Murray, J. D., Zerbi, V. & Wang, X.-J. Multimodal gradients across mouse cortex. *PNAS* **116**, 4689–4695 (2019).

65. Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).

66. Gomez, J., Zhen, Z. & Weiner, K. S. Human visual cortex is organized along two genetically opposed hierarchical gradients with unique developmental and evolutionary origins. *PLOS Biology* **17**, e3000362 (2019).

67. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *PNAS* **113**, 12574–12579 (2016).

68. Wagstyl, K., Ronan, L., Goodyer, I. M. & Fletcher, P. C. Cortical thickness gradients in structural hierarchies. *Neuroimage* **111**, 241–250 (2015).

69. Larivière, S. *et al.* Multiscale Structure–Function Gradients in the Neonatal Connectome. *Cereb Cortex* (2019) doi:10.1093/cercor/bhz069.

70. Huntenburg, J. M. *et al.* A Systematic Relationship Between Functional Connectivity and Intracortical Myelin in the Human Cerebral Cortex. *Cereb Cortex* **27**, 981–997 (2017).

71. Burt, J. B. *et al.* Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).

72. Weber, B., Keller, A. L., Reichold, J. & Logothetis, N. K. The Microvascular System of the Striate and Extrastriate Visual Cortex of the Macaque. *Cereb Cortex* **18**, 2318–2330 (2008).

73. Fonta, C. & Imbert, M. Vascularization in the Primate Visual Cortex during Development. *Cereb Cortex* **12**, 199–211 (2002).

74. Tieman, S. B., Möllers, S., Tieman, D. G. & White, J. The blood supply of the cat's visual cortex and its postnatal development. *Brain Research* **998**, 100–112 (2004).

75. Bystron, I., Blakemore, C. & Rakic, P. Development of the human cerebral cortex: Boulder Committee revisited. *Nat. Rev. Neurosci.* **9**, 110–122 (2008).

76. Travis, K., Ford, K. & Jacobs, B. Regional dendritic variation in neonatal human cortex: a quantitative Golgi study. *Dev. Neurosci.* **27**, 277–287 (2005).

77. Jakovcevski, I. & Zecevic, N. Sequence of oligodendrocyte development in the human fetal telencephalon. *Glia* **49**, 480–491 (2005).

78. Jakovcevski, I., Filipovic, R., Mo, Z., Rakic, S. & Zecevic, N. Oligodendrocyte Development and the Onset of Myelination in the Human Fetal Brain. *Front Neuroanat* **3**, (2009).

79. Elston, G. N. & Fujita, I. Pyramidal cell development: postnatal spinogenesis, dendritic growth, axon growth, and electrophysiology. *Front Neuroanat* **8**, 78 (2014).

80. Elston, G. N., Oga, T. & Fujita, I. Spinogenesis and Pruning Scales across Functional Hierarchies. *J. Neurosci.* **29**, 3271–3275 (2009).

81. Bianchi, S. *et al.* Synaptogenesis and development of pyramidal neuron dendritic morphology in the chimpanzee neocortex resembles humans. *Proc. Natl. Acad. Sci. U.S.A.* **110 Suppl 2**, 10395–10401 (2013).

82. Huttenlocher, P. R. & Dabholkar, A. S. Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology* **387**, 167–178 (1997).

83. Guillery, R. W. Is postnatal neocortical maturation hierarchical? *Trends in Neurosciences* **28**, 512–517 (2005).

84. Batalle, D. *et al.* Different patterns of cortical maturation before and after 38 weeks gestational age demonstrated by diffusion MRI in vivo. *NeuroImage* **185**, 764–775 (2019).

85. Dubois, J. *et al.* Mapping the early cortical folding process in the preterm newborn brain. *Cereb.Cortex* **18**, 1444–1454 (2008).

86. Bouyssi-Kobar, M. *et al.* Regional microstructural organization of the cerebral cortex is affected by preterm birth. *NeuroImage: Clinical* **18**, 871–880 (2018).

87. Antonson, A. M., Balakrishnan, B., Radlowski, E. C., Petr, G. & Johnson, R. W. Altered Hippocampal Gene Expression and Morphology in Fetal Piglets following Maternal Respiratory Viral Infection. *DNE* **40**, 104–119 (2018).

88. Garbett, K. A., Hsiao, E. Y., Kálmán, S., Patterson, P. H. & Mirnics, K. Effects of maternal immune activation on gene expression patterns in the fetal brain. *Transl Psychiatry* **2**, e98 (2012).

89. Lussier, S. & Stevens, H. Delays in GABAergic Interneuron Development and Behavioral Inhibition after Prenatal Stress. *Dev Neurobiol* **76**, 1078–1091 (2016).

90. Glasser, M. F. & Van Essen, D. C. Mapping Human Cortical Areas In Vivo Based on Myelin Content as Revealed by T1- and T2-Weighted MRI. *J Neurosci* **31**, 11597–11616 (2011).

91. Uddin, Md. N., Figley, T. D., Solar, K. G., Shatil, A. S. & Figley, C. R. Comparisons between multi-component myelin water fraction, T1w/T2w ratio, and diffusion tensor imaging measures in healthy human brain structures. *Sci Rep* **9**, (2019).

92. Shams, Z., Norris, D. G. & Marques, J. P. A comparison of in vivo MRI based cortical myelin mapping using T1w/T2w and R1 mapping at 3T. *PLoS One* **14**, (2019).

93. Bozek, J. *et al.* Construction of a neonatal cortical surface atlas using Multimodal Surface Matching in the Developing Human Connectome Project. *Neuroimage* **179**, 11–29 (2018).

94. Pecheva, D. *et al.* Neural Transcription Correlates of Multimodal Cortical Phenotypes during Development. *Cereb Cortex* doi:10.1093/cercor/bhz271.

95. Barres, B. A. *et al.* A crucial role for neurotrophin-3 in oligodendrocyte development. *Nature* **367**, 371–375 (1994).

96. Feltri, M. L., Suter, U. & Relvas, J. B. The function of RhoGTPases in axon ensheathment and myelination. *Glia* **56**, 1508–1517 (2008).

97. Ishii, A., Furusho, M. & Bansal, R. Sustained Activation of ERK1/2 MAPK in Oligodendrocytes and Schwann Cells Enhances Myelin Growth and Stimulates Oligodendrocyte Progenitor Expansion. *J Neurosci* **33**, 175–186 (2013).

98. Boardman, J. P. *et al.* Common Genetic Variants and Risk of Brain Injury After Preterm Birth. *Pediatrics* **133**, e1655–e1663 (2014).

99. Krishnan, M. L. *et al.* Machine learning shows association between genetic variability in PPARG and cerebral connectivity in preterm infants. *PNAS* **114**, 13744–13749 (2017).

100. Basser, P. J. & Pierpaoli, C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J.Magn.Reson.B* **111**, 209–219 (1996).

101. Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* **61**, 1000–1016 (2012).

102. Kuklisova-Murgasova, M., Quaghebeur, G., Rutherford, M. A., Hajnal, J. V. & Schnabel, J. A. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. *Med Image Anal* **16**, 1550–1564 (2012).

103. Zhu, K. *et al.* Hybrid-Space SENSE Reconstruction for Simultaneous Multi-Slice MRI. *IEEE Trans Med Imaging* **35**, 1824–1836 (2016).

104. Tustison, N. J. *et al.* N4ITK: Improved N3 Bias Correction. *IEEE Trans Med Imaging* **29**, 1310–1320 (2010).

105. Smith, S. M. Fast robust automated brain extraction. *Human Brain Mapping* **17**, 143–155 (2002).

106. Makropoulos, A. *et al.* Automatic whole brain MRI segmentation of the developing neonatal brain. *IEEE Trans Med Imaging* **33**, 1818–1831 (2014).

107. Schuh, A. *et al.* A deformable model for the reconstruction of the neonatal cortex. in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* 800–803 (2017). doi:10.1109/ISBI.2017.7950639.

108. Robinson, E. C. *et al.* MSM: a new flexible framework for Multimodal Surface Matching. *Neuroimage* **100**, 414–426 (2014).

109. Robinson, E. C. *et al.* Multimodal surface matching with higher-order smoothness constraints. *Neuroimage* **167**, 453–465 (2018).

110. Veraart, J., Fieremans, E. & Novikov, D. S. Diffusion MRI noise mapping using random matrix theory. *Magnetic Resonance in Medicine* **76**, 1582–1593 (2016).

111. Kellner, E., Dhital, B., Kiselev, V. G. & Reisert, M. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic Resonance in Medicine* **76**, 1574–1581 (2016).

112. Christiaens, D. *et al.* Scattered slice SHARD reconstruction for motion correction in multi-shell diffusion MRI of the neonatal brain. *arXiv:1905.02996 [physics]* (2019).

113. Tournier, J.-D. *et al.* MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* **202**, 116137 (2019).

114. Kunz, N. *et al.* Assessing white matter microstructure of the newborn with multi-shell diffusion MRI and biophysical compartment models. *NeuroImage* **96**, 288–299 (2014).

115. Habegger, L. *et al.* RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. *Bioinformatics* **27**, 281–283 (2011).

116. Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).

117. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

118. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* **7**, 3 (2006).

119. Huang, H. *et al.* Coupling Diffusion Imaging with Histological and Gene Expression Analysis to Examine the Dynamics of Cortical Areas across the Fetal Period of Human Brain Development. *Cereb Cortex* **23**, 2620–2631 (2013).

120. SciPy.org — SciPy.org. https://www.scipy.org/.

121. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825−2830 (2011).
122. nlme: Nonlinear Mixed-Effects Models in nlme: Linear and Nonlinear Mixed Effects Models. https://rdrr.io/cran/nlme/man/nlme.html.
123. Wood, S. *Generalized Additive Models: An Introduction with R*. (CRC Press, 2017).
124. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).
125. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17 (2005).
126. 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).