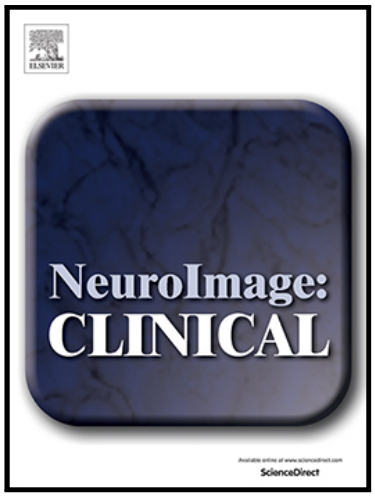
# Neonatal morphometric similarity mapping for predicting brain age and characterizing neuroanatomic variation associated with preterm birth

# Journal Pre-proof




Neonatal morphometric similarity mapping for predicting brain age and characterizing neuroanatomic variation associated with preterm birth

Paola Galdi, Manuel Blesa, David Q. Stoye, Gemma Sullivan, Gillian J. Lamb, Alan J. Quigley, Michael J. Thrippleton, Mark E. Bastin, James P. Boardman

PII: S2213-1582(20)30032-2
DOI: https://doi.org/10.1016/j.nicl.2020.102195
Reference: YNICL 102195

To appear in: *NeuroImage: Clinical*

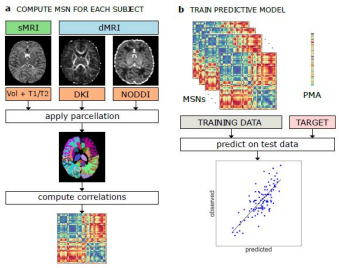
Received date: 23 October 2019
Revised date: 14 January 2020
Accepted date: 21 January 2020

Please cite this article as: Paola Galdi, Manuel Blesa, David Q. Stoye, Gemma Sullivan, Gillian J. Lamb, Alan J. Quigley, Michael J. Thrippleton, Mark E. Bastin, James P. Boardman, Neonatal morphometric similarity mapping for predicting brain age and characterizing neuroanatomic variation associated with preterm birth, *NeuroImage: Clinical* (2020), doi: https://doi.org/10.1016/j.nicl.2020.102195

**Highlights**

- Multiple MRI features are integrated in a single model to study brain maturation in newborns.
- Morphometric similarity networks (MSNs) provide a whole-brain description of the structural properties of neonatal brain.
- The information encoded in MSNs is predictive of chronological brain age in the perinatal period.
- MSNs provide a novel data-driven method for investigating neuroanatomic variation associated with preterm birth.

## Graphical Abstract



# Neonatal morphometric similarity mapping for predicting brain age and characterizing neuroanatomic variation associated with preterm birth

Paola Galdi[a,1,*], Manuel Blesa[a,1], David Q. Stoye[a], Gemma Sullivan[a], Gillian J. Lamb[a], Alan J. Quigley[b], Michael J. Thrippleton[c,d], Mark E. Bastin[c], James P. Boardman[a,c]

[a] *MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh EH16 4TJ, UK*
[b] *Department of Radiology, Royal Hospital for Sick Children, Edinburgh EH9 1LF, UK*
[c] *Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH16 4SB, UK*
[d] *Edinburgh Imaging, University of Edinburgh, Edinburgh EH16 4SB, UK*



**Abstract**

Multi-contrast MRI captures information about brain macro- and micro-structure which can be combined in an integrated model to obtain a detailed "fingerprint" of the anatomical properties of an individual's brain. Inter-regional similarities between features derived from structural and diffusion MRI, including regional volumes, diffusion tensor metrics, neurite orientation dispersion and density imaging measures, can be modelled as morphometric similarity networks (MSNs). Here, individual MSNs were derived from 105 neonates (59 preterm and 46 term) who were scanned between 38 and 45 weeks postmenstrual age (PMA). Inter-regional similarities were used as predictors in a regression model of age at the time of scanning and in a classification model to discriminate between preterm and term infant brains. When tested on unseen data, the regression model predicted PMA at scan with a mean absolute error of 0.70 ± 0.56 weeks, and the classification model achieved 92% accuracy. We conclude that MSNs predict chronological brain age accurately; and they provide a data-driven approach to identify networks that characterise typical maturation and those that contribute most to neuroanatomic variation associated with preterm

*Correspondence: Paola Galdi, Queen's Medical Research Institute, 47 Little France Crescent, Edinburgh EH16 4TJ, UK. Email: paola.galdi@ed.ac.uk
[1]These authors contributed equally to the work.

birth.

*Keywords:* `morphometric similarity networks, preterm, developing brain, brain age, multi-modal data, MRI`

---

**Highlights**

1. Multiple MRI features are integrated in a single model to study brain maturation in newborns.
2. Morphometric similarity networks (MSNs) provide a whole-brain description of the structural properties of neonatal brain.
3. The information encoded in MSNs is predictive of chronological brain age in the perinatal period.
4. MSNs provide a novel data-driven method for investigating neuroanatomic variation associated with preterm birth.

## 1. Introduction

Preterm birth is closely associated with increased risk of neurodevelopmental, cognitive and psychiatric impairment that extends across the life course (Nosarti et al., 2012; Anderson, 2014; Mathewson et al., 2017; Van Lieshout et al., 2018). Structural and diffusion MRI (sMRI and dMRI) support the conceptualisation of atypical brain growth after preterm birth as a process characterised by micro-structural alteration of connective pathways due to impaired myelination and neuronal dysmaturation (Boardman et al., 2006; Anjari et al., 2007; Counsell et al., 2008; Ball et al., 2013b; Back and Miller, 2014; Van Den Heuvel et al., 2015; Eaton-Rosen et al., 2015; Thompson et al., 2016; Batalle et al., 2017; Telford et al., 2017; Batalle et al., 2018); this leads to a "dysconnectivity phenotype" that could form the basis for long term functional impairment (Boardman et al., 2010; Caldinelli et al., 2017; Keunen et al., 2017; Cao et al., 2017; Batalle et al., 2018b). However, there has not been a unified approach that incorporates information from sMRI and dMRI to study brain maturation

in the perinatal period so the set of image features that best capture brain maturation, and support image classification, are unknown.

The majority of neonatal connectomics studies have used single modes of data such as dMRI tractography (Brown et al., 2014; Batalle et al., 2017; Blesa et al., 2019) or resting-state functional connectivity (Ball et al., 2016; Smyser et al., 2016a). An alternative connectome model is the structural covariance network (SCN) approach (Alexander-Bloch et al., 2013) in which covariance between regional measurements is calculated across subjects, resulting in a single network for the entire population. Other approaches have constructed subject-specific SCNs (Li et al., 2017; Mahjoub et al., 2018) or higher order morphological networks to model the relationship between ROIs across different views (Soussia and Rekik, 2018), but these techniques have been restricted to the use of morphometric variables available through standard structural T1-weighted MRI sequences and by using a single metric (e.g. cortical thickness) to assess the "connectivity" between nodes (Shi et al., 2012).

Based on observations that integrating data from different MRI sequences enhances anatomic characterization (Melbourne et al., 2014; Kulikova et al., 2015; Ball et al., 2017; Thompson et al., 2018a), we investigated whether whole-brain structural connectomes derived from multi-modal data within a prediction framework can capture novel information about perinatal brain development. We used morphometric similarity networks (MSNs) to model inter-regional correlations of multiple macro- and micro-structural multi-contrast MRI variables in a single individual. This approach was originally devised to study how human cortical networks underpin individual differences in psychological functions (Seidlitz et al., 2018), and we adapted it to describe both cortical and subcortical regions in the developing brain. The method works by computing for each region of interest (ROI) a number of metrics derived from different MRI sequences which are arranged in a vector. The aim is to obtain a multidimensional description of the structural properties of the ROIs. The MSN is then built considering the ROIs as nodes and modelling connection strength as the correlation between pairs of ROI vectors, thus integrating in a single connectome

the ensemble of imaging features. The pattern of inter-regional correlations can be conceptualised as a "fingerprint" of an individual's brain.

We investigated the utility of MSNs for describing brain maturation, and for patient classification. The edges of individual MSNs were used to train two predictive models: a regression model to predict postmenstrual age (PMA) at scan and identify the set of image features that best model chronological brain age; and a classification model to discriminate between preterm infants at term equivalent age and term neonates, and thereby identify the networks that explain neuroanatomic variation associated with preterm birth. We hypothesized that predictive models based on MSNs, which integrate information from multiple data modalities, would outperform models based on single metrics and single data modalities.

## 2. Material and methods

### *2.1. Participants and data acquisition*

Participants were recruited as part of a longitudinal study designed to investigate the effects of preterm birth on brain structure and long term outcome. The study was conducted according to the principles of the Declaration of Helsinki, and ethical approval was obtained from the UK National Research Ethics Service. Parents provided written informed consent. One hundred and twelve neonates underwent MRI at term equivalent age at the Edinburgh Imaging Facility: Royal Infirmary of Edinburgh, University of Edinburgh, UK, and 105 had multi-modal imaging suitable for MSN analysis (7 acquisitions did not yield usable datasets across all modalities due to motion or wakefulness during one or more sequences). The study group contained 46 term and 59 preterm infants (details are provided in Table 1). The distribution of PMA at scan for all participants, for the term and preterm groups, and the distribution by gender are shown in Fig. 1. Of the preterm infants, 12 had bronchopulmonary dysplasia, 3 had necrotising enterocolitis and 3 required treatment for retinopathy of prematurity.

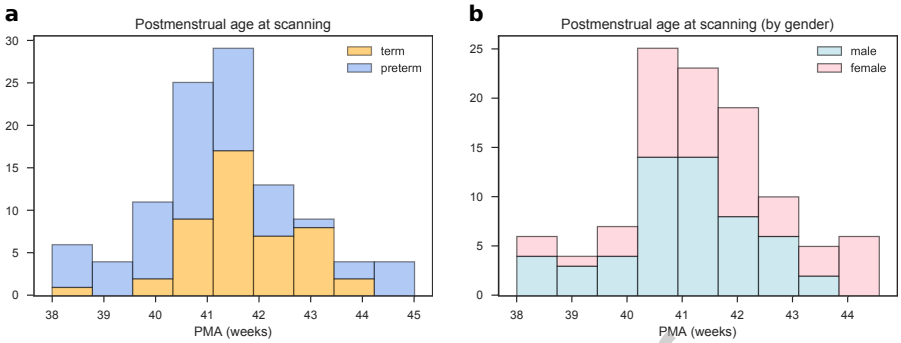



Figure 1: Distribution of postmenstrual age at scan for all subjects. a) Age distribution for the for term (blue) and preterm (orange) groups. b) Age distribution for male (blue) and female (pink) participants.

Table 1: Participant characteristics. The last column reports the $p$ values of the group differences computed with the Wilcoxon rank-sum test for continuous variables and with the chi-squared test for categorical variables.

| | preterm (N=59) | term (N=46) | all (N=105) | preterm vs. term |
|---|---|---|---|---|
| PMA at birth (weeks) | 23.42-32.00 | 37.00-42.00 | 23.42-42.00 | $p = 1.88 \times 10^{-18}$ |
| Birth weight (grams) | 454-2100 | 2556-4560 | 454-4560 | $p = 1.93 \times 10^{-18}$ |
| PMA at scan (weeks) | 38.00-44.56 | 38.28-43.84 | 38.00-44.56 | $p = 0.0035$ |
| M:F ratio | 29:30 | 26:20 | 55:50 | $p = 0.4532$ |

PMA = Postmenstrual age, M = male, F = female.

A Siemens MAGNETOM Prisma 3 T MRI clinical scanner (Siemens Healthcare Erlangen, Germany) and 16-channel phased-array paediatric head coil were used to acquire: 3D T1-weighted MPRAGE (T1w) (acquired voxel size = 1mm isotropic) with TI 1100 ms, TE 4.69 ms and TR 1970 ms; 3D T2-weighted SPACE (T2w) (voxel size = 1mm isotropic) with TE 409 ms and TR 3200 ms; and axial dMRI. dMRI was acquired in two separate acquisitions to reduce the time needed to re-acquire any data lost to motion artefact: the first acquisition consisted of 8 baseline volumes (b = 0 s/mm$^2$ [b0]) and 64 volumes with b = 750 s/mm$^2$, the second consisted of 8 b0, 3 volumes with b = 200 s/mm$^2$, 6 volumes with b = 500 s/mm$^2$ and 64 volumes with b = 2500 s/mm$^2$; an optimal angular coverage for the sampling scheme was applied (Caruyer et al.,

2013). In addition, an acquisition of 3 b0 volumes with an inverse phase encoding direction was performed. All dMRI images were acquired using single-shot spin-echo echo planar imaging (EPI) with 2-fold simultaneous multislice and 2-fold in-plane parallel imaging acceleration and 2 mm isotropic voxels; all three diffusion acquisitions had the same parameters (TR/TE 3400/78.0 ms).

Infants were fed and wrapped and allowed to sleep naturally in the scanner. Feeds were timed to increase the likelihood of post-prandial sleep, flexible earplugs and neonatal earmuffs (MiniMuffs, Natus) were used for acoustic protection, and a soothing environment was created in terms of light and noise. Pulse oximetry, electrocardiography and temperature were monitored. All scans were supervised by a doctor or nurse trained in neonatal resuscitation. Each acquisition was inspected contemporaneously for motion artefact and repeated if there had been movement but the baby was still sleeping; dMRI acquisitions were repeated if signal loss was seen in 3 or more volumes. The majority of the cohort had one or more sequences repeated in order to acquire the best possible quality data for processing.

Conventional images were reported by an experienced paediatric radiologist (A.J.Q.) using a structured system (Leuchter et al., 2014; Woodward et al., 2006), and none of the images included in the final sample ($N = 105$) showed evidence of focal parenchymal injury (defined as post-haemorrhagic ventricular dilatation, porencephalic cyst or cystic periventricular leukomalacia), or central nervous system malformation.

### *2.2. Data preprocessing*

All the following preprocessing steps, including maps calculation and quality check, were performed using dcm2niix, FSL, MRtrix, MIRTK, ANTs, Connectome Workbench and cuDIMOT (Smith et al., 2004; Avants et al., 2011; Marcus et al., 2011; Makropoulos et al., 2014; Li et al., 2016; Hernandez-Fernandez et al., 2019; Tournier et al., 2019).

First, all DICOM image files (dMRI and sMRI) were converted to NIFTI (Li et al., 2016). Structural data were preprocessed using the developing Human

Connectome Project (dHCP) minimal structural processing pipeline for neonatal data (Makropoulos et al., 2018). Briefly, the T1w image was co-registered to the T2w image, both were corrected for bias field inhomogeinities (Tustison et al., 2010) and an initial brain mask was created (Smith, 2002). Following this, the brain was segmented into different tissue types (CSF: cerebrospinal fluid; WM: white matter; cGM: cortical grey matter; GM: subcortical grey matter) using the Draw-EM algorithm (Makropoulos et al., 2014). Twenty manually labelled atlases (Gousias et al., 2012) were then registered to each subject using a multi-channel registration approach, where the different channels of the registration were the original intensity T2-weighted images and GM probability maps. These GM probability maps were derived from an initial tissue segmentation, performed using tissue priors propagated through registration of a preterm probabilistic tissue atlas (Serag et al., 2012). The framework produces several output files, but for this study only the aligned T1w and the T2w images and the parcellation in 87 ROIs were used (Makropoulos et al., 2018). Note that from these 87 ROIs six were removed: the background, the unlabelled brain area (mainly internal capsule), the CSF, the lateral ventricles (left and right) and the corpus callosum (see section 2.4).

Diffusion MRI processing was performed as follows: for each subject the two dMRI acquisitions were first concatenated and then denoised using a Marchenko-Pastur-PCA-based algorithm (Veraart et al., 2016,b); the eddy current, head movement and EPI geometric distortions were corrected using outlier replacement and slice-to-volume registration with TOPUP and EDDY (Andersson et al., 2003; Smith et al., 2004; Andersson and Sotiropoulos, 2016; Andersson et al., 2016, 2017); bias field inhomogeneity correction was performed by calculating the bias field of the mean b0 volume and applying the correction to all the volumes (Tustison et al., 2010). This framework only differs from the optimal pipeline for diffusion preprocessing presented in Maximov et al. (2019) in that we did not perform the final smoothing or the gibbs-ring removal (Kellner et al., 2016) due to the nature of the data (partial fourier space acquisition).

The mean b0 EPI volume of each subject was co-registered to their structural

T2w volume using boundary-based registration (Greve and Fischl, 2009), then the inverse transformation was used to propagate ROI labels to dMRI space, with a modified bbrslope parameter of 0.5, which is used for neonatal data (Toulmin et al., 2015).

For each ROI, two metrics were computed in structural space: ROI volume and the mean T1w/T2w signal ratio (Glasser and Van Essen, 2011). The other ten metrics were calculated in native diffusion space: five metrics derived from the diffusion kurtosis (DK) model (Jensen et al., 2005) and five derived from the Neurite Orientation Dispersion and Density Imaging model (NODDI) (Zhang et al., 2012; Tariq et al., 2016).

### *2.3. Feature extraction*

#### *2.3.1. Structural metrics*

ROI volumes were calculated without normalising for the whole brain volume, as they are used only to compute inter-regional similarities within subjects. The mean T1w/T2w signal ratio was calculated before the bias field correction. The T1w/T2w ratio was used because it enhances myelin contrast and mathematically cancels the signal intensity bias related to the sensitivity profile of radio frequency receiver coils (Glasser and Van Essen, 2011).

#### *2.3.2. Diffusion kurtosis metrics*

The diffusion kurtosis (DK) model is an expansion of the diffusion tensor model. In addition to the diffusion tensor, the DK model quantifies the degree to which water diffusion in biological tissues is non-Gaussian using the kurtosis tensor. The reason for this is that the Gaussian displacement assumption underlying the diffusion tensor breaks at high b-values (Jensen et al., 2005). On the kurtosis component, we only focus on the mean value along all diffusion directions.

The metrics obtained from the DK model for each ROI are the means of: the fractional anisotropy (FA), mean, axial and radial diffusivity (MD, RD, AD) and kurtosis (MK). The MK map quantifies the deviation from Gaussianity of water

molecule displacement and can reflect different degrees of tissue heterogeneity (Steven et al., 2014).

### *2.3.3. NODDI metrics*

We included NODDI metrics alongside the more commonly adopted diffusion tensor measures as previous studies have shown that NODDI indices are sensitive to underlying biological changes in the brain and provide more specific microstructural characteristics, in agreement with histology (Grussu et al., 2017; Batalle et al., 2018).

For the NODDI measures, the Bingham distribution was employed (Tariq et al., 2016) as it allows extra flexibility by describing fibre dispersion along two orthogonal axes. From this NODDI implementation we obtain five metrics: intracellular volume fraction ($\upsilon_{\mathrm{ic}}$), isotropic volume fraction ($\upsilon_{\mathrm{iso}}$), the orientation dispersion index along the primary and secondary directions ($\mathrm{ODI_P}$ and $\mathrm{ODI_S}$) and the overall orientation dispersion index ($\mathrm{ODI_{TOT}}$).

One limitation of this model is that it requires fixing a value for the diffusivity along the axons. However, optimal values for this parameter are region-dependent (Karmacharya et al., 2018) and the default value may be suboptimal for the neonatal population as it has been optimised using an adult cohort (Zhang et al., 2012; Karmacharya et al., 2018). Several studies have been reporting NODDI values for neonates using default (or unspecified) parameters (Batalle et al., 2018; Bastiani et al., 2018; Karmacharya et al., 2018) or modified ones (Kunz et al., 2014; Jelescu et al., 2015). As our goal was not to report NODDI values for the different areas, and because of the lack of reference values for this population, we calculated NODDI maps using default parameters (Batalle et al., 2018).

## *2.4. Data Quality Control*

The parcellations obtained after the processing were visually inspected and parcels corresponding to CSF and background parcels were excluded because they do not represent brain tissue. We observed a poor segmentation of the

corpus callosum in part of the subjects, but we did not find any anomalies in the rest of the parcels. This effect could be caused by different factors: a) this area is problematic to segment due to the proximity to CSF and its small thickness (see for example Otsuka et al. (2019)); b) the framework we used was optimised for the dHCP data that have a very high resolution (0.5 $mm^3$ isotropic) and data quality, making the partial volume effect more noticeable in data with a resolution of 1 $mm^3$; c) or susceptibility artifacts. Instead of removing the subjects with a poor segmentation, we decided to remove the corpus callosum from the model, aiming at maximising the number of subjects. As a result of the whole quality check, we include the whole population ($N = 105$) and each network is composed of 81 nodes (ROIs).

For the dMRI data we use eddy QC (Bastiani et al., 2019). The quality control is performed at subject level and group level. Eddy QC provides several measures related to the rotation, translation and outliers of the images. In addition, it also computes the signal-to-noise (SNR) ratio maps of the b0 volumes and the contrast-to-noise (CNR) ratio maps for the different b-values. These maps can be used at group level to visualise the quality of the data (Bastiani et al., 2018). The results show that the overall quality of the data-set was good (Fig. 2). For eddy QC to work, we removed the b-value = 200 $s/mm^2$ only from the quality control. This is because the low number of volumes with this b-value sometimes leads the Gaussian process performed by eddy to produce a perfect fit, which makes the CNR maps unrealistic.

Fig. 2 shows two representative subjects, one from the top quartile of the SNR and CNR distributions (green star) and one from the bottom quartile (red star). In the first panel we can see where they are placed in terms of SNR and CNR over the overall population. The second panel shows the SNR maps (for the b0) and the CNR maps (for the rest of b-values). The bottom panel of the Fig. 2 shows the b0 before and after the processing of the selected subjects. It is possible to observe the effect of the different steps involved, such as the EPI geometric corrections or the bias field inhomogeneity correction. Supplementary Figs. S8 and S9 report the above results for the term and preterm population

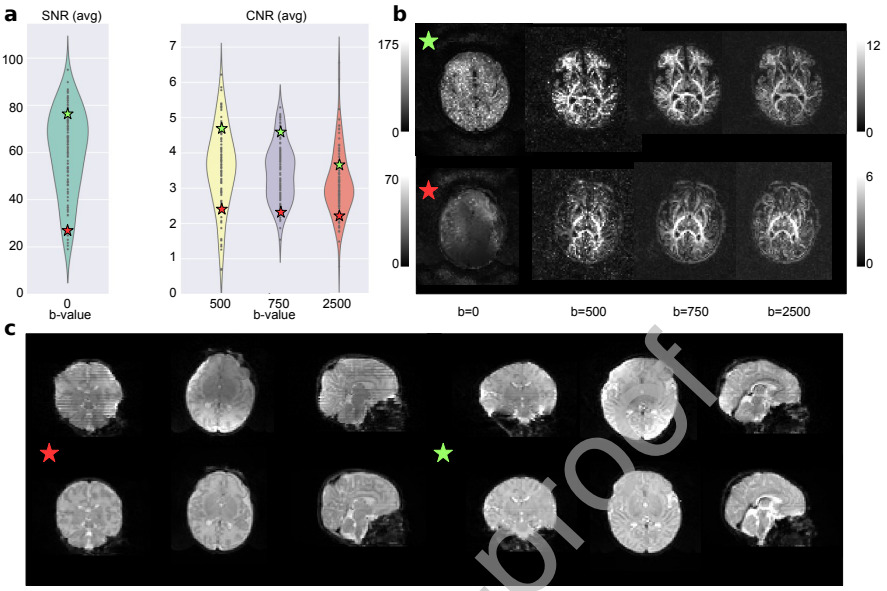



Figure 2: Quality control results. a) Results for the overall population with two selected subjects, one from the top quartile of the SNR and CNR distributions (green star) and the other from the bottom quartile (red star). b) The SNR and CNR maps for the selected subjects. c) The b0 of both subjects before and after the processing pipeline.

respectively.

Following Bastiani et al. (2019), for each volume, motion is quantified by averaging voxel displacement across all voxels (computed as 3 translations and 3 rotations around the x, y and z axes). Absolute displacement is computed with respect to the reference volume, while relative displacement is computed with respect to the previous volume. A summary measure for each subject is calculated as the average (absolute or relative) displacement across all volumes. In Supplementary Fig. S10 we show the distribution of absolute and relative motion for the term and the preterm groups. We compared the distributions with a Wilcoxon rank-sum test and found no difference between the relative motion scores ($W = 1330$, $p = 0.43$) and a significant difference between the absolute motion scores ($W = 1720$, $p = 0.02$). However, as the violin plot shows, this difference is driven by the presence of outliers.

### 2.5. Experimental design and statistical analysis

The models and the analyses described in this section were implemented in Python (v3.6.4) using open source libraries and frameworks for scientific computing, including SciPy (v1.0.0), Numpy (v1.14.0), Statsmodels (v0.8.0), Pandas (v0.22.0), Scikit-learn (v0.19.1) and Matplotlib (v2.1.2) (Jones et al., 2001; Hunter, 2007; Seabold and Perktold, 2010; McKinney et al., 2010; Pedregosa et al., 2011; Van Der Walt et al., 2011).

#### 2.5.1. Network Construction

The MSN for each subject was constructed starting from 81 ROIs; each of the ROI metrics was normalised (z-scored) and Pearson correlations were computed between the vectors of metrics from each pair of ROIs. In this way, the nodes of each network are the ROIs and the edges represent the morphometric similarity between the two related ROIs (Fig. 3). In the following, the terms "edge", "connection" and "inter-regional similarity" are used interchangeably to refer to the correlation between the regional metrics of a pair of ROIs.

#### 2.5.2. Confounding variables

Early exposure to the extrauterine environment due to preterm birth exposes infants to several processes that are known to impact brain maturation (e.g. specific co-morbidities such as bronchopulmonary dysplasia and necrotising enterocolitis (Barnett et al., 2018)), and other processes and diseases that can modify brain maturation (for example gestational age at birth, chorioamnionitis, fetal growth restriction, nutritional insufficiency, pain and medication exposures (Duerden et al., 2016; Anblagan et al., 2016; Barnett et al., 2018; Schneider et al., 2018; Duerden et al., 2018; Blesa et al., 2019)). In addition, there may be as yet unknown environmental risks to brain structural connectivity and genomic and epigenomic factors may interact with gestational age at birth to confer risk (Batalle et al., 2017, 2018b; Boardman et al., 2014; Sparrow et al., 2016; Krishnan et al., 2017). Therefore, it is not possible to define a preterm infant cohort without any exposures to processes that could influence

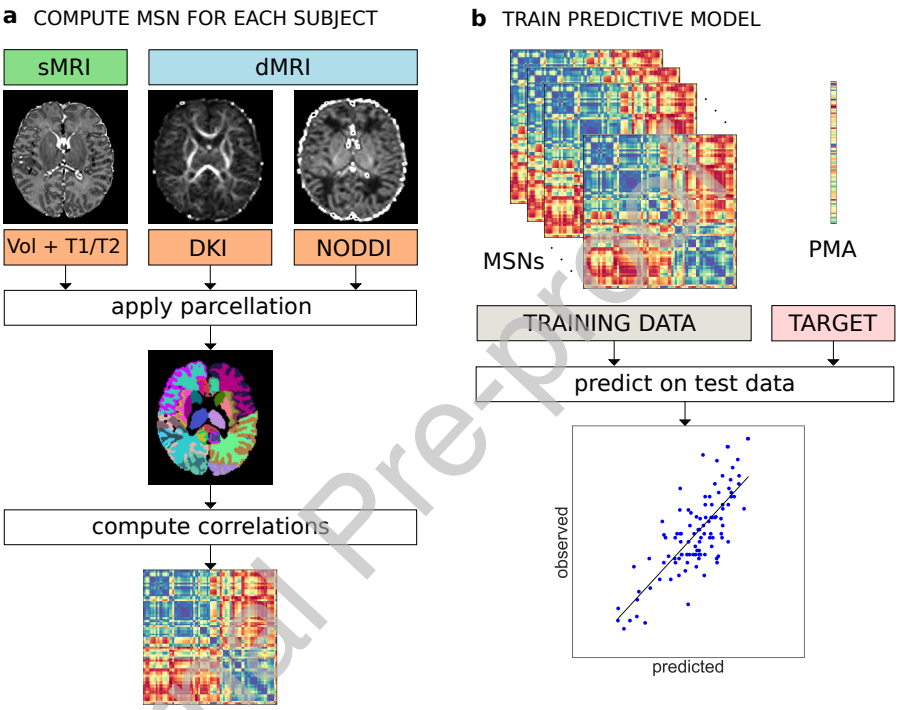



Figure 3: a) Individual MSN construction. Different metrics are extracted from dMRI and sMRI data. The same parcellation is applied to all image types and the average metric values are computed for each ROI. A MSN (represented here as a connectivity matrix) is built by computing the Pearson correlation between the vectors of metrics of each pair of ROIs. b) Training of a predictive model (here for PMA at scan) from individual MSNs. The inter-regional correlations are used as predictor variables in a machine learning model. The performance of the model is evaluated on an independent test set.

brain maturation. As our intention was to develop an integrated approach for characterising dysmaturation in a study group representative of the target population, rather than to investigate possible drivers of dysmaturation, we did not control for any of the above factors.

We did however find that the preterm group was characterised by higher in-scanner motion than the term-group, hence we considered absolute displacement as a confounder (section 2.4). We also observed a positive correlation ($\rho = 0.27, p = 0.0048$) between PMA at scan and PMA at birth and a negative correlation ($\rho = -0.22, p = 0.0233$) between PMA at scan and gender (coded as a binary variable where 0 indicates female infants and 1 male infants), implying that in our sample term subjects and female subjects tend to have their scan acquired at a later age (see also fig. 1). To control for potential bias, we used these confounders as predictors and compared their predictive performance with our network-based features. We tested the interaction between gender and prematurity in a linear regression model of PMA at scan, but the interaction term was not significant ($p = 0.9634$). Birthweight was not included explicitly as a confounder due to its collinearity with PMA at birth.

*2.5.3. Regression model for age*

We trained a linear regression model with elastic net regularisation to predict PMA at scan – i.e. chronological brain age – in both preterm and term infants starting from individual MSNs. This model was chosen for its ability to cope with a high number of features (Zou and Hastie, 2005). For each subject, the edges of the MSN (inter-regional correlations) were concatenated to form a feature vector to be given as input to the regression model. Since the connectivity matrix representing the MSN is symmetric, we considered only the upper triangular matrix for each subject. Gender and age at birth were included in the model to control for their possible confounding effects. The prediction performances were evaluated with a leave-one-out cross-validation (LOOCV) scheme, by computing the mean absolute error (MAE) averaged across subjects. Within each fold of the LOOCV, the parameters of the elastic net were selected with

a nested 3-fold cross-validation loop; the folds were stratified in percentiles to include samples covering the whole age range in each of the folds. Permutation testing was used for the statistical validation of the model performance: the null distribution was built by running the age prediction analysis on 1000 random permutation of the PMA.

#### *2.5.4. Classification model*

A Support Vector Machine (SVM) classifier with linear kernel was trained to discriminate between preterm and term infants. As per the regression model, the input for each subject consisted of inter-regional connections taken from the upper triangular connectivity matrix and the performances were evaluated with LOOCV. Age at the time of scanning, gender and motion were included as additional covariates.While in the case of regression the elastic net regularisation performs automatically a variable selection step, recursive feature elimination (RFE) was applied in combination with SVM to select the best subset of connections. Model selection was implemented using nested cross validation: an outer 3-fold cross-validation loop was used to select the SVM parameters and an inner 4-fold cross-validation loop was used for RFE. Folds were stratified to include the same proportion of term and preterm subjects. The accuracy of the model was computed as the number of correctly classified subjects across the leave-one-out folds over the total number of subjects in the test set. The null distribution was built by repeating the exact same analysis 1000 times after randomly assigning subjects to the term and the preterm group.

#### *2.5.5. Feature selection*

After the preprocessing phase, twelve different metrics were available for each ROI. To study which combination of features produced better performance in the prediction tasks, we implemented a sequential backward-forward feature selection scheme. Starting from the full set of features, at each iteration we compare the performances of different models built by removing in turn each of the features from the current set of candidate features. We then exclude from

the next iteration the feature whose subtraction caused the least increase in prediction error (down to three features, for a total of 73 combinations). The rationale behind this scheme is to explore the space of possible models without enumerating all possible solutions, thus reducing the computational demands compared to an exhaustive search. The procedure was performed separately for the regression and the classification models.

#### *2.5.6. Cross-validation strategy*

We adopted LOOCV to select the best performing model in both the age prediction and the classification tasks as this scheme enabled maximum size of the training set and therefore best use of available data, but this strategy might induce high variance in the estimation of prediction accuracy (Kohavi, 1995; Efron, 1983). In the context of brain decoding (i.e. predictions from brain images or signals), LOOCV was shown to produce overly optimistic estimates of prediction accuracy in the within-subject setting (i.e. when all samples are highly correlated because they come from the same subject). In the between-subject setting (as in this work), the performance of LOOCV is similar to schemes involving random splits and mostly determined by sample size (Varoquaux et al., 2017; Varoquaux, 2018). To assess the stability of our results with respect to the chosen cross-validation scheme, we report the prediction accuracy computed with a 10 repeated stratified 5-fold scheme (10-5-fold) for all the models selected with LOOCV.

#### *2.5.7. Comparison with individual metrics and single data modalities models*

We compared the performances of the best performing models based on MSNs with three classes of baseline models: a) models based on single global brain metrics (total brain volume and median FA in the WM); b) models based on individual metrics, where instead of similarities, predictors are the concatenation of all regional values for each of the individual metrics used to build MSNs; c) single data modality MSNs, i.e. models built on structural features only (Volume and T1/T2), on DKI features only, and on NODDI features only.

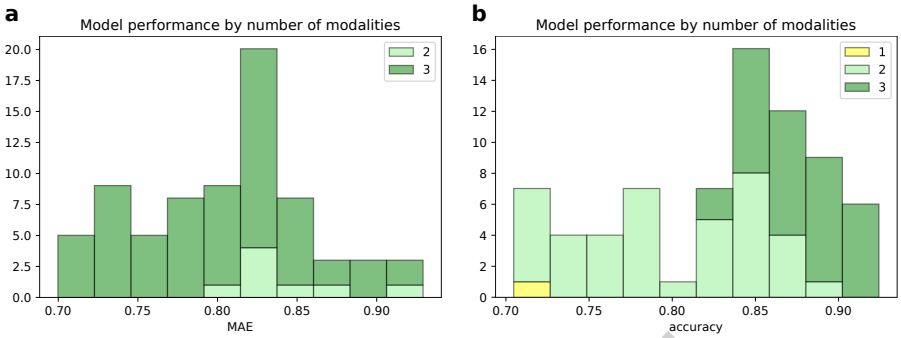



Figure 4: Histograms of the performance of the 73 models compared in the backward feature selection scheme for the age prediction task (a) and for the classification task (b). Bars are grouped by the number of modalities included in the models.

### *2.6. Data and code availability*

Source code implementing the methods described in this paper is available upon request to the corresponding author. The preprocessed and anonymised data used in the analyses can be requested through the Brains Image Bank (`https://www.brainsimagebank.ac.uk/`) (Job et al., 2017).

## 3. Results

### *3.1. Feature selection*

In Fig. 4 we report two histograms summarising the LOOCV performance of the 73 different models compared per each task in the backward feature selection scheme. In both cases, we can observe that the models based on all three data modalities achieved better results in terms of prediction accuracy. The performances of each of the compared model are reported in Supplementary Figs. S1 and S3 for the age prediction and for the classification models, respectively.

The best performing model for age prediction, which was adopted for all subsequent analyses, was based on seven features (Volume, FA, MD, AD, MK, $v_{\mathrm{iso}}$, $\mathrm{ODI_P}$). Fig. 5a shows the average MSN matrix computed across all subjects for the selected set of features and the matrix of correlation between inter-regional similarities and PMA at scan across subjects. The average MSN matrix shows

four main blocks that correspond roughly to positive correlations between ROIs within GM and between ROIs within WM, and to negative correlation between WM ROIs and GM ROIs, indicating that ROIs within GM (and within WM) share similar structural properties, while GM and WM regional descriptors tend to be anti-correlated. The four-block structure is recognisable also in the matrix reporting correlations with chronological age: with increasing age regions within GM or within WM become more similar with each other, while the dissimilarities between GM and WM ROIs increases.

The best classifier model was based on eleven out of the twelve features (all except $\mathrm{ODI_S}$), so compared to the age prediction model, four additional features were included: T1/T2, RD, $v_{\mathrm{ic}}$ and $\mathrm{ODI_{TOT}}$. The average MSN computed with the selected features and the matrix of correlation with PMA at birth is shown in Fig. 5 (panels b and c). Comparing panel b and d of Fig. 5, it is apparent that while the patterns of correlation with PMA at scan and at birth are similar within GM and WM, subcortical ROIs show an opposite trend: with increasing PMA at scan subcortical ROIs tend to become more similar to WM ROIs and more dissimilar to GM ROIs, but the similarity between subcortical ROIs and cortical GM is positively correlated to age at birth.

### *3.2. Prediction results*

The best regression model selected with LOOCV predicted chronological age (PMA at scan) with a MAE of $0.70 \pm 0.56$ weeks on the test data, and a correlation between the predicted and the actual age equal to $r = 0.78$ ($p = 1.71 \times 10^{-22}$) (Supplementary Fig. S5). The results of the permutation test are shown in Fig. 6 and Supplementary Fig. S6. The confounding variables (gender and age at birth) were not selected by the internal feature selection procedure, hence the predictions were based on network features alone. To test whether there was any systematic difference in the predicted age between the term and the preterm group, we compared the error distributions with a Wilcoxon rank-sum test, but the result was not significant ($W = 1108,\ p = 0.1085$). For comparison, we evaluated the predictive performance of a linear regression model using

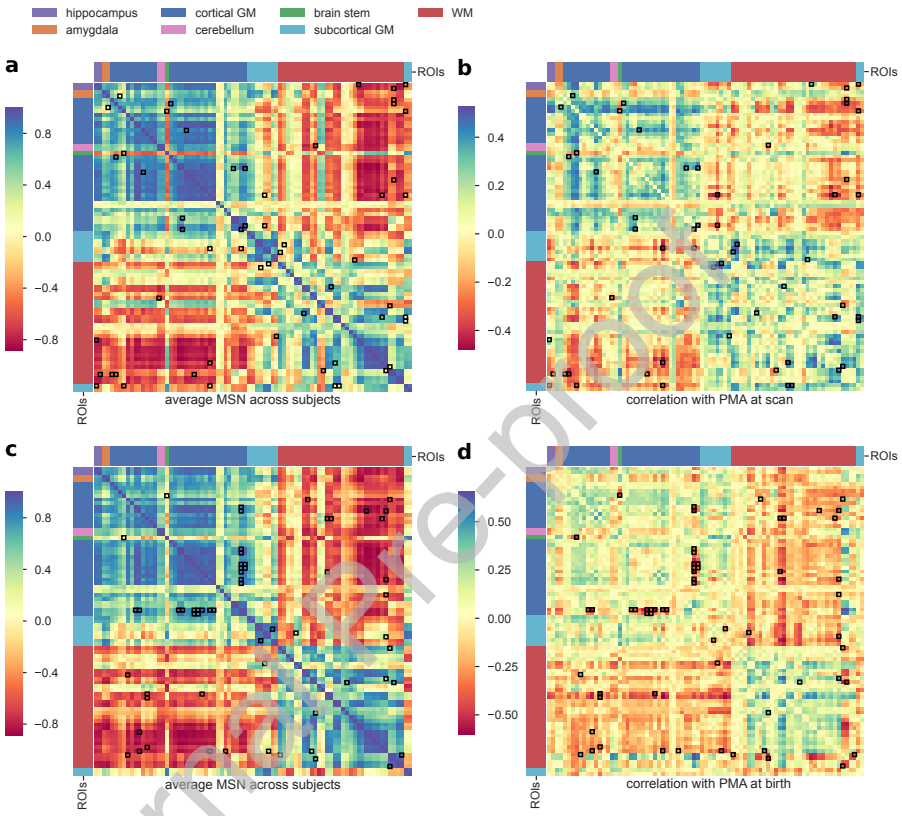



Figure 5: a) Average MSN computed across all subjects using the combination of features selected through the backward feature selection scheme for the age prediction task (Volume, FA, MD, AD, MK, $v_{\mathrm{iso}}$, $\mathrm{ODI_P}$). b) Correlation between each connection weight (inter-regional similarity) shown in (a) and PMA at scan across subjects. c) Average MSN computed across all subjects using the combination of features selected through the backward feature selection scheme for the classification task (Volume, T1/T2, FA, MD, AD, RD, MK, $v_{\mathrm{ic}}$, $v_{\mathrm{iso}}$, $\mathrm{ODI_P}$, $\mathrm{ODI_{TOT}}$). d) Correlation between each connection weight (inter-regional similarity) shown in (c) and PMA at birth across subjects. Connections that were identified as predictive features by the models are highlighted in black. ROIs are ordered as in Supplementary Table S1.

only gender and PMA at birth as independent variables, that achieved a MAE of $1.03 \pm 0.88$ weeks. A Wilcoxon signed-rank test confirmed that the latter model achieved a significantly greater error ($W = 1633,\ p = 0.0001$). Also models based on single global metrics and single-modality MSNs models provided poorer predictive performance than the selected multi-modality MSNs model (brain volume: MAE= $0.93 \pm 0.68$, $R = 0.58$; median FA: MAE= $0.88 \pm 0.63$, $R = 0.58$; structural: MAE= $1.08 \pm 0.79$, $R = 0.32$; DKI: MAE= $0.94 \pm 0.70$, $R = 0.57$; NODDI: MAE= $0.88 \pm 0.69$, $R = 0.61$) and this was confirmed by a Wilcoxon signed-rank test (brain volume: $W = 1813,\ p = 0.0019$; median FA: $W = 2045,\ p = 0.0184$; structural: $W = 1361,\ p = 2.76 \times 10^{-06}$; DKI: $W = 1734,\ p = 0.0004$; NODDI: $W = 1811,\ p = 0.0009$). Conversely, the baseline model based on the ensemble on individual metrics used to build the best performing MSN model achieved similar performances (MAE: $0.72 \pm 0.56$, $R = 0.77$). A scatter plot of the residuals of the two models (Supplementary Fig. S11) showed a linear trend, indicating that the two models share a similar information content.

Supplementary Fig. S2 shows the results computed with 10-5-fold cross-validation in. All compared models performed similarly under the 10-5-fold scheme, and in general worse than with the LOOCV scheme, with the selected model achieving a MAE of $1 \pm 0.2$ weeks (Supplementary Fig. S7).

To study which connections contributed the most to chronological age prediction, we selected only edges which were assigned a non-zero coefficient in at least 99% of cross-validation folds. These edges are shown in the chord diagram in Fig. 7 (realised with Circos, Krzywinski et al. (2009)), and are colour coded to distinguish between inter-regional similarities that increase or decrease with age, to highlight networks of regions whose morphological properties are converging (gray) or that tend to differentiate with increasing age (red). Intuitively, these edges connect ROIs whose anatomical and micro-structural properties are changing more than others between 38 and 45 weeks PMA, making the ROIs more or less similar. In other words, it is the relative timing of maturation of different brain tissues to determine the relevance of a connection in the age

prediction task. The selected connections are located in both cortical (frontal, temporal, parietal and occipital lobes; insular and posterior cingulate cortex) and subcortical regions (thalamus, subthalamic and lentiform nuclei), in the brain stem and in the cerebellum. These areas have been previously associated with age-related changes and preterm birth (Boardman et al., 2006; Ball et al., 2013a; Batalle et al., 2017). For comparison, we report in Supplementary Table S2 the regional metrics selected as most predictive of age in the baseline model based on individual metrics.

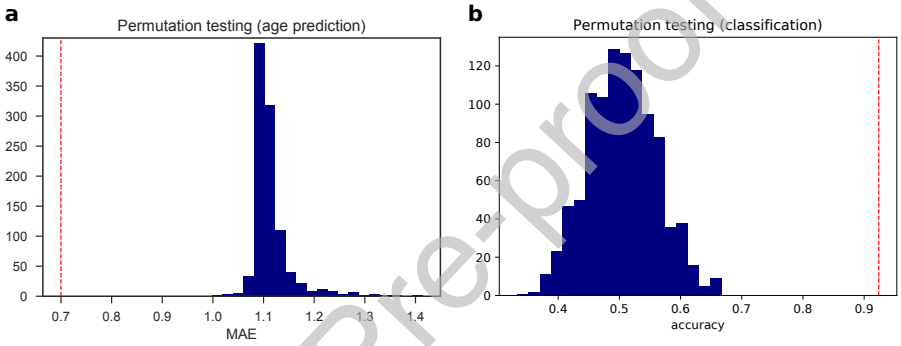



Figure 6: Null distributions computed over 1000 random permutations of the target variable for the age prediction (a) and the classification tasks (b). The red dotted lines indicate the performances of our models.

The best classifier discriminated between term and preterm infants with a 92% LOOCV accuracy (Fig. 6). None of the confounders were included among the selected features. A logistic regression model built on age at scan and gender did not achieve significant accuracy ($56\%, p = 0.091$), while adding motion to the predictors produced a 61% accuracy, slightly above chance level ($p = 0.03$), but it should be noted that a model based on motion only was 59% accurate ($p = 0.02$). Models based on global features achieved 55% accuracy for total brain volume and 56% accuracy for median FA. Models built on single data modalities attained 65% accuracy for structural features only, 89% accuracy for DKI features only, and 88% accuracy for NODDI features only. Results computed with 10-5-fold cross-validation are shown in Supplementary Fig. S4.

The best classifier selected with LOOCV also achieved top accuracy with 10-5-fold (accuracy 90%, Supplementary Fig. S7).

The network of regions that showed the most divergent pattern of structural brain properties in preterm versus term infants comprised the brain stem, the thalamus and the subthalamic nucleus; WM regions in the frontal and insular lobes; GM regions in the occipital lobe; both WM and GM regions in the temporal and parietal lobes and in the posterior cingulate cortex. The chord diagram of edges selected by 99% of the models is shown in Fig. 8, in red where inter-regional similarities are greater in the term group and in gray where they are greater in the preterm group. For comparison, Supplementary Table S3 lists the regional metrics selected by the baseline model based on individual metrics, that obtained a 94% accuracy.

### *3.3. Testing for asymmetry*

In both chord diagrams (Figs. 7 and 8) we observed more edges in the right hemisphere than in the left one. Both elastic net and SVM models perform a feature selection step to exclude features that are correlated and that carry redundant information in order to improve prediction performance, hence it might be the case that the models selected the right connections and discarded the left ones precisely because they had a similar information content. Additionally, in the leave-one-out cross-validation scheme the training sets only differ by two samples in each fold, hence models might be similar across folds.

To test the hypothesis that the two hemispheres carry a different information content, we performed two experiments. First, we repeated the same analyses extracting inter-regional similarities from either the right or the left hemisphere. We compared the performance obtained with the regression and classification models on the different subsets of features used in the backward feature selection scheme in the main analyses. We found that for the age prediction model a Wilcoxon signed-rank test testing the hypothesis that the prediction error was higher using only connections from the left hemisphere was significant ($W = 156, p = 2.57 \times 10^{-11}$), while there was no statistically significant differ-

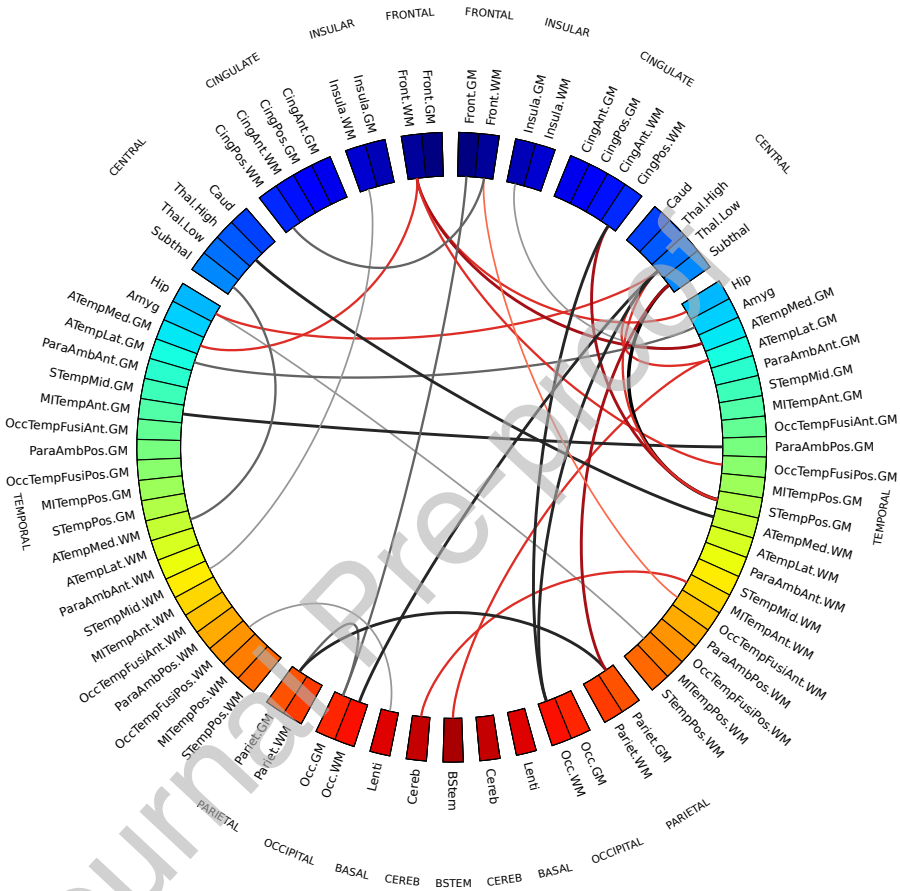



Figure 7: Chord diagram showing MSN edges used for age prediction in at least 99% of regression models in the cross-validation folds. Connections shown in gray are inter-regional similarities that increase with chronological age, while connections in red are inter-regional similarities that decrease with chronological age. The edge width is proportional to the correlation between inter-regional similarities and PMA. The left side of the diagram corresponds to the left side of the brain. Abbreviations for ROI names are explained in Supplementary Table S1.

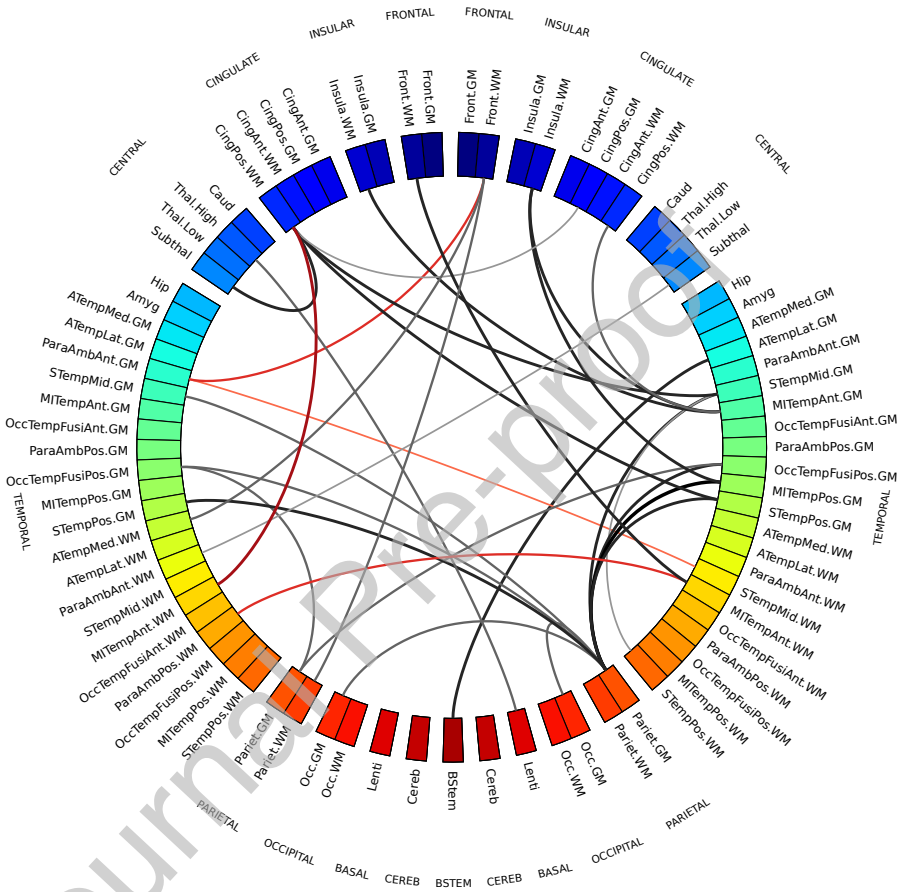



Figure 8: MSN edges showing a divergent pattern of morphological properties in term and preterm infants in at least 99% of classification models in the cross-validation folds. Gray connections indicate inter-regional similarities that are greater in the preterm group, while red connections are greater in the term group. The edge width is proportional to the correlation between inter-regional similarities and prematurity. The left side of the diagram corresponds to the left side of the brain. Abbreviations for ROI names are explained in in Supplementary Table S1.

ence in the case of the classification model. These results replicated also when using 10-5-fold cross-validation (age prediction: $W = 160, p = 2.98 \times 10^{-11}$; no significant difference in classification). We also compared the residuals obtained using either the right or the left hemisphere for age prediction with the set of features selected with backward feature selection (Supplementary Fig. S11) and found that the residuals of the fitted models are linearly correlated, suggesting that the two hemispheres do carry a similar information content, but one presents clearer signal than the other. We then used permutation testing to test the "interchangeability" of right and left regions: starting from the subsets of imaging metrics selected in the main analyses for the age prediction and classification models, we generated two null distributions by randomly swapping a subset of homotopic brain regions between the right and left hemisphere, and then repeating the exact same analyses 1000 times. We then counted how many times in the random models there was a disproportion of inter-regional similarities selected in the right hemisphere equal or greater than the one we observed with our models. If the right and left are "interchangeable", the number of inter-regional similarities selected should remain the same on average. We found that in the age prediction task, under the null distribution, the disproportion of predictive connections in the right hemisphere was associated with a $p = 0.036$, while in the classification task the disproportion was not significant ($p = 0.166$). This implies that at least for age prediction the two hemispheres are not interchangeable, suggesting again that the right hemispheres has a stronger signal. A similar trend was observed under the 10-5-fold cross-validation scheme, but in this case we could not reject the null hypothesis that inter-regional similarities are selected with the same frequency from both hemispheres ($p = 0.098$).

## 4. Discussion

These results show that the information encoded in MSNs is predictive of chronological brain age in the neonatal period and that MSNs provide a novel data-driven method for investigating neuroanatomic variation associated with

preterm birth. MSNs were built by combining features from different imaging sequences that describe complementary aspects of brain structure that have been previously studied in isolation (Makropoulos et al., 2016; Batalle et al., 2017) and the resulting predictive models achieved a high accuracy for age prediction and classification. By comparing the performance of MSNs features with basic demographic information (age at birth and gender) and simple metrics such as total brain volume and median white matter FA, we also showed that integrating imaging data provides relevant additional information to characterise brain age. Although we cannot exclude the possibility that some of the variability shared with age at birth, gender or brain volume is encoded in the imaging variables, the comparative analysis and the permutation testing results showed that the observed variance cannot be completely explained by demographic variables or simpler metrics alone. However, a high accuracy is not the only goal of the proposed method: once we have determined that the model is able to learn a relationship between the MSN features and age or prematurity, we can interrogate it to find out which features, regions and structures are involved in the predictions, thus allowing for further inferences.

We anticipate that the main clinical and research utilities of MSNs will be to investigate divergent maturational patterns in the context of perinatal environmental, genetic and clinical exposures, leading to improved understanding of antecedents to, and consequences of, atypical brain development. For these purposes a prediction tool with average 5 days error is highly precise compared with other methods for assessing brain maturation, which usually rely upon simple linear regression, use single image features, or broad classifications of prematurity (Toews et al., 2012; Brown et al., 2017; Batalle et al., 2018; Deprez et al., 2018; Bouyssi-Kobar et al., 2018; Ouyang et al., 2018).

The regions identified as most predictive have been previously associated with age-related changes and preterm birth (Boardman et al., 2006; Ball et al., 2013a; Batalle et al., 2017; Bouyssi-Kobar et al., 2018). These data suggest that to fully describe morphological variation in the developing brain it may be advantageous to adopt a holistic approach, leveraging the additional information

that can be derived from integrating multi-contrast MRI data. The main motivation for using a network-based approach is to obtain a whole-brain description of a developmental pattern. By using topologically integrated features instead of single metrics it is possible to access an additional layer of information that is not explicitly encoded in the individual metrics, i.e. how the relationships between metrics vary in different parts of the brain. Working with correlations instead of an ensemble of heterogeneous metrics also aids interpretation, as the focus is shifted from the values of single metrics across the brain, each influenced by disparate factors, to similarities between brain regions, which is a more relatable concept. Additionally, the adoption of a network model has proven to be a useful abstraction to capture the modular organisation of the brain: in the original work introducing MSNs to study microscale cortical organization in adults, the authors demonstrated that regions that were similar in MSNs were more likely to belong to the same cytoarchitectonic class, to be axonally connected and to have high levels of co-expressions of genes specialised for neural functions (Seidlitz et al., 2018). Another reason for working with similarities instead of single regional metrics is methodological: computing edge weights as inter-regional similarities enables an integrated representation of several metrics in a single network; to work with the original features directly would mean either working with several networks (thus requiring a further step to integrate them and aggravating the problems related with the "curse of dimensionality") or concatenating all the features in a single predictive model (thus excluding the interactions between metrics from the model).

Table 2: Results from previous works in the age prediction task.

| | **Age span** | **Model** | **Error/Accuracy** |
|---|---|---|---|
| Brown et al. 2017 | 27-45 weeks PMA | FA-weighted structural connectivity | MAE = 1.6 weeks |
| Ouyang et al. 2019 | 31.5-41.7 weeks PMA | cortical FA and MK (mean kurtosis) | FA: r = .92; MK: r = .63 |
| Deprez et al. 2018 | 29-44 weeks PMA | spatio-temporal growth models for myelin-like signals in the thalami and brainstem | Thalami: MAE = 1.41 weeks<br>Brainstem: MAE = 2.56 weeks |
| Toews et al. 2012 | 8-590 days from birth | scale-invariant T1w features | MAE = 72 days |
| Wu et al. 2019 | 14-48 days from birth | cortical measures | MAE = 11.1 ± 0.3 days |

PMA = postmenstrual age, MAE = mean absolute error, r = Pearson's coefficient between actual and predicted age.

Our data are consistent with previous studies of perinatal brain age prediction based on a single type of data or a single metric. For example, Brown et al. (2017) used dMRI tractography to predict brain dysmaturation in preterm infants with brain injury and abnormal developmental outcome and found that altered connectivity in the posterior cingulate gyrus and the inferior orbitofrontal cortex were associated with a delayed maturation; both of these regions are included in the networks identified by our model. Regional FA, MD, MK, and $v_{\mathrm{ic}}$ are each predictive of age (Genc et al., 2017; Karmacharya et al., 2018; Ouyang et al., 2019), and the first three measures were selected in our age predicition model. Growth of the thalami and brainstem, defined in terms of myelin-like signals from T2-weighted images, successfully predicted age between 29 and 44 weeks (Deprez et al., 2018) and these regions are included in the networks most predictive of age in the current study. In Toews et al. (2012), scale-invariant image features were extracted from T1-weighted MRI data of 92 subjects over an age range of 8-590 days to build a developmental model that was used to predict age of new subjects; and Ceschin et al. (2018) proposed a deep learning approach to detect subcortical brain dysmaturation from T2-weighted fast spin echo images in infants with congenital hearth disease. Wu et al. (2019) used cortical features extracted from structural images to predict age of 50 healthy subjects with 251 longitudinal MRI scans from 14 to 797 days; in accordance with our results, the regions reported to be important for age prediction were bilateral medial orbitofrontal, parahippocampal, temporal pole, right superior parietal and posterior cingulate cortex. Although our results are not directly comparable with the above works because of the heterogeneity of employed models, validation techniques and population variation (different age ranges), our prediction error is among the lowest reported (see Table 2 for a summary of previous results), but it should be noted that there is a strong positive correlation between the reported MAEs and the age range of the samples. In addition, many works have identified imaging biomarkers associated with preterm birth, such as brain tissue volume (Alexander et al., 2018; Gui et al., 2019), myelin content (Melbourne et al., 2016), and diffusion tensor metrics (Anjari et al.,

2007; Bouyssi-Kobar et al., 2018).

The connections most predictive of age revealed that brain maturation is characterised by morphological convergence of some networks and divergence of others (Fig. 7). These connections mostly involve fronto-temporal and subcortical ROIs, which suggests that the micro- and macro-structural properties of these regions are highly dynamic between 38-45 weeks. Among these, inter-regional similarities within GM and WM increase with age, similarities between cortical GM and WM decrease, while subcortical ROIs become more similar to WM and more dissimilar to cortical GM. This is consistent with previous findings on the different trends in development of the thalamus and the cortex (Eaton-Rosen et al., 2015). Additionally, in a study of early development of structural networks (Batalle et al., 2017), connections to and from deep grey matter are reported to show the most rapid developmental changes between 25-45 weeks, while intra-frontal, frontal to cingulate, frontal to caudate and inter-hemispheric connections are reported to mature more slowly.

Conversely, the inter-regional similarities selected by the SVM classifier to discriminate between term and preterm (Fig. 5) are more distributed across cortical GM and WM and are for the most part greater in the preterm group. The fact that in the term group these cortical ROIs are less homogeneous in terms of structural properties could be interpreted as a sign that in term infants these regions are at a different stage of maturation where their morphological profile is consolidating along specialised developmental trajectories. It has been previously suggested that the rapid maturation of cortical structures occurring in the perinatal period is vulnerable to the effects of preterm birth (Kostović and Jovanov-Milošević, 2006; Ball et al., 2011, 2013b; Smyser et al., 2016b).

The differences between networks identified for age prediction and for preterm classification indicate that atypical brain development after preterm birth is not solely a problem of delayed maturation, but it is characterised by a specific signature. Indeed, while the age prediction networks capture changes occurring in both the preterm and the term group, the classification networks highlight where there are group-wise differences, and they do not match: in the case of a delayed

maturation we would have observed differences in the same regions undergoing age-related changes. MSN variations associated with preterm birth affected brain stem, thalami, sub-thalamic nuclei, WM regions in the frontal and insular lobes, GM regions in the occipital lobe, and WM and GM regions in the temporal and parietal lobes and in the posterior cingulate cortex. This distribution of structural variation is consistent with previous reports of regional alteration in brain volume and dMRI parameters based on single contrasts (Boardman et al., 2006; Bonifacio et al., 2010; Ball et al., 2013a; Brown et al., 2017; Batalle et al., 2017; Alexander et al., 2018; Thompson et al., 2018b; Bouyssi-Kobar et al., 2018). Furthermore, compared to the age prediction model, the MSNs used for preterm classification are based on four additional metrics: T1/T2, related to myelination; RD, measuring water dispersion; $v_{ic}$ describing neurite density; and $\mathrm{ODI}_{\mathrm{TOT}}$, associated with the fanning of WM tracts. All these metrics contribute to characterise the micro-structural alterations associated with preterm birth (Eaton-Rosen et al., 2015; Melbourne et al., 2016; Batalle et al., 2018; Thompson et al., 2018b; Bouyssi-Kobar et al., 2018).

We observed a disproportion in the distribution of the connections selected by our models, with a preference for the right hemisphere, hinting at the existence of lateralization in the maturational process. An asymmetry in the development of the right hemisphere in neonates was previously reported in Dubois et al. (2010); Yap et al. (2011); Wu et al. (2019), and our experiments (section 3.3) partially supported the hypothesis that the right hemisphere plays a relevant role in the context of age prediction.

### *4.1. Limitations*

This work has some limitations. First, compared with the original work on MSNs (Seidlitz et al., 2018), we did not have a multi-parametric mapping sequence (Weiskopf et al., 2013); however, because the model is extensible, information from other contrasts could be added and evaluated for their effect on prediction. The MSN model could also be applied to study the properties of cortical gray matter (such as thickness, sulcal depth or curvature), that have been

previously reported to be predictive of age in children (Brown et al., 2012) and could contribute significantly in characterising the newborn brain. However, metrics that only apply to selected structures (e.g. the cortex) cannot be used in a whole brain analysis, as to compute inter-regional similarities each region needs to be described by the same set of metrics. This particular study was designed based on prior knowledge that typical development and atypical development associated with preterm birth are characterised by global changes (Ball et al., 2013a; Anderson, 2014; Eaton-Rosen et al., 2015; Melbourne et al., 2014), and MSNs integrating dMRI and sMRI data were chosen to study generalised processes across the whole brain.

Second, we used a motion correction technique that attenuates the impact of head motion on structural connectivity (Andersson and Sotiropoulos, 2016; Baum et al., 2018), and we found that scanner motion was not contributing significantly to prediction accuracy; however we cannot rule out a possible confounding effect of motion on the estimation of regional metrics.

Third, the preterm study population was representative of survivors of modern neonatal intensive care in terms of gestational age range and prevalence of co-morbidities of preterm birth that may influence brain maturation, but it is still possible that the results were influenced by biological variability specific to the cohort. A replication study will be required to determine whether the patterns of dysmaturation we found are generalisable.

Finally, we assessed the performance of our models with both LOOCV and 10-5-fold schemes in order to investigate the stability of our findings with respect to the chosen cross-validation scheme and we observed some variability in the general trends of the results. The disagreement we found might derive from the limited size of the training set in the case of the repeated-5-fold scheme (all models tended to perform worse, suggesting there were not enough samples for learning), and this was indeed the reason why our first choice was the leave-one-out scheme. As it is always the case when working with machine learning, increasing the sample size would increase the power of the models, thereby reducing the margin of error and the risk of overfitting, with the result that

both schemes should converge to similar findings.

### *4.2. Conclusions*

Combining multiple imaging features in a single model enabled a detailed description of the morphological properties of the developing brain that was used inside a predictive framework to identify two networks of regions: the first, predominantly located in subcortical and fronto-temporal areas, that contributed most to age prediction: the second, comprising mostly frontal, parietal, temporal and insular regions, that discriminated between preterm and term born infant brains. Both predictive models performed best when structural, diffusion tensor-derived and NODDI metrics were combined, which demonstrates the importance of integrating different biomarkers to generate a global picture of the developing human brain. The achieved accuracy supports the hypothesis that studying the interaction between regional metrics can shed light on the mechanics of development.

Morphology, structural connectivity and maturation are all influenced by genetics, co-morbidities of preterm birth, and nutrition (Boardman et al., 2014; Anblagan et al., 2016; Sparrow et al., 2016; Krishnan et al., 2016; Ball et al., 2017; Alexander et al., 2018; Blesa et al., 2019). In future work MSNs could offer new understanding of the impact of these factors on integrated measures of brain development, and the relationship between neonatal MSNs and functional outcome could bring novel insights into the neural bases of cognition and behaviour, by identifying networks of regions associated with later development. MSNs could also enable a direct comparison with functional networks extracted from fMRI, to explore how structure and function interplay in the neonatal period, and study how well the two network models together explain individual variability in developmental outcome.

## Conflicts of interest

Authors declare no conflict of interests.

## Acknowledgements

We are grateful to the families who consented to take part in the study. This work was supported by Theirworld (www.theirworld.org) and was undertaken in the MRC Centre for Reproductive Health, which is funded by MRC Centre Grant (MRC G1002033). MJT was supported by NHS Lothian Research and Development Office. Participants were scanned in the University of Edinburgh Imaging Research MRI Facility at the Royal Infirmary of Edinburgh which was established with funding from The Wellcome Trust, Dunhill Medical Trust, Edinburgh and Lothians Research Foundation, Theirworld, The Muir Maxwell Trust and many other sources; we thank the University's imaging research staff for providing the infant scanning.

## References

## References

Alexander, B., Kelly, C.E., Adamson, C., Beare, R., Zannino, D., Chen, J., Murray, A.L., Loh, W.Y., Matthews, L.G., Warfield, S.K., Anderson, P.J., Doyle, L.W., Seal, M.L., Spittle, A.J., Cheong, J.L., Thompson, D.K., 2018. Changes in neonatal regional brain volume associated with preterm birth and perinatal factors. NeuroImage doi:10.1016/j.neuroimage.2018.07.021.

Alexander-Bloch, A., Giedd, J.N., Bullmore, E., 2013. Imaging structural covariance between human brain regions. Nature Reviews Neuroscience 14, 322–336. doi:10.1038/nrn3465.

Anblagan, D., Pataky, R., Evans, M.J., Telford, E.J., Serag, A., Sparrow, S., Piyasena, C., Semple, S.I., Wilkinson, A.G., Bastin, M.E., Boardman, J.P., 2016. Association between preterm brain injury and exposure to chorioamnionitis during fetal life. Scientific Reports 6, 37932. doi:10.1038/srep37932.

Anderson, P.J., 2014. Neuropsychological outcomes of children born very preterm. Seminars in Fetal and Neonatal Medicine 19, 90–96. doi:10.1016/J.SINY.2013.11.012.

Andersson, J.L., Graham, M.S., Drobnjak, I., Zhang, H., Filippini, N., Bastiani, M., 2017. Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. NeuroImage 152, 450–466. doi:10.1016/j.neuroimage.2017.02.085.

Andersson, J.L., Graham, M.S., Zsoldos, E., Sotiropoulos, S.N., 2016. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. NeuroImage 141, 556–572. doi:10.1016/j.neuroimage.2016.06.058.

Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. NeuroImage 20, 870 – 888. doi:https://doi.org/10.1016/S1053-8119(03)00336-7.

Andersson, J.L., Sotiropoulos, S.N., 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. NeuroImage 125, 1063–1078. doi:10.1016/j.neuroimage.2015.10.019.

Anjari, M., Srinivasan, L., Allsop, J.M., Hajnal, J.V., Rutherford, M.A., Edwards, A.D., Counsell, S.J., 2007. Diffusion tensor imaging with tract-based spatial statistics reveals local white matter abnormalities in preterm infants. NeuroImage 35, 1021–1027. doi:10.1016/J.NEUROIMAGE.2007.01.035.

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ants similarity metric performance in brain image registration. NeuroImage 54, 2033 – 2044. doi:https://doi.org/10.1016/j.neuroimage.2010.09.025.

Back, S.A., Miller, S.P., 2014. Brain injury in premature neonates: A primary cerebral dysmaturation disorder? Annals of Neurology 75, 469–486.

Ball, G., Aljabar, P., Arichi, T., Tusor, N., Cox, D., Merchant, N., Nongena, P., Hajnal, J., Edwards, A., Counsell, S., 2016. Machine-learning to characterise

neonatal functional connectivity in the preterm brain. NeuroImage 124, 267–275. doi:10.1016/J.NEUROIMAGE.2015.08.055.

Ball, G., Aljabar, P., Nongena, P., Kennea, N., Gonzalez-Cinca, N., et al., 2017. Multimodal image analysis of clinical influences on preterm brain development. Annals of Neurology 82, 233–246. doi:10.1002/ana.24995.

Ball, G., Boardman, J.P., Aljabar, P., Pandit, A., Arichi, T., Merchant, N., Rueckert, D., Edwards, A.D., Counsell, S.J., 2013a. The influence of preterm birth on the developing thalamocortical connectome. Cortex 49, 1711 – 1721. doi:https://doi.org/10.1016/j.cortex.2012.07.006.

Ball, G., Boardman, J.P., Rueckert, D., Aljabar, P., Arichi, T., Merchant, N., Gousias, I.S., Edwards, A.D., Counsell, S.J., 2011. The effect of preterm birth on thalamic and cortical development. Cerebral Cortex 22, 1016–1024.

Ball, G., Srinivasan, L., Aljabar, P., Counsell, S.J., Durighel, G., Hajnal, J.V., Rutherford, M.A., Edwards, A.D., 2013b. Development of cortical microstructure in the preterm human brain. Proceedings of the National Academy of Sciences 110, 9541–9546. doi:10.1073/PNAS.1301652110.

Barnett, M.L., Tusor, N., Ball, G., Chew, A., Falconer, S., Aljabar, P., Kimpton, J.A., Kennea, N., Rutherford, M., David Edwards, A., Counsell, S.J., 2018. Exploring the multiple-hit hypothesis of preterm white matter damage using diffusion MRI. NeuroImage: Clinical 17, 596–606. doi:10.1016/J.NICL.2017.11.017.

Bastiani, M., Andersson, J., Cordero-Grande, L., Murgasova, M., Hutter, J., Price, A.N., Makropoulos, A., Fitzgibbon, S.P., Hughes, E., Rueckert, D., Victor, S., Rutherford, M., Edwards, A.D., Smith, S., Tournier, J.D., Hajnal, J.V., Jbabdi, S., Sotiropoulos, S.N., 2018. Automated processing pipeline for neonatal diffusion MRI in the developing human connectome project. NeuroImage doi:https://doi.org/10.1016/j.neuroimage.2018.05.064.

Bastiani, M., Cottaar, M., Fitzgibbon, S.P., Suri, S., Alfaro-Almagro, F., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L., 2019. 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. doi:https://doi.org/10.1016/j.neuroimage.2018.09.073.

Batalle, D., Edwards, A.D., O'Muircheartaigh, J., 2018b. Annual research review: Not just a small adult brain: understanding later neurodevelopment through imaging the neonatal brain. Journal of Child Psychology and Psychiatry 59, 350–371. doi:10.1111/jcpp.12838.

Batalle, D., Hughes, E.J., Zhang, H., Tournier, J.D., Tusor, N., others., 2017. Early development of structural networks and the impact of prematurity on brain connectivity. NeuroImage 149, 379–392. doi:10.1016/j.neuroimage.2017.01.065.

Batalle, D., O'Muircheartaigh, J., Makropoulos, A., Kelly, C.J., Dimitrova, R., Hughes, E.J., Hajnal, J.V., Zhang, H., Alexander, D.C., Edwards, A.D., Counsell, S.J., 2018. Different patterns of cortical maturation before and after 38 weeks gestational age demonstrated by diffusion MRI in vivo. NeuroImage doi:https://doi.org/10.1016/j.neuroimage.2018.05.046.

Baum, G.L., Roalf, D.R., Cook, P.A., Ciric, R., Rosen, A.F., Xia, C., Elliott, M.A., Ruparel, K., Verma, R., Tunç, B., et al., 2018. The impact of in-scanner head motion on structural connectivity derived from diffusion mri. Neuroimage 173, 275–286.

Blesa, M., Sullivan, G., Anblagan, D., Telford, E.J., Quigley, A.J., Sparrow, S.A., Serag, A., Semple, S.I., Bastin, M.E., Boardman, J.P., 2019. Early breast milk exposure modifies brain connectivity in preterm infants. NeuroImage 184, 431 – 439. doi:https://doi.org/10.1016/j.neuroimage.2018.09.045.

Boardman, J., Craven, C., Valappil, S., Counsell, S., Dyet, L., Rueckert, D., Aljabar, P., Rutherford, M., Chew, A., Allsop, J., Cowan, F., Edwards, A.,

2010. A common neonatal image phenotype predicts adverse neurodevelopmental outcome in children born preterm. NeuroImage 52, 409 – 414. doi:https://doi.org/10.1016/j.neuroimage.2010.04.261.

Boardman, J.P., Counsell, S.J., Rueckert, D., Kapellou, O., Bhatia, K.K., Aljabar, P., Hajnal, J., Allsop, J.M., Rutherford, M.A., Edwards, A.D., 2006. Abnormal deep grey matter development following preterm birth detected using deformation-based morphometry. NeuroImage 32, 70–78. doi:https://doi.org/10.1016/j.neuroimage.2006.03.029.

Boardman, J.P., Walley, A., Ball, G., Takousis, P., Krishnan, M.L., Hughes-Carre, L., Aljabar, P., Serag, A., King, C., Merchant, N., Srinivasan, L., Froguel, P., Hajnal, J., Rueckert, D., Counsell, S., Edwards, A.D., 2014. Common genetic variants and risk of brain injury after preterm birth. Pediatrics 133, e1655–e1663. doi:10.1542/peds.2013-3011.

Bonifacio, S.L., Glass, H.C., Chau, V., Berman, J.I., Xu, D., Brant, R., Barkovich, A.J., Poskitt, K.J., Miller, S.P., Ferriero, D.M., 2010. Extreme premature birth is not associated with impaired development of brain microstructure. The Journal of Pediatrics 157, 726–732.e1. doi:10.1016/J.JPEDS.2010.05.026.

Bouyssi-Kobar, M., Brossard-Racine, M., Jacobs, M., Murnick, J., Chang, T., Limperopoulos, C., 2018. Regional microstructural organization of the cerebral cortex is affected by preterm birth. NeuroImage: Clinical 18, 871–880. doi:10.1016/j.nicl.2018.03.020.

Brown, C.J., Miller, S.P., Booth, B.G., Andrews, S., Chau, V., Poskitt, K.J., Hamarneh, G., 2014. Structural network analysis of brain development in young preterm neonates. NeuroImage 101, 667 – 680. doi:https://doi.org/10.1016/j.neuroimage.2014.07.030.

Brown, C.J., Moriarty, K.P., Miller, S.P., Booth, B.G., et al., 2017. Prediction of brain network age and factors of delayed maturation in very preterm

infants, in: Lecture Notes in Computer Science, pp. 84–91. doi:10.1007/978-3-319-66182-7_10.

Brown, T., Kuperman, J., Chung, Y., Erhart, M., McCabe, C., Hagler, D., Venkatraman, V., Akshoomoff, N., Amaral, D., Bloss, C., Casey, B., Chang, L., Ernst, T., Frazier, J., Gruen, J., Kaufmann, W., Kenet, T., Kennedy, D., Murray, S., Sowell, E., Jernigan, T., Dale, A., 2012. Neuroanatomical Assessment of Biological Maturity. Current Biology 22, 1693–1698. doi:10.1016/J.CUB.2012.07.002.

Caldinelli, C., Froudist-Walsh, S., Karolis, V., Tseng, C.E., Allin, M.P., Walshe, M., Cuddy, M., Murray, R.M., Nosarti, C., 2017. White matter alterations to cingulum and fornix following very preterm birth and their relationship with cognitive functions. NeuroImage 150, 373 – 382. doi:https://doi.org/10.1016/j.neuroimage.2017.02.026.

Cao, M., Huang, H., He, Y., 2017. Developmental connectomics from infancy through early childhood. Trends in neurosciences 40, 494 – 506.

Caruyer, E., Lenglet, C., Sapiro, G., Deriche, R., 2013. Design of multishell sampling schemes with uniform coverage in diffusion MRI. Magnetic Resonance in Medicine 69, 1534–1540. doi:10.1002/mrm.24736.

Ceschin, R., Zahner, A., Reynolds, W., Gaesser, J., Zuccoli, G., Lo, C.W., Gopalakrishnan, V., Panigrahy, A., 2018. A computational framework for the detection of subcortical brain dysmaturation in neonatal MRI using 3D Convolutional Neural Networks. NeuroImage 178, 183–197. doi:10.1016/J.NEUROIMAGE.2018.05.049.

Counsell, S.J., Edwards, A.D., Chew, A.T.M., Cowan, F.M., Boardman, J.P., Allsop, J.M., Hajnal, J.V., Srinivasan, L., Dyet, L.E., Rutherford, M.A., Anjari, M., 2008. Specific relations between neurodevelopmental abilities and white matter microstructure in children born preterm. Brain 131, 3201–3208.

Deprez, M., Wang, S., Ledig, C., Hajnal, J., Counsell, S., Schnabel, J., 2018. Segmentation of myelin-like signals on clinical MR images for age estimation in preterm infants. bioRxiv doi:10.1101/357749.

Dubois, J., Benders, M., Lazeyras, F., Borradori-Tolsa, C., Leuchter, R.H.V., Mangin, J., Hüppi, P., 2010. Structural asymmetries of perisylvian regions in the preterm newborn. NeuroImage 52, 32–42. doi:10.1016/J.NEUROIMAGE.2010.03.054.

Duerden, E.G., Grunau, R.E., Guo, T., Foong, J., Pearson, A., Au-Young, S., Lavoie, R., Chakravarty, M.M., Chau, V., Synnes, A., Miller, S.P., 2018. Early procedural pain is associated with regionally-specific alterations in thalamic development in preterm neonates. Journal of Neuroscience 38, 878–886. URL: https://www.jneurosci.org/content/38/4/878, doi:10.1523/JNEUROSCI.0867-17.2017, arXiv:https://www.jneurosci.org/content/38/4/878.full.pdf.

Duerden, E.G., Guo, T., Dodbiba, L., Chakravarty, M.M., Chau, V., Poskitt, K.J., Synnes, A., Grunau, R.E., Miller, S.P., 2016. Midazolam dose correlates with abnormal hippocampal growth and neurodevelopmental outcome in preterm infants. Annals of Neurology 79, 548–559. doi:10.1002/ana.24601.

Eaton-Rosen, Z., Melbourne, A., Orasanu, E., Cardoso, M.J., Modat, M., Bainbridge, A., Kendall, G.S., Robertson, N.J., Marlow, N., Ourselin, S., 2015. Longitudinal measurement of the developing grey matter in preterm subjects using multi-modal MRI. NeuroImage 111, 580–589. doi:10.1016/J.NEUROIMAGE.2015.02.010.

Efron, B., 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American statistical association 78, 316–331.

Genc, S., Malpas, C.B., Holland, S.K., Beare, R., Silk, T.J., 2017. Neurite density index is sensitive to age related differences in the developing brain. NeuroImage 148, 373–380. doi:10.1016/j.neuroimage.2017.01.023.

Glasser, M.F., Van Essen, D.C., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-Weighted MRI. Journal of Neuroscience 31, 11597–11616. doi:10.1523/JNEUROSCI.2180-11.2011.

Gousias, I.S., Edwards, A.D., Rutherford, M.A., Counsell, S.J., Hajnal, J.V., Rueckert, D., Hammers, A., 2012. Magnetic resonance imaging of the newborn brain: Manual segmentation of labelled atlases in term-born and preterm infants. NeuroImage 62, 1499 – 1509. doi:https://doi.org/10.1016/j.neuroimage.2012.05.083.

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. NeuroImage 48, 63–72. doi:10.1016/j.neuroimage.2009.06.060.

Grussu, F., Schneider, T., Tur, C., Yates, R.L., Tachrount, M., Ianuş, A., Yiannakas, M.C., Newcombe, J., Zhang, H., Alexander, D.C., et al., 2017. Neurite dispersion: a new marker of multiple sclerosis spinal cord pathology? Annals of clinical and translational neurology 4, 663–679.

Gui, L., Loukas, S., Lazeyras, F., Hüppi, P., Meskaldji, D., Borradori Tolsa, C., 2019. Longitudinal study of neonatal brain tissue volumes in preterm infants and their ability to predict neurodevelopmental outcome. NeuroImage 185, 728–741. doi:10.1016/J.NEUROIMAGE.2018.06.034.

Hernandez-Fernandez, M., Reguly, I., Jbabdi, S., Giles, M., Smith, S., Sotiropoulos, S.N., 2019. Using gpus to accelerate computational diffusion MRI: From microstructure estimation to tractography and connectomes. NeuroImage 188, 598 – 615. doi:https://doi.org/10.1016/j.neuroimage.2018.12.015.

Hunter, J.D., 2007. Matplotlib: A 2d graphics environment. Computing in science & engineering 9, 90.

Jelescu, I.O., Veraart, J., Adisetiyo, V., Milla, S.S., Novikov, D.S., Fieremans, E., 2015. One diffusion acquisition and different white

matter models: How does microstructure change in human early development based on wmti and noddi? NeuroImage 107, 242 – 256. URL: http://www.sciencedirect.com/science/article/pii/S1053811914010015, doi:https://doi.org/10.1016/j.neuroimage.2014.12.009.

Jensen, J., Helpern, J., Ramani, A., Lu, H., Kaczynski, K., 2005. Diffusional kurtosis imaging: The quantification of nongaussian water diffusion by means of magnetic resonance imaging. Magnetic Resonance in Medicine 53, 1432–1440. doi:10.1002/mrm.20508.

Job, D.E., Dickie, D.A., Rodriguez, D., Robson, A., Danso, S., Pernet, C., Bastin, M.E., Boardman, J.P., Murray, A.D., Ahearn, T., Waiter, G.D., Staff, R.T., Deary, I.J., Shenkin, S.D., Wardlaw, J.M., 2017. A brain imaging repository of normal structural mri across the life course: Brain images of normal subjects (brains). NeuroImage 144, 299 – 304. URL: http://www.sciencedirect.com/science/article/pii/S1053811916000331, doi:https://doi.org/10.1016/j.neuroimage.2016.01.027. data Sharing Part II.

Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open source scientific tools for Python,. http://www.scipy.org/.

Karmacharya, S., Gagoski, B., Ning, L., Vyas, R., Cheng, H.H., Soul, J., Newberger, J.W., Shenton, M.E., Rathi, Y., Grant, P.E., 2018. Advanced diffusion imaging for assessing normal white matter development in neonates and characterizing aberrant development in congenital heart disease. NeuroImage: Clinical 19, 360–373. doi:10.1016/j.nicl.2018.04.032.

Kellner, E., Dhital, B., Kiselev, V.G., Reisert, M., 2016. Gibbs-ringing artifact removal based on local subvoxel-shifts. Magnetic Resonance in Medicine 76, 1574–1581. doi:10.1002/mrm.26054.

Keunen, K., Benders, M.J., Leemans, A., Fieret-Van Stam, P.C., Scholtens, L.H., Viergever, M.A., Kahn, R.S., Groenendaal, F., de Vries, L.S., van den

Heuvel, M.P., 2017. White matter maturation in the neonatal brain is predictive of school age cognitive capacities in children born very preterm. Developmental Medicine & Child Neurology 59, 939–946.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the 14th international joint conference on Artificial intelligence-Volume 2, Morgan Kaufmann Publishers Inc.. pp. 1137–1143.

Kostović, I., Jovanov-Milošević, N., 2006. The development of cerebral connections during the first 2045 weeks' gestation. Seminars in Fetal and Neonatal Medicine 11, 415–422. doi:10.1016/J.SINY.2006.07.001.

Krishnan, M.L., Van Steenwinckel, J., Schang, A.L., Yan, J., Arnadottir, J., Le Charpentier, T., Csaba, Z., Dournaud, P., Cipriani, Constance Auvynet, S., Titomanlio, L., Pansiot, J., Ball, G., Boardman, J.P., Walley, A.J., Saxena, A., Mirza, G., Fleiss, B., Edwards, A.D., Petretto, E., Gressens, P., 2017. Integrative genomics of microglia implicates dlg4 (psd95) in the white matter development of preterm infants. Nature Communications 8.

Krishnan, M.L., Wang, Z., Silver, M., Boardman, J.P., Ball, G., Counsell, S.J., Walley, A.J., Montana, G., Edwards, A.D., 2016. Possible relationship between common genetic variation and white matter development in a pilot study of preterm infants. Brain and behavior 6, e00434. doi:10.1002/brb3.434.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. Genome research 19, 1639–45. doi:10.1101/gr.092759.109.

Kulikova, S., Hertz-Pannier, L., Dehaene-Lambertz, G., Buzmakov, A., Poupon, C., Dubois, J., 2015. Multi-parametric evaluation of the white matter maturation. Brain Structure and Function 220, 3657–3672. doi:10.1007/s00429-014-0881-y.

Kunz, N., Zhang, H., Vasung, L., O'Brien, K.R., Assaf, Y., Lazeyras, F., Alexander, D.C., Hüppi, P.S., 2014. Assessing white matter microstructure of the newborn with multi-shell diffusion MRI and biophysical compartment models. NeuroImage 96, 288–299. doi:10.1016/j.neuroimage.2014.03.057.

Leuchter, R.H.V., Gui, L., Poncet, A., Hagmann, C., Lodygensky, G.A., Martin, E., Koller, B., Darqu, A., Bucher, H.U., Hppi, P.S., 2014. Association Between Early Administration of High-Dose Erythropoietin in Preterm Infants and Brain MRI Abnormality at Term-Equivalent Age. JAMA 312, 817–824.

Li, W., Yang, C., Shi, F., Wu, S., Wang, Q., Nie, Y., Zhang, X., 2017. Construction of individual morphological brain networks with multiple morphometric features. Frontiers in Neuroanatomy 11. doi:10.3389/fnana.2017.00034.

Li, X., Morgan, P.S., Ashburner, J., Smith, J., Rorden, C., 2016. The first step for neuroimaging data analysis: Dicom to nifti conversion. Journal of Neuroscience Methods 264, 47 – 56. doi:https://doi.org/10.1016/j.jneumeth.2016.03.001.

Mahjoub, I., Mahjoub, M.A., Rekik, I., 2018. Brain multiplexes reveal morphological connectional biomarkers fingerprinting late brain dementia states. Scientific Reports 8, 4103. doi:10.1038/s41598-018-21568-7.

Makropoulos, A., Aljabar, P., Wright, R., Hüning, B., Merchant, N., et al., 2016. Regional growth and atlasing of the developing human brain. NeuroImage 125, 456–478. doi:10.1016/j.neuroimage.2015.10.047.

Makropoulos, A., Gousias, I.S., Ledig, C., Aljabar, P., Serag, A., Hajnal, J.V., Edwards, A.D., Counsell, S.J., Rueckert, D., 2014. Automatic whole brain MRI segmentation of the developing neonatal brain. IEEE Transactions on Medical Imaging 33, 1818–1831. doi:10.1109/TMI.2014.2322280.

Makropoulos, A., Robinson, E.C., Schuh, A., Wright, R., Fitzgibbon, S., et al., 2018. The developing human connectome project: A minimal processing

pipeline for neonatal cortical surface reconstruction. NeuroImage 173, 88–112. doi:10.1016/j.neuroimage.2018.01.054.

Marcus, D., Harwell, J., Olsen, T., Hodge, M., Glasser, M., Prior, F., Jenkinson, M., Laumann, T., Curtiss, S., Van Essen, D., 2011. Informatics and data mining tools and strategies for the human connectome project. Frontiers in Neuroinformatics 5, 4.

Mathewson, K., Chow, C., Dobson, K., Pope, E., Schmidt, L., Van Lieshout, R., 2017. Mental health of extremely low birth weight survivors: A systematic review and meta-analysis. Psychological Bulletin 143, 347 – 383.

Maximov, I.I., Alnaes, D., Westlye, L.T., 2019. Towards an optimised processing pipeline for diffusion MRI data: Effects of artefact corrections on diffusion metrics and their age associations in UK Biobank. bioRxiv .

McKinney, W., et al., 2010. Data structures for statistical computing in python, in: Proceedings of the 9th Python in Science Conference, Austin, TX. pp. 51–56.

Melbourne, A., Eaton-Rosen, Z., Orasanu, E., Price, D., Bainbridge, A., Cardoso, M.J., Kendall, G.S., Robertson, N.J., Marlow, N., Ourselin, S., 2016. Longitudinal development in the preterm thalamus and posterior white matter: MRI correlations between diffusion weighted imaging and T2 relaxometry. Human Brain Mapping 37, 2479–2492.

Melbourne, A., Kendall, G.S., Cardoso, M.J., Gunny, R., Robertson, N.J., Marlow, N., Ourselin, S., 2014. Preterm birth affects the developmental synergy between cortical folding and cortical connectivity observed on multimodal MRI. NeuroImage 89, 23–34. doi:10.1016/J.NEUROIMAGE.2013.11.048.

Nosarti, C., Reichenberg, A., Murray, R.M., Cnattingius, S., Lambe, M.P., Yin, L., MacCabe, J., Rifkin, L., Hultman, C.M., 2012. Preterm Birth and Psychiatric Disorders in Young Adult Life. Archives of General Psychiatry 69, 610–617.

Otsuka, Y., Chang, L., Kawasaki, Y., Wu, D., Ceritoglu, C., Oishi, K., Ernst, T., Miller, M., Mori, S., Oishi, K., 2019. A multi-atlas label fusion tool for neonatal brain mri parcellation and quantification. Journal of Neuroimaging .

Ouyang, M., Dubois, J., Yu, Q., Mukherjee, P., Huang, H., 2018. Delineation of early brain development from fetuses to infants with diffusion MRI and beyond. NeuroImage doi:10.1016/j.neuroimage.2018.04.017.

Ouyang, M., Jeon, T., Sotiras, A., Peng, Q., Mishra, V., Halovanic, C., Chen, M., Chalak, L., Rollins, N., Roberts, T.P.L., Davatzikos, C., Huang, H., 2019. Differential cortical microstructural maturation in the preterm human brain with diffusion kurtosis and tensor imaging. Proceedings of the National Academy of Sciences , 201812156doi:10.1073/PNAS.1812156116.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. Journal of machine learning research 12, 2825–2830.

Schneider, J., Fischer Fumeaux, C.J., Duerden, E.G., Guo, T., Foong, J., Graz, M.B., Hagmann, P., Chakravarty, M.M., Hüppi, P.S., Beauport, L., Truttmann, A.C., Miller, S.P., 2018. Nutrient intake in the first two weeks of life and brain growth in preterm neonates. Pediatrics 141. doi:10.1542/peds.2017-2169.

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and statistical modeling with python, in: Proceedings of the 9th Python in Science Conference, Scipy. p. 61.

Seidlitz, J., Váša, F., Shinn, M., Romero-Garcia, R., Whitaker, K.J., et al., 2018. Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation. Neuron 97, 231–247.e7. doi:10.1016/j.neuron.2017.11.039.

Serag, A., Aljabar, P., Ball, G., Counsell, S.J., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D., 2012. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. NeuroImage 59, 2255 – 2265. doi:https://doi.org/10.1016/j.neuroimage.2011.09.062.

Shi, F., Yap, P.T., Gao, W., Lin, W., Gilmore, J.H., Shen, D., 2012. Altered structural connectivity in neonates at genetic risk for schizophrenia: A combined study using morphological and white matter networks. NeuroImage 62, 1622–1633. doi:10.1016/j.neuroimage.2012.05.026.

Smith, S.M., 2002. Fast robust automated brain extraction. Human Brain Mapping 17, 143–155. doi:10.1002/hbm.10062.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., Luca, M.D., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural mr image analysis and implementation as fsl. NeuroImage 23, S208 – S219. doi:https://doi.org/10.1016/j.neuroimage.2004.07.051. mathematics in Brain Imaging.

Smyser, C.D., Dosenbach, N.U., Smyser, T.A., Snyder, A.Z., Rogers, C.E., Inder, T.E., Schlaggar, B.L., Neil, J.J., 2016a. Prediction of brain maturity in infants using machine-learning algorithms. NeuroImage 136, 1–9. doi:10.1016/J.NEUROIMAGE.2016.05.029.

Smyser, T.A., Smyser, C.D., Rogers, C.E., Gillespie, S.K., Inder, T.E., Neil, J.J., 2016b. Cortical gray and adjacent white matter demonstrate synchronous maturation in very preterm infants. Cerebral Cortex 26, 3370–3378. doi:10.1093/cercor/bhv164.

Soussia, M., Rekik, I., 2018. Unsupervised manifold learning using high-order morphological brain networks derived from T1-w MRI for autism diagnosis. Frontiers in Neuroinformatics 12, 70. doi:10.3389/fninf.2018.00070.

Sparrow, S., Manning, J.R., Cartier, J., Anblagan, D., Bastin, M.E., Piyasena, C., Pataky, R., Moore, E.J., Semple, S.I., Wilkinson, A.G., Evans, M., Drake, A.J., Boardman, J.P., 2016. Epigenomic profiling of preterm infants reveals DNA methylation differences at sites associated with neural function. Translational psychiatry 6, e716. doi:10.1038/tp.2015.210.

Steven, A.J., Zhuo, J., Melhem, E.R., 2014. Diffusion kurtosis imaging: an emerging technique for evaluating the microstructural environment of the brain. AJR. American journal of roentgenology 202 1, W26–33.

Tariq, M., Schneider, T., Alexander, D.C., Wheeler-Kingshott, C.A.G., Zhang, H., 2016. Bingham-NODDI: Mapping anisotropic orientation dispersion of neurites using diffusion MRI. NeuroImage 133, 207 – 223. doi:https://doi.org/10.1016/j.neuroimage.2016.01.046.

Telford, E.J., Cox, S.R., Fletcher-Watson, S., Anblagan, D., Sparrow, S., Pataky, R., Quigley, A., Semple, S.I., Bastin, M.E., Boardman, J.P., 2017. A latent measure explains substantial variance in white matter microstructure across the newborn human brain. Brain Structure and Function 222, 4023–4033.

Thompson, D.K., Chen, J., Beare, R., Adamson, C.L., Ellis, R., Ahmadzai, Z.M., Kelly, C.E., Lee, K.J., Zalesky, A., Yang, J.Y., Hunt, R.W., Cheong, J.L., Inder, T.E., Doyle, L.W., Seal, M.L., Anderson, P.J., 2016. Structural connectivity relates to perinatal factors and functional impairment at 7years in children born very preterm. NeuroImage 134, 328 – 337. doi:https://doi.org/10.1016/j.neuroimage.2016.03.070.

Thompson, D.K., Kelly, C.E., Chen, J., Beare, R., Alexander, B., Seal, M.L., Lee, K., Matthews, L.G., Anderson, P.J., Doyle, L.W., Spittle, A.J., Cheong, J.L., 2018a. Early life predictors of brain development at term-equivalent age in infants born across the gestational age spectrum. NeuroImage doi:10.1016/j.neuroimage.2018.04.031.

Thompson, D.K., Kelly, C.E., Chen, J., Beare, R., Alexander, B., Seal, M.L., Lee, K.J., Matthews, L.G., Anderson, P.J., Doyle, L.W., Cheong, J.L., Spittle, A.J., 2018b. Characterisation of brain volume and microstructure at term-equivalent age in infants born across the gestational age spectrum. NeuroImage: Clinical , 101630.

Toews, M., Wells, W.M., Zöllei, L., 2012. A feature-based developmental model of the infant brain in structural MRI, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 204–211.

Toulmin, H., Beckmann, C.F., O'Muircheartaigh, J., Ball, G., Nongena, P., Makropoulos, A., Ederies, A., Counsell, S.J., Kennea, N., Arichi, T., Tusor, N., Rutherford, M.A., Azzopardi, D., Gonzalez-Cinca, N., Hajnal, J.V., Edwards, A.D., 2015. Specialization and integration of functional thalamocortical connectivity in the human infant. Proceedings of the National Academy of Sciences 112, 6485–6490.

Tournier, J.D., Smith, R.E., Raffelt, D.A., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.H., Connelly, A., 2019. Mrtrix3: A fast, flexible and open software framework for medical image processing and visualisation. bioRxiv doi:10.1101/551739.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: Improved N3 bias correction. IEEE Transactions on Medical Imaging 29, 1310–1320. doi:10.1109/TMI.2010.2046908.

Van Den Heuvel, M.P., Kersbergen, K.J., De Reus, M.A., Keunen, K., et al., 2015. The neonatal connectome during preterm brain development. Cerebral Cortex 25, 3000–3013. doi:10.1093/cercor/bhu095.

Van Der Walt, S., Colbert, S.C., Varoquaux, G., 2011. The numpy array: a structure for efficient numerical computation. Computing in Science & Engineering 13, 22.

Van Lieshout, R.J., Ferro, M.A., Schmidt, L.A., Boyle, M.H., Saigal, S., Morrison, K.M., Mathewson, K.J., 2018. Trajectories of psychopathology in extremely low birth weight survivors from early adolescence to adulthood: a 20-year longitudinal study. Journal of Child Psychology and Psychiatry 59, 1192–1200.

Varoquaux, G., 2018. Cross-validation failure: small sample sizes lead to large error bars. Neuroimage 180, 68–77.

Varoquaux, G., Raamana, P.R., Engemann, D.A., Hoyos-Idrobo, A., Schwartz, Y., Thirion, B., 2017. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. NeuroImage 145, 166–179.

Veraart, J., Fieremans, E., Novikov, D.S., 2016. Diffusion MRI noise mapping using random matrix theory. Magnetic Resonance in Medicine 76, 1582–1593. doi:10.1002/mrm.26059.

Veraart, J., Novikov, D.S., Christiaens, D., Ades-aron, B., Sijbers, J., Fieremans, E., 2016b. Denoising of diffusion MRI using random matrix theory. NeuroImage 142, 394–406. doi:10.1016/j.neuroimage.2016.08.016.

Weiskopf, N., Suckling, J., Williams, G., Correia, M.M., Inkster, B., Tait, R., Ooi, C., Bullmore, E.T., Lutti, A., 2013. Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation. Frontiers in Neuroscience 7. doi:10.3389/fnins.2013.00095.

Woodward, L.J., Anderson, P.J., Austin, N.C., Howard, K., Inder, T.E., 2006. Neonatal mri to predict neurodevelopmental outcomes in preterm infants. New England Journal of Medicine 355, 685–694.

Wu, Z., Li, G., Shen, D., Hu, D., Lin, W., 2019. Hierarchical rough-to-fine model for infant age prediction based on cortical features. IEEE Journal of Biomedical and Health Informatics , 1–1doi:10.1109/jbhi.2019.2897020.

Yap, P.T., Fan, Y., Chen, Y., Gilmore, J.H., Lin, W., Shen, D., 2011. Development trends of white matter connectivity in the first years of life. PLoS ONE 6, e24678. doi:10.1371/journal.pone.0024678.

Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C., 2012. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. NeuroImage 61, 1000–1016. doi:10.1016/j.neuroimage.2012.03.072.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301–320.

**Credit Author Statement**

Paola Galdi: Conceptualization, Methodology, Software, Writing - Original Draft
Manuel Blesa: Conceptualization, Methodology, Software, Writing - Original Draft
David Q. Stoye: Resources, Data Curation, Writing - Review & Editing
Gemma Sullivan: Resources, Data Curation, Writing - Review & Editing
Gillian J. Lamb: Resources, Data Curation, Writing - Review & Editing
Alan J. Quigley: Resources, Data Curation, Writing - Review & Editing
Michael J. Thrippleton: Resources, Data Curation, Writing - Review & Editing
Mark E. Bastin: Resources, Data Curation, Writing - Review & Editing, Supervision
James P. Boardman: Resources, Writing - Review & Editing, Supervision, Funding acquisition