



Published in final edited form as:

*Pediatr Res.* 2021 May ; 89(6): 1452–1460. doi:10.1038/s41390-020-01134-6.

## White matter tracts related to memory and emotion in very preterm children

Claire E. Kelly<sup>1,2,\*</sup>, Deanne K. Thompson<sup>1,2,3,4</sup>, Malcolm Cooper<sup>1,2</sup>, Jenny Pham<sup>1,2</sup>, Thanh D. Nguyen<sup>1,2</sup>, Joseph Y.M. Yang<sup>2,4,5,6</sup>, Gareth Ball<sup>2</sup>, Chris Adamson<sup>2</sup>, Andrea L. Murray<sup>1,2</sup>, Jian Chen<sup>2</sup>, Terrie E. Inder<sup>7</sup>, Jeanie L.Y. Cheong<sup>1,8,9</sup>, Lex W. Doyle<sup>1,4,8,9</sup>, Peter J. Anderson<sup>1,10</sup>

<sup>1</sup>Victorian Infant Brain Studies (VIBeS), Murdoch Children's Research Institute, Melbourne, Australia

<sup>2</sup>Developmental Imaging, Murdoch Children's Research Institute, Melbourne, Australia

<sup>3</sup>Florey Institute of Neuroscience and Mental Health, Melbourne, Australia

<sup>4</sup>Department of Paediatrics, The University of Melbourne, Melbourne, Australia

<sup>5</sup>Neuroscience Advanced Clinical Imaging Suite (NACIS), Department of Neurosurgery, The Royal Children's Hospital, Melbourne, Australia

<sup>6</sup>Neuroscience Research, Murdoch Children's Research Institute, Melbourne, Australia

<sup>7</sup>Department of Pediatric Newborn Medicine, Brigham and Women's Hospital, Boston, MA

<sup>8</sup>Newborn Research, The Royal Women's Hospital, Melbourne, Australia

<sup>9</sup>Department of Obstetrics and Gynaecology, The University of Melbourne, Melbourne, Australia

<sup>10</sup>Turner Institute for Brain and Mental Health and School of Psychological Sciences, Monash University, Melbourne, Australia

### Abstract

**Background**—Very preterm (VP) children are at risk of memory and emotional impairments, however the neural correlates remain incompletely defined. This study investigated the effect of VP birth on white matter tracts traditionally related to episodic memory and emotion.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

**\*Corresponding author:** Claire Kelly, Victorian Infant Brain Studies (VIBeS), Murdoch Children's Research Institute, The Royal Children's Hospital, 50 Flemington Road, Parkville, Victoria, Australia, 3052. [claire.kelly@mcri.edu.au](mailto:claire.kelly@mcri.edu.au)

**Author contributions:** Each author has met the Pediatric Research authorship requirements listed below. All authors (CEK, DKT, MC, JP, TDN, JYMY, GB, CA, ALM, JC, TEI, JLYC, LWD, PJA) completed the following: substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published.

**Disclosure statement:** The authors have no disclosures.

**Category of study:** Basic Science

**Patient consent:** Written informed consent was obtained from all parents or caregivers.

**Methods**—The cingulum, fornix, uncinate fasciculus, medial forebrain bundle and anterior thalamic radiation were reconstructed using tractography in 144 VP children and 33 full-term controls at age 7 years.

**Results**—Compared with controls, VP children had higher axial, radial and mean diffusivities and neurite orientation dispersion, and lower volume and neurite density in the fornix, along with higher neurite orientation dispersion in the medial forebrain bundle. Support vector classification models based on tract measures significantly classified VP children and controls. Higher fractional anisotropy and lower diffusivities in the cingulum, uncinate fasciculus, medial forebrain bundle and anterior thalamic radiation were associated with better episodic memory, independent of key perinatal risk factors. Support vector regression models using tract measures did not predict episodic memory and emotional outcomes.

**Conclusions**—Altered tract structure is related to adverse episodic memory outcomes in VP children, but further research is required to determine the ability of tract structure to predict outcomes of individual children.

---

## Introduction

Children born very preterm (VP, <32 weeks' GA) have higher rates of neurodevelopmental delays compared with children born full-term (FT, ≥37 weeks' GA), including memory impairments and emotional difficulties (1, 2). The mechanisms responsible are not clear but may be better understood by characterising the neuroanatomical underpinnings using MRI.

Episodic memory and emotion are thought to be supported by a complex cortical and subcortical circuitry involving the mesial temporal structures, medial forebrain, hypothalamus, thalamus, orbitofrontal lobe and brainstem, interconnected by white matter tracts including the fornix, stria terminalis, mammillo-thalamic tract, anterior thalamic radiation, cingulum, uncinate fasciculus, and medial forebrain bundle (3). VP birth has been associated with reduced hippocampal volume (4, 5), but we previously found that hippocampal volume was not associated with concurrent memory performance in VP children at age 7 years (5). Disruptions of the hippocampal efferent and afferent tracts, and the broader system involved in episodic memory and emotion, might better explain episodic memory and emotional impairments in VP children. Some studies using tractography have found that the fornix and cingulum (6, 7), uncinate fasciculus (8, 9), and anterior thalamic radiation (9) have altered microstructure in preterm children and young adults compared with FT controls. Only a few studies have reported associations between fornix and cingulum volume and microstructure and memory outcomes in preterm young adults (6, 7). Additional studies would be beneficial to characterise a broader range of tracts involved in episodic memory and emotion, and their relationships with episodic memory and emotional outcomes.

Previous research in this area generally examined differences in brain structure between groups of preterm and FT children, and associations between brain structure and neurodevelopmental outcomes in preterm children at a group level. Machine learning methods enable a more individualised approach (10, 11). Machine learning has previously been used to classify preterm and FT individuals based on MRI data (11–13), and to

investigate whether MRI data can predict the longer-term neurodevelopmental functioning of individual preterm children (for example (14–16)). Support vector machines are a type of supervised machine learning algorithm that have been applied successfully and relatively frequently on cohorts of preterm children (11, 14, 15). We aimed to expand the application of support vector machine learning to investigate the effects of VP birth on tracts traditionally related to episodic memory and emotion in childhood.

This study uses data collected as part of a large longitudinal cohort study with previous publications (for example (1, 4, 5, 15, 17)). The current manuscript is the first to examine a broad range of tracts involved in episodic memory and emotion using tractography, and is the first to apply advanced machine learning methods to analyse tracts. In summary, our first aim was to investigate the volume and microstructure of several of the major tracts traditionally involved in episodic memory and emotion in VP children compared with FT children at age 7 years, including the cingulum, fornix, uncinate fasciculus, medial forebrain bundle and anterior thalamic radiation. Our second aim was to investigate whether the structure of these tracts is related to episodic memory and emotional outcomes in VP and FT children. While the specific roles of individual tracts in episodic memory and emotion are still not completely understood, we hypothesised that cingulum and fornix volume and microstructure would be most strongly related to episodic memory, while uncinate fasciculus and medial forebrain bundle volume and microstructure would be most strongly related to emotional outcome.

## Methods

### Participants

VP infants (born <30 weeks' GA and/or <1250 g;  $n = 224$ ) and FT infants (born ≥ 37 weeks' GA;  $n = 46$ ) were recruited during the neonatal period, between July 2001 and December 2003, from the Royal Women's Hospital, Melbourne. Infants with genetic or congenital abnormalities likely to affect development were excluded. At age 7 years, participants were invited to attend a follow-up study. Of the originally recruited participants, 198 VP participants (88% of the original cohort) and 43 FT participants (93% of the original cohort) attended the follow-up, of whom 159 VP and 36 FT participants had MRI. The study was approved by the Royal Children's Hospital, Melbourne, Human Research Ethics Committee, and written informed consent was obtained from all parents or caregivers.

### Magnetic resonance imaging

Participants underwent MRI in a 3T scanner (Siemens Tim Trio, Erlangen, Germany). All participants were scanned with two diffusion-weighted (single-shot twice-refocused echo planar imaging) sequences. The first sequence (' $b1200$ ') had multiple  $b$ -values ranging from 50 to 1200  $\text{s/mm}^2$  in increments of 50 (the same  $b$ -values were used for all children), 25 gradient directions in total (i.e. one direction per  $b$ -value, except  $b = 300$  and  $b = 600$  had two directions), one  $b = 0$   $\text{s/mm}^2$  volume, repetition time = 12000 ms, echo time = 96 ms, field of view =  $250 \times 250$  mm, matrix size =  $144 \times 144$ , and voxel size =  $1.7 \text{ mm}^3$ . This was an advanced acquisition scheme at the time (2008 – 2011), and several studies utilising this type of multiple  $b$ -value sequence have been published (for example (18)). The second

sequence ('*b3000*') had *b*-values of 3000 s/mm<sup>2</sup>, 45 gradient directions, six *b* = 0 s/mm<sup>2</sup> volumes, repetition time = 7400 ms, echo time = 106 ms, field of view = 240 × 240 mm, matrix size = 104 × 104, and voxel size = 2.3 mm<sup>3</sup>.

## Tractography

We reconstructed the white matter tracts from the *b3000* images using MRtrix version 0.2. Tractography was performed manually over several years. While MRtrix3 became publicly available during the tractography (November 2013), we did not update the version in order to avoid introducing inconsistencies into the analysis. Tractography was based on the regions of interest detailed in Supplementary Table 1. After placing the regions of interest, we delineated the tracts using a probabilistic algorithm based on fibre orientation distributions provided by constrained spherical deconvolution (*l*<sub>max</sub> 6) (19) (Figure 1).

Tractography was repeated on a subset of randomly selected participants to determine intra-rater reliability for the single operator who delineated each tract. Intra-rater reliability for the cingulum was >0.8 (17). Intra-rater reliabilities for the remaining tracts are reported in Supplementary Table 2, and were all >0.8.

## Tract measures

Diffusion tensor measures [fractional anisotropy (FA) and axial (AD), radial (RD), and mean (MD) diffusivities] were generated from the *b1200* images. This involved correcting the images for motion and eddy current-induced distortions, incorporating *b*-matrix reorientation, and fitting the diffusion tensor model based on the weighted linear least squares method, using ExploreDTI software. We did not remove any volumes affected by motion, however we excluded participants if their diffusion images exhibited severe motion artefact based on visual inspection. Motion parameters for this sequence were similar between the groups (Supplementary Table 3).

Prior to Neurite Orientation Dispersion and Density Imaging (NODDI), we processed the images using the Functional MRI of the Brain Software Library (FSL). This involved: 1. linearly registering the *b3000* images to the *b1200* images using the FSL linear image registration tool (FLIRT) (20); 2. merging the two sequences together; 3. motion correcting the combined sequences using the 'eddy\_correct' tool (incorporating *b*-vector reorientation); 4. normalising each sequence separately by its *b0* image/s to attempt to account for the different TEs and TRs between the sequences. Finally, we fitted the NODDI model to the combined sequences using the NODDI Matlab toolbox version 0.9 (21). NODDI is a multi-compartment tissue microstructure model. The intracellular compartment is modelled as sticks, whose volume fraction represents the neurite density index (range 0–1). The orientation distribution of the sticks is modelled by a Watson distribution, which provides a measure of the neurite orientation dispersion index (range 0–1, from coherently aligned axons with low dispersion to bending, fanning or crossing axons with high dispersion) (21).

Diffusion tensor and NODDI images in *b1200* space were aligned to tract (*b3000*) space using FLIRT (20). Tractography reconstructions from MRtrix were converted to probability (track density) images, thresholded at 0.01 (voxels with <10 out of 1000 streamlines were removed) and binarised. Resulting tract masks were multiplied by diffusion tensor images

and NODDI images that were in  $b3000$  space. Diffusion tensor and NODDI values were averaged across all the voxels in each tract, and tract volumes were calculated (number of voxels in the tract masks multiplied by voxel volume in  $\text{cm}^3$ ).

### Episodic memory and emotional outcome assessments

Verbal memory and learning was assessed at age 7 years using the California Verbal Learning Task- Children's Version (22). The children were required to recall a list of 15 words, which was presented over 5 trials. The variable of interest was the number of correctly recalled words over the 5 trials (verbal episodic memory and learning). Visual memory and learning was assessed using the Dot Locations Test from the Children's Memory Scale (23). The children were required to recall the position of 6 dots presented on a  $3 \times 3$  grid, which was presented over 3 trials. The variable of interest was the number of correct locations recalled over 3 trials (visual episodic memory and learning). Age standardised scores were used, with higher scores indicating better performance. Emotional problems were assessed using the Emotional Symptoms subscale of the parent-report Strengths and Difficulties Questionnaire (24). This subscale consists of five items assessing somatic complaints (e.g. headaches), anxiety (e.g. worries, fears, nervousness), and mood (e.g. unhappy, downhearted) (24). Higher scores indicate more emotional problems.

## Statistical Analysis

### Linear regression

Linear regressions were performed using Stata version 14. All linear regression models were fitted using generalised estimating equations to account for correlations between data from multiple births. Additionally, all linear regression analyses were performed separately for all the tract measures (left and right FA, AD, RD, MD, neurite orientation dispersion, neurite density and tract volume); hence the total number of participants with data for each tract could be included in each analysis.

Firstly, we compared tract measures between the VP and FT groups, adjusted for age at MRI and sex, and analyses with tract volumes were additionally adjusted for intracranial volume, to determine whether findings were independent of overall head size. We repeated these analyses with a sex-by-group interaction term, to investigate whether the group differences in tract measures differed between males and females.

Secondly, we explored associations between the tract measures and episodic memory and emotional outcomes in all children. We adjusted for age at neurodevelopmental assessment, sex, GA at birth and global brain abnormality score on term MRI (25), given these factors have previously been associated with neurodevelopmental outcomes in VP children (1, 26). Analyses with tract volumes were also adjusted for intracranial volume. We repeated these analyses with a group-by-tract measure interaction term, to investigate whether the tract measure-outcome relationships differed between the VP and FT groups. We also repeated these analyses with a sex-by-tract measure interaction term, to investigate whether the tract measure-outcome relationships differed between males and females.

All linear regression analyses were false discovery rate (FDR)-corrected for the number of tracts and tract measures (12 by 7 = 84 comparisons); only results that were significant at  $p < 0.05$ , FDR-corrected are reported.

## Machine learning

Machine learning involves building statistical models that can learn from and make predictions about data (10). A set of features (that is, independent variables) is used to predict outcome measurements (that is, dependent variables). Supervised learning involves using ‘training’ data, in which both features and outcomes are known for a set of study participants, to build a model which can be applied to new ‘test’ data to predict unknown outcomes for a set of participants. Support vector machines are a type of supervised learning algorithm. Support vector classification can be used to predict categorical outcomes, while support vector regression can be used to predict continuous outcomes (10).

Support vector machine analyses were performed using scikit-learn version 0.17.1 (27), implemented in Python (version 3.5.2). Linear support vector classification was used to investigate the ability of all the tract measures combined to classify VP and FT children. We used all tract measures as features; hence these analyses were restricted to children who had data for all tracts (116 VP children and 28 FT children). All feature measurements were normalised by demeaning and scaling to unit variance prior to model training. Age at MRI, sex and intracranial volume were regressed from all the features prior to using them in the classification model. We tested internal model generalisation ability using a five-fold stratified cross-validation with an 80%/20% train/test split. Stratified folds ensure that the proportions of VP and FT participants in the training and test subsets in each fold are approximately the same as the proportions of VP and FT participants in the entire cohort, and no participants are present in multiple folds. The linear support vector classification model adopted in this study used  $L_2$  penalty, squared-hinge loss functions and a balanced class weighting given the unbalanced number of VP and FT participants. The best value of the tuning parameter C, which controls the trade-off between training data classification accuracy and generalisability, was chosen according to a nested three-fold cross-validation loop within each training fold using a grid search with values of  $10^{-3}$ ,  $10^{-2}$ ,  $10^{-1}$ ,  $10^0$ ,  $10^1$ ,  $10^2$ ,  $10^3$ .

Accuracy and balanced accuracy scores were calculated for each fold with a final estimate of model accuracy given by the average across the five folds. We calculated balanced accuracy as the proportions of VP and FT children were not equal. Balanced accuracy is the average of the true positive rate (proportion of VP correctly predicted) and true negative rate (proportion of FT correctly predicted), and should therefore better reflect the accuracy across both groups (28). To test whether the accuracy scores were significant, we performed permutation testing, with 5000 permutations.

We performed support vector regression to investigate the ability of the tract measures to predict episodic memory and emotional outcomes in the VP and FT children. Features were standardised as above, and five-fold cross-validation was used to evaluate model generalisation ability. The best value of C was chosen from the values of  $10^0$ ,  $10^1$ ,  $10^2$  using three-fold nested cross-validation. Model accuracy was determined using the Pearson’s

correlation between true episodic memory or emotional scores and predicted scores averaged over the five folds. Model significance was tested using 5000 permutations.

## Results

### Participant characteristics

Of the 159 VP and 36 FT children who had MRI, we excluded 18 children (15 VP, 3 FT). This was because the diffusion MRI acquisitions of these children were not completed, were incorrect and/or were affected by motion artefact, for the *b1200* ( $n=3$ ), *b3000* ( $n=13$ ), or both the *b1200* and *b3000* ( $n=2$ ) sequences. This meant we generated data (tractography, diffusion tensor and NODDI data) for a maximum of 144 VP and 33 FT participants. All of these participants had cingulum and medial forebrain bundle data. We excluded additional participants' data for other tracts, because certain image artefacts and brain structural abnormalities are not uniform across the brain and affected certain tracts more than others (for example, cardiac pulsation artefact affected the fornix due to its proximity to the ventricles), and certain tracts are more difficult to delineate than others due to unique characteristics of each tract such as size, thickness, and curvature (29). Of the 144 VP and 33 FT participants with cingulum and medial forebrain bundle data, we excluded 22 (20 VP, 2 FT) children for the fornix, 12 (9 VP, 3 FT) children for the uncinate fasciculus, and five (5 VP, 0 FT) children for the anterior thalamic radiation (some of these excluded VP children overlapped). This left 144 VP and 33 FT children with data for the cingulum and medial forebrain bundle, 124 VP and 31 FT children with data for the fornix, 135 VP and 30 FT children with data for the uncinate fasciculus, and 139 VP and 33 FT children with data for the anterior thalamic radiation. All these children were included in each separate linear regression analysis of each tract measure. Because some of the VP children excluded for the fornix, uncinate fasciculus and anterior thalamic radiation overlapped, we were left with a total of 116 VP and 28 FT children who had complete data for all tracts, and these data were included in the machine learning analyses.

Rates of major brain injuries (cystic periventricular leukomalacia and intraventricular haemorrhage) were low in the VP group, although as expected the VP group had more medical complications than the FT group (Table 1). Age at MRI and the proportion of males and females were similar between groups. The VP group had poorer verbal and visual episodic memory scores, and trended towards having greater emotional symptoms, at age 7 years compared with the FT group (Table 1).

Fewer VP participants had postnatal corticosteroids, and more VP participants were from multiple births, compared with VP nonparticipants. Otherwise, characteristics were similar between participants and nonparticipants (data not shown).

### Very preterm children compared with full-term children

**Linear regression**—Bilaterally, the fornix had significantly higher MD, AD, RD and neurite orientation dispersion, and lower volume in the VP children compared with the FT children. Additionally, the left fornix had significantly lower neurite density in the VP children compared with the FT children (Figure 2). The left superior-lateral medial forebrain

bundle had significantly higher neurite orientation dispersion in the VP children compared with the FT children (Figure 2). Volume of the anterior thalamic radiation was significantly higher in VP children compared with FT children (Figure 2). There were no significant group-by-sex interactions.

**Support vector classification**—After regressing age at MRI, sex and intracranial volume from the features, the tract measures significantly predicted group status with a cross-validated mean (95% CI) accuracy of 72.9 (56.2, 89.5) %,  $p = 0.009$ , and a cross-validated mean (95% CI) balanced accuracy of 62.7 (37.7, 87.6) %,  $p = 0.009$ . Fewer FT children [16/28 (57.1%)] were correctly classified compared with VP children [103/116 (88.8%)].

### Relationships with neurodevelopmental outcomes

**Linear regression**—Higher FA and lower MD and RD in the left uncinate fasciculus, lower MD and RD in the left cingulum, lower MD and RD in the right superior-lateral medial forebrain bundle, and lower RD in the right anterior thalamic radiation were significantly associated with better verbal episodic memory and learning in both groups, independent of GA at birth and neonatal brain abnormality score (Figure 3). After adjustment for the potentially confounding variables and multiple comparison correction, associations between tract measures and visual episodic memory and learning and emotional symptoms were not significant (data not shown). All tract measure-outcome associations did not significantly differ between the VP and FT groups, except for the relationship between left superior-lateral medial forebrain bundle AD and emotional symptoms (stronger positive association in the FT group versus weaker negative relationship in the VP group; data not shown). All tract measure-outcome associations did not differ significantly between males and females.

**Support vector regression**—The support vector regression model using all tract measures did not significantly predict episodic memory and emotional scores in the VP and FT children (all  $p$ -values for correlations between true and predicted scores were  $>0.3$ ; Supplementary Figure 1).

## Discussion

### Effect of very preterm birth on white matter tracts

The structure of major white matter tracts traditionally related to episodic memory and emotion, the fornix and medial forebrain bundle, was altered in VP children compared with FT children at age 7 years. This could be caused by ischaemic or inflammatory insults, or secondary alterations in white matter development (31). Understanding the underlying cellular changes is challenging given the indirect nature of the diffusion measures. The NODDI measures used in the current study provided increased specificity to cellular properties (neurite density and orientation dispersion) compared with diffusion tensor measures (21). Our results are consistent with many previous studies that have reported volume and microstructural differences in various white matter tracts between preterm and FT children, including in the tracts examined in the current paper (6–9). Our findings for the



fornix are largely consistent with previous studies, which found lower fornix volume in preterm children compared with FT children (6, 7), as well as altered fornix microstructure in preterm children compared with FT children, although our differences were in the AD, RD, MD, neurite density and neurite orientation dispersion parameters, whereas the previous studies found no difference in RD, but differences in an alternative microstructure measure, hindrance modulated orientational anisotropy (HMOA) (6, 7). This difference in findings could be related to the differences in scanners, sequences, and image processing pipelines between studies. In line with our study, previous studies found no microstructural differences in the cingulum between preterm and FT groups (6, 7, 9), although in contrast to our study, previous studies found lower cingulum volume in the preterm group compared with the FT group (6, 7). There are also some conflicting findings in studies on the uncinate fasciculus. We found no significant differences in uncinate fasciculus volume and microstructure between VP and FT children, whereas other studies found lower FA in the uncinate fasciculus in preterm children compared with FT controls (8, 9). Additionally, a previous study found higher FA in the anterior thalamic radiation in preterm children compared with FT controls (9), whereas we found no microstructural differences in this tract between VP and FT children, although we did find higher anterior thalamic radiation volume in the VP group compared with the FT group. These differences in results could be related to differing analysis methods and sample characteristics between studies. In particular, all previous studies were based on older participants than the current sample of 7-year-old children (19-year-olds in (6), 30-year-olds in (7), 12-year-olds in (8), and 9–17-year-olds in (9)). Tracts such as the cingulum, uncinate fasciculus and anterior thalamic radiation demonstrate ongoing development over adolescence and young adulthood such as increasing myelination and axon density (32), and hence group differences in tract measures are likely to change over time, and developmental delays related to VP birth may become more apparent with age in these developmentally-sensitive tracts.

Our support vector classification model used tract data to significantly classify VP and FT children. This suggests that VP-born individuals have distinct white matter tract development from FT-born individuals, and this can be detected by machine learning techniques, even 7 years after birth. While this analysis was purely for research purposes because the GA of an individual is already known at birth, accurate classification is important as it demonstrates that the support vector machine model is able to predict based on the tract measures whether a child has white matter structure that is more indicative of a VP or FT individual (11). Our classification performance is similar to previous studies (11–13). These previous studies used grey matter data from structural MRI to achieve 86–93% accuracy in classifying preterm and FT 12–17-year-olds (11), and used functional MRI data to achieve >80% accuracy in classifying preterm and FT infants (12, 13). Hence, data from multiple MRI modalities and brain regions appear able to classify VP from FT children using machine-based learning. This success justifies further investigation into whether machine learning algorithms can be used to predict the neurodevelopmental outcomes of VP individuals.

## Relationship between white matter tracts and episodic memory and emotion

Although weaker than the differences in tract structure between the VP and FT children, the linear regression results provided some evidence for associations between the tracts and verbal episodic memory and learning outcome in VP and FT children, independent of several key perinatal risk factors known to influence long-term neurodevelopmental outcomes. Our findings provide further support that these tracts have a role in episodic memory (3), and that alterations to these tracts may contribute to episodic memory difficulties in VP children (6, 7). Interestingly, fornix microstructure differed between VP and FT children, but was not related to episodic memory outcome, whereas other tracts that did not differ between VP and FT children (cingulum and uncinate fasciculus) were related to episodic memory. Additionally, the medial forebrain bundle (superior-lateral aspect) microstructure both differed between VP and FT children and was related to episodic memory outcome. The mechanisms behind these associations can only be speculated, but may reflect importance of the cingulate and orbitofrontal cortices in episodic memory in VP and FT children. The cingulum, anterior thalamic radiation, uncinate fasciculus and superior-lateral medial forebrain bundle have direct connections to these regions, whereas the fornix does not (3, 33).

Despite linear regressions revealing some associations between the tract measures and verbal episodic memory outcome, support vector regression results suggested that the ability of the tract measures as a whole to predict concurrent episodic memory outcome is limited. This discrepancy is an important finding. The effect sizes from linear regression were modest, and there were few strong associations from linear regression compared with the number of tests performed. Additionally, the statistical power of neuroimaging studies in general tends to be low, which can lead to overestimated effect sizes (34). Consequently, the strength of the reported associations from linear regression may be upwardly biased. This could explain why these associations perform poorly when testing in small, out-of-sample datasets using support vector regression. Additionally, there may be little extra information to be gained by including all the tracts in one machine learning model, due to shared variance between the tracts, as investigated previously using principal component analysis (35). Unlike the current findings, previous machine learning studies have found that MRI data can predict neurodevelopmental outcomes of preterm children (14–16). These studies used whole brain structural and diffusion MRI data from neonates, which may have better predictive ability compared with our specific white matter tracts studied later in childhood. However, our aim was not to achieve better predictive ability than previous studies, but rather to expand the use of machine learning to improve our understanding of the effects of VP birth on tracts that are traditionally related to episodic memory and emotion in childhood.

## Strengths and limitations

Strengths of our study include the large sample of children followed up until age 7 years, and the use of advanced statistical analyses in addition to traditional regression analyses. However, loss of participants to follow up or not having useable MRI data meant that a maximum of 144 (64%) of the original cohort of VP children and 33 (72%) of the original cohort of FT children could be included in the current study, and numbers were even smaller for the support vector machine analyses, which required complete data from all participants.

This reduced our power to predict neurodevelopmental outcomes and limits the generalisability of our results. The number of VP children included in our study is generally similar or larger than other studies in the field, and the number of FT controls included in our study is generally similar to other studies in the field (6–9), however we acknowledge that we had fewer FT controls than VP children and that the unbalanced groups could affect the support vector classification. We attempted to address the unbalanced groups by performing balanced class weighting within the support vector classification, and calculating balanced accuracy, which is a better reflection of accuracy for both groups. Future studies could recruit more FT controls to further minimise any bias of unbalanced samples on classification. Additionally, while we used K-Fold cross-validation with a partially nested cross-validation for parameter tuning, future studies could use nested cross-validation to reduce bias on support vector machine performance estimates (36). Given our interest in examining the neural correlates of emotional outcomes, we selected the Emotional Symptoms subscale of the Strengths and Difficulties Questionnaire. However, we acknowledge this is a brief screening measure and more comprehensive measures of emotional outcome are required to replicate our findings. Many of the tracts delineated, particularly the fornix, are located close to the lateral ventricles, exacerbating the risk of cerebrospinal fluid contamination of the microstructural measures. We tested whether cerebrospinal fluid contamination could be influencing our group differences from linear regression by adjusting for cerebrospinal fluid volume obtained using FreeSurfer, and results were similar (data not shown). Future studies could improve tractography further by using advanced methods for estimating the fibre orientation distribution, such as Multi-Shell Multi-Tissue constrained spherical deconvolution or Single-Shell 3-Tissue constrained spherical deconvolution (37, 38).

## Conclusions

The volume and microstructure of some tracts traditionally related to episodic memory and emotion are altered in VP children compared with FT children. This altered structure can be detected by machine learning techniques. Specific tracts, including the cingulum, uncinate fasciculus, anterior thalamic radiation and medial forebrain bundle, were associated with episodic memory outcome in VP and FT children, independent of several important perinatal risk factors. Despite this, the ability of tract structure to predict the episodic memory outcome of individual children was limited.

This study contributes important information towards understanding the long-term neural changes following VP birth, and how these might contribute to adverse episodic memory outcomes. Studies applying machine learning methods are critical for improving knowledge in the field. More broadly, such studies are a necessary first step towards understanding whether MRI data can aid in the identification of individual VP children who are at risk of adverse neurodevelopmental outcomes, and who may benefit from interventions to improve their outcomes. Future studies using larger datasets acquired as early as possible and testing a broader range of machine learning algorithms would be worthwhile for advancing knowledge in the field.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

This research was conducted within the VIBeS and Developmental Imaging research groups, located at the Murdoch Children's Research Institute, as well as the Children's MRI Centre, located at the Royal Children's Hospital, Melbourne, Victoria, Australia. We thank members of the VIBeS and Developmental Imaging teams, including Marilyn Bear for recruitment of the children in this study, the Royal Children's Hospital Medical Imaging staff for their assistance and expertise in the collection of the MRI data included in this study, Wai Yen Loh for development of a pipeline to segment the subcortical brain regions that were used in this study, and the families and children for participating in this study.

### Financial support

This study was supported by the National Health and Medical Research Council of Australia (Project Grant No. 237117 to TEI and LWD; Project Grant No. 491209 to PJA, TEI, LWD; Centre of Research Excellence No. 1060733 to LWD, PJA, JLYC and DKT; Early Career Fellowship No. 1012236 to DKT and 1053787 to JLYC; Career Development Fellowship No. 1085754 to DKT and No. 1141354 to JLYC; Senior Research Fellowship No. 1081288 to PJA). This study was also supported by the National Institutes of Health (NIH; grant R01 HD05709801, P30 HD062171, and UL1TR000448 to TEI). Additional support came from The Royal Children's Hospital Foundation devoted to raising funds for research at The Royal Children's Hospital (grant no. RCH1000 to JYMY), as well as the Murdoch Children's Research Institute, the Royal Children's Hospital, Department of Paediatrics at The University of Melbourne, the Victorian Government's Operational Infrastructure Support Program, and the United Cerebral Palsy Foundation, The United States (TEI). The funding sources had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

## References

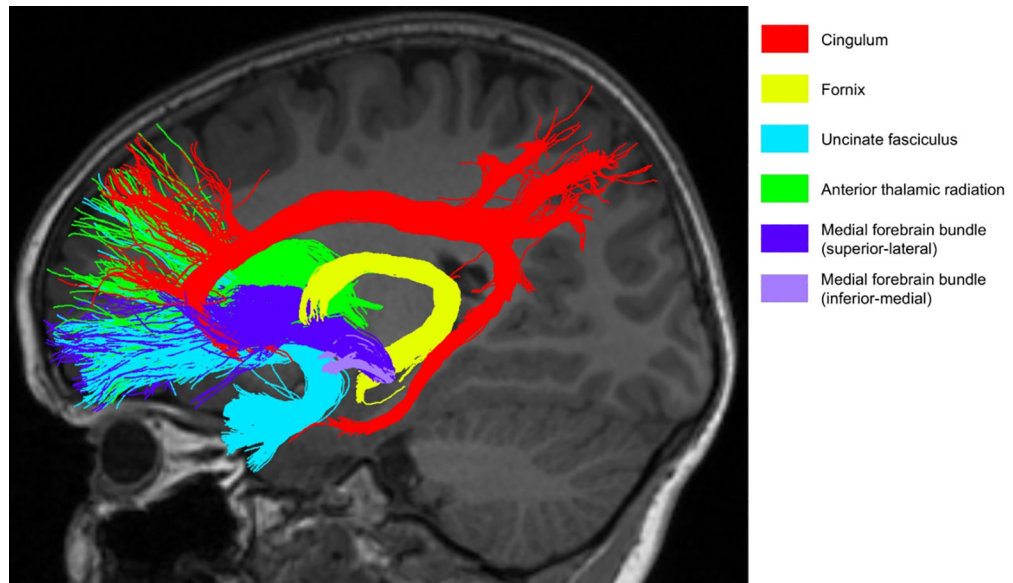
1. Omizzolo C, et al. Neonatal brain abnormalities and memory and learning outcomes at 7 years in children born very preterm. *Memory*. 22, 605–615 (2014). [PubMed: 23805915]
2. Treyvaud K, et al. Social-emotional difficulties in very preterm and term 2 year olds predict specific social-emotional problems at the age of 5 years. *J Pediatr Psychol*. 37, 779–785 (2012). [PubMed: 22408055]
3. Catani M, Dell'acqua F and Thiebaut de Schotten M. A revised limbic system model for memory, emotion and behaviour. *Neurosci Biobehav Rev*. 37, 1724–1737 (2013). [PubMed: 23850593]
4. Thompson DK, et al. Neonate hippocampal volumes: prematurity, perinatal predictors, and 2-year outcome. *Ann Neurol*. 63, 642–651 (2008). [PubMed: 18384167]
5. Omizzolo C, et al. Hippocampal volume and memory and learning outcomes at 7 years in children born very preterm. *J Int Neuropsychol Soc*. 19, 1065–1075 (2013). [PubMed: 23947431]
6. Caldinelli C, et al. White matter alterations to cingulum and fornix following very preterm birth and their relationship with cognitive functions. *Neuroimage*. 150, 373–382 (2017). [PubMed: 28216430]
7. Tseng CJ, et al. A multimodal imaging study of recognition memory in very preterm born adults. *Hum Brain Mapp*. 38, 644–655 (2017). [PubMed: 27647705]
8. Constable RT, et al. Prematurely born children demonstrate white matter microstructural differences at 12 years of age, relative to term control subjects: an investigation of group and gender effects. *Pediatrics*. 121, 306–316 (2008). [PubMed: 18245422]
9. Travis KE, Adams JN, Ben-Shachar M and Feldman HM. Decreased and Increased Anisotropy along Major Cerebral White Matter Tracts in Preterm Children and Adolescents. *PLoS One* 10, e0142860 (2015). [PubMed: 26560745]
10. Hastie T, Tibshirani R and Friedman JJH. *The elements of statistical learning, volume 1.* (Springer, New York, 2001).
11. Chu C, Lagercrantz H, Forssberg H and Nagy Z. Investigating the use of support vector machine classification on structural brain images of preterm-born teenagers as a biological marker. *PLoS One*. 10, e0123108 (2015). [PubMed: 25837791]

12. Ball G, et al. Machine-learning to characterise neonatal functional connectivity in the preterm brain. *Neuroimage*. 124, 267–275 (2016). [PubMed: 26341027]
13. Smyser CD, et al. Prediction of brain maturity in infants using machine-learning algorithms. *Neuroimage*. 136, 1–9 (2016). [PubMed: 27179605]
14. Moeskops P, et al. Prediction of cognitive and motor outcome of preterm infants based on automatic quantitative descriptors from neonatal MR brain images. *Sci Rep*. 7, 2163 (2017). [PubMed: 28526882]
15. Ullman H, et al. Neonatal MRI is associated with future cognition and academic achievement in preterm children. *Brain*. 138, 3251–3262 (2015). [PubMed: 26329284]
16. Schadl K, et al. Prediction of cognitive and motor development in preterm children using exhaustive feature selection and cross-validation of near-term white matter microstructure. *Neuroimage Clin*. 17, 667–679 (2018). [PubMed: 29234600]
17. Murray AL, et al. White matter abnormalities and impaired attention abilities in children born very preterm. *Neuroimage*. 124, 75–84 (2016). [PubMed: 26318524]
18. Smyser TA, et al. Cortical Gray and Adjacent White Matter Demonstrate Synchronous Maturation in Very Preterm Infants. *Cereb Cortex*. 26, 3370–3378 (2016). [PubMed: 26209848]
19. Tourmier JD, Calamante F and Connelly A. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage*. 35, 1459–1472 (2007). [PubMed: 17379540]
20. Jenkinson M, Bannister P, Brady M and Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. 17, 825–841 (2002). [PubMed: 12377157]
21. Zhang H, Schneider T, Wheeler-Kingshott CA and Alexander DC. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*. 61, 1000–1016 (2012). [PubMed: 22484410]
22. Delis DC, Kramer JH, Kaplan E and Ober BA. California Verbal Learning Test — Children’s Version (CVLT-C). (Harcourt Assessment, 1994).
23. Cohen MJ. Children’s Memory Scale. (The Psychological Corporation, San Antonio, 1997).
24. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 38, 581–586 (1997). [PubMed: 9255702]
25. Kidokoro H, Neil JJ and Inder TE. New MR imaging assessment tool to define brain abnormalities in very preterm infants at term. *Am J Neuroradiol*. 34, 2208–2214 (2013). [PubMed: 23620070]
26. Bhutta AT, Cleves MA, Casey PH, Cradock MM and Anand KJ. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA*. 288, 728–737 (2002). [PubMed: 12169077]
27. Pedregosa F, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 12, 2825–2830 (2011).
28. Brodersen KH, Ong CS, Stephan KE and Buhmann JM. The Balanced Accuracy and Its Posterior Distribution. 20th International Conference on Pattern Recognition (ICPR), 2010. Istanbul, Turkey, (2010).
29. Wasserthal J, Neher P and Maier-Hein KH. Tract Seg - Fast and accurate white matter tract segmentation. *Neuroimage*. 183, 239–253 (2018). [PubMed: 30086412]
30. Papile LA, Burstein J, Burstein R and Koffler H. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. *J Pediatr*. 92, 529–534 (1978). [PubMed: 305471]
31. Volpe JJ. Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. *Lancet Neurol*. 8, 110–124 (2009). [PubMed: 19081519]
32. Lebel C and Beaulieu C. Longitudinal development of human brain wiring continues from childhood into adulthood. *J Neurosci*. 31, 10937–10947 (2011). [PubMed: 21795544]
33. Nosarti C and Froudust-Walsh S. Alterations in development of hippocampal and cortical memory mechanisms following very preterm birth. *Dev Med Child Neurol*. 58 (Suppl 4), 35–45 (2016). [PubMed: 27027606]

34. Button KS, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci.* 14, 365–376 (2013). [PubMed: 23571845]
35. Telford EJ, et al. A latent measure explains substantial variance in white matter microstructure across the newborn human brain. *Brain Struct Funct.* 222, 4023–4033 (2017). [PubMed: 28589258]
36. Vabalas A, Gowen E, Poliakoff E and Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS One.* 14, e0224365 (2019). [PubMed: 31697686]
37. Dhollander T and Connelly A. A novel iterative approach to reap the benefits of multi-tissue CSD from just single-shell (+b=0) diffusion MRI data. 24th International Society of Magnetic Resonance in Medicine Singapore, (2016).
38. Jeurissen B, Tournier JD, Dhollander T, Connelly A and Sijbers J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *Neuroimage.* 103, 411–426 (2014). [PubMed: 25109526]

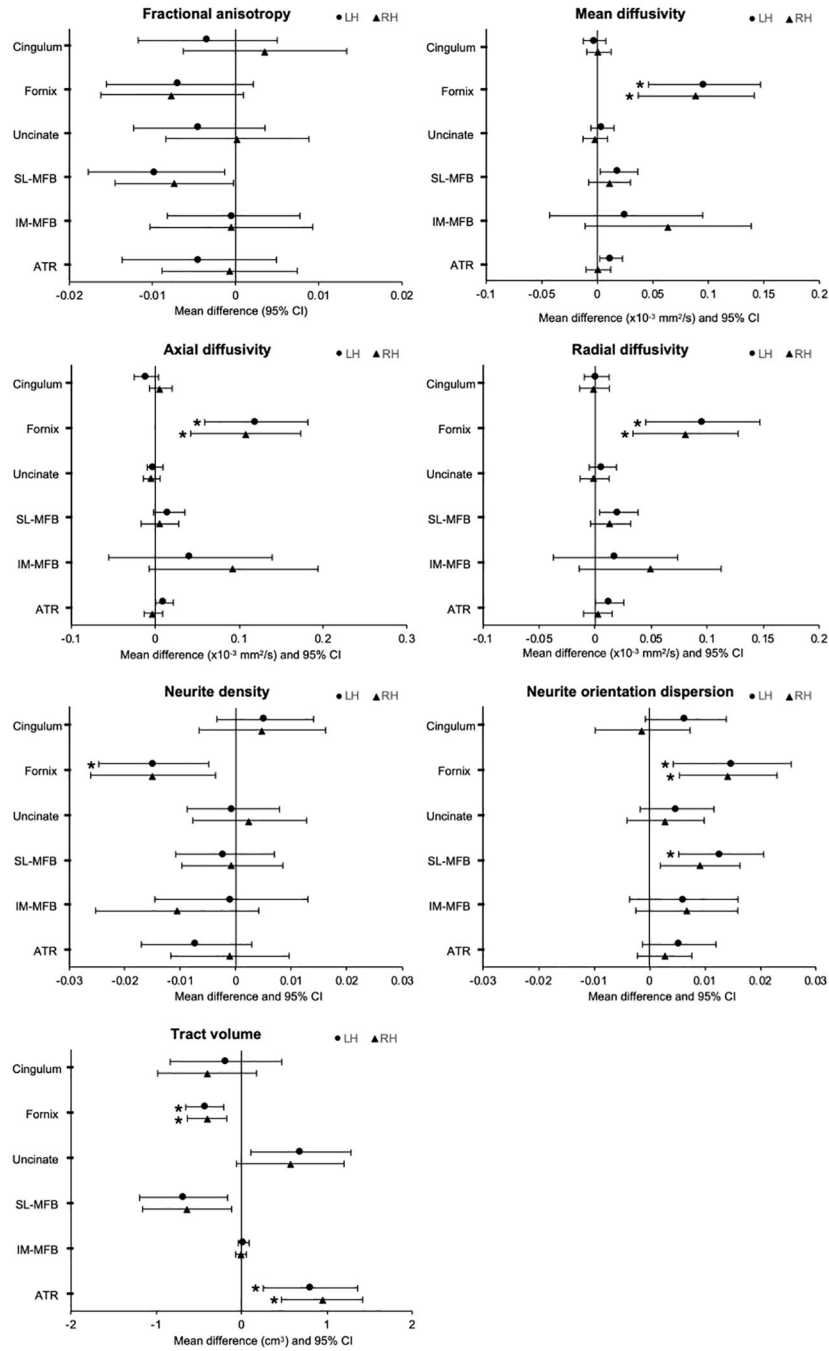
**Impact**

- We studied white matter fibre tracts thought to be involved in episodic memory and emotion in very preterm and full-term children using diffusion MRI and machine learning
- Very preterm children have altered fornix and medial forebrain bundle structure compared with full-term children
- Altered tract structure can be detected using machine learning, which accurately classified very preterm and full-term children using tract data
- Altered cingulum, uncinate fasciculus, medial forebrain bundle and anterior thalamic radiation structure was associated with poorer episodic memory skills using linear regression
- The ability of tract structure to predict episodic memory and emotional outcomes of individual children based on support vector regression was limited



**Figure 1.** Tractography reconstructions in one representative full-term child. Tracts are shown for the right hemisphere only. We have attempted to reconstruct most of the major tracts that are shown in Catani et al.'s diagram of the limbic system, and the colours of our tracts match those in Figure 8 in Catani et al (3).





**Figure 2. Comparisons of tract measures between very preterm and full-term children, from linear regression models.**

Points represent mean differences between groups, and error bars represent 95% CI. Circle points indicate analyses with left hemisphere (LH) tract measures; triangle points indicate analyses with right hemisphere (RH) tract measures. Analyses are adjusted for age at magnetic resonance imaging and sex (and intracranial volume for tract volume measures only). Asterisks indicate the mean difference was significant at  $p < 0.05$ , false discovery rate-corrected for the number of tracts and the number of tract measures. Units of axial, radial and mean diffusivities:  $\times 10^{-3} \text{ mm}^2/\text{s}$ ; units of tract volumes:  $\text{cm}^3$ . ATR= anterior thalamic

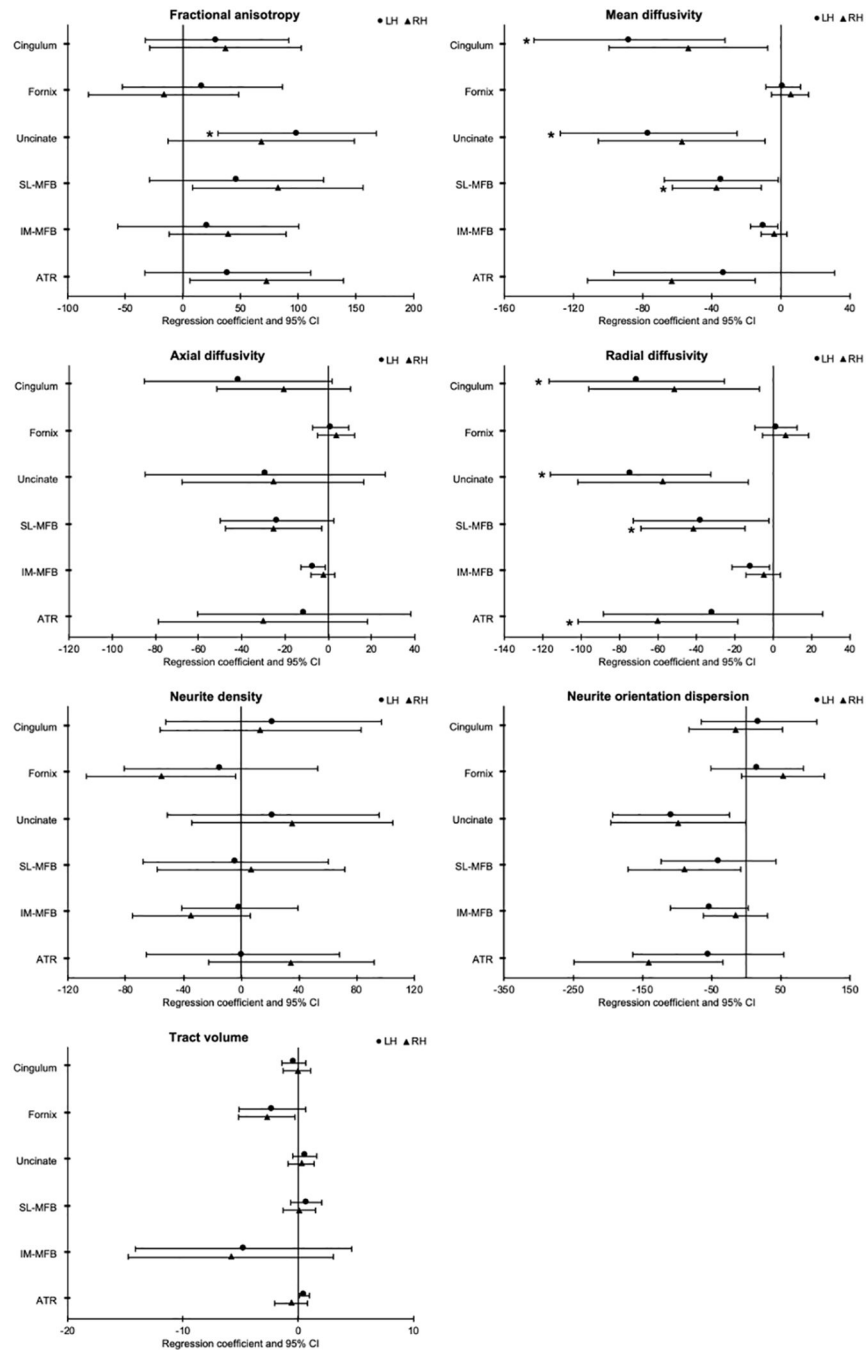
radiation; IM-MFB= inferior-medial medial forebrain bundle; SL-MFB= superior-lateral medial forebrain bundle.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 3. Associations between tract measures and verbal episodic memory in very preterm and full-term children, from linear regression models.**

Points are regression coefficients (change in the verbal episodic memory score per unit change in the tract measure), and error bars are 95% CI. Circle points indicate analyses with left hemisphere (LH) tract measures; triangle points indicate analyses with right hemisphere (RH) tract measures. Analyses are adjusted for age at neurodevelopmental assessment, sex, gestational age at birth, and neonatal brain abnormality score, as well as intracranial volume for the tract volume plot only. Asterisks indicate the association was significant at  $p < 0.05$ , false discovery rate-corrected for the number of tracts and the number of tract measures.

Units of axial, radial and mean diffusivities:  $\times 10^{-3}$  mm<sup>2</sup>/s; units of tract volumes: cm<sup>3</sup>.  
ATR= anterior thalamic radiation; IM-MFB= inferior-medial medial forebrain bundle; SL-MFB= superior-lateral medial forebrain bundle.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1.**

Perinatal medical characteristics and 7-year neurodevelopmental outcomes of the participants included in the current study, contrasted between the very preterm and full-term groups.

Characteristic	Very preterm, <i>n</i> =144	Full-term, <i>n</i> =33	Mean difference or odds ratio <sup>a</sup> (95% CI)	<i>P</i> -value
Perinatal characteristics <sup>b</sup>				
Gestational age at birth in weeks, mean (SD, min-max)	27.5 (1.9, 22–32)	38.9 (1.3, 37–41)	-11.33 (-12.01, -10.66)	NA
Birth weight in g, mean (SD, min-max)	977 (223, 414–1425)	3244 (501, 2390–4290)	-2267 (-2379, -2155)	NA
Birth weight SD score, mean (SD)	-0.50 (0.91)	0.01 (0.92)	-0.51 (-0.86, -0.16)	NA
Small for gestational age, <i>n</i> (%)	12 (8)	1 (3)	2.91 (0.36, 23.20)	0.313
Males, <i>n</i> (%)	70 (49)	16 (48)	0.99 (0.47, 2.12)	0.990
Cystic periventricular leukomalacia, <i>n</i> (%)	5 (3)	0 (0)	NA	NA
Intraventricular haemorrhage grade 3/4, <i>n</i> (%)	5 (3)	0 (0)	NA	NA
Antenatal corticosteroids, <i>n</i> (%)	126 (88) <sup>c</sup>	0 (0)	NA	NA
Postnatal corticosteroids, <i>n</i> (%)	8 (6) <sup>c</sup>	0 (0)	NA	NA
Bronchopulmonary dysplasia, <i>n</i> (%)	44 (31)	0 (0)	NA	NA
Infection, <i>n</i> (%)	46 (32)	1 (3)	15.02 (1.99, 113.34)	0.009
Multiple birth, <i>n</i> (%)	70 (49)	2 (6)	14.66 (3.38, 63.56)	<0.001
7-year characteristics				
Corrected age at MRI in years, mean (SD, min-max)	7.5 (0.2, 6.6–8.1)	7.6 (0.2, 7.2–8.0)	-0.06 (-0.16, 0.03)	0.155
Intracranial volume in cm <sup>3</sup> , mean (SD)	1397 (118)	1503 (103)	-106 (-150, -62)	<0.001
Corrected age at neurodevelopmental assessment in years, mean (SD)	7.5 (0.2)	7.6 (0.2)	-0.06 (-0.15, 0.03)	0.162
Full-scale IQ score <sup>d</sup> , mean (SD)	99.2 (12.9)	109.5 (11.4)	-10.26 (-15.09, -5.43)	<0.001
Full-scale IQ score < 85, <i>n</i> (%)	16 (11)	1 (3)	4 (0.5, 31.3)	0.187
Score for verbal memory/learning, mean (SD)	49.6 (10.3)	53.7 (12.8)	-4.14 (-8.25, -0.04)	0.048
Score for visual memory/learning, mean (SD)	9.6 (3.2)	11.4 (2.3)	-1.83 (-2.99, -0.68)	0.002
Score for emotional symptoms, mean (SD)	2.4 (2.3) <sup>e</sup>	1.6 (1.9) <sup>f</sup>	0.80 (-0.10, 1.71)	0.082

<sup>a</sup>When mean (SD) per group are reported, corresponding mean differences (95% CI) between groups are provided. When *n* (%) per group are reported, corresponding odds ratios (95% CI) are provided, which represent the odds of the outcome in the very preterm group relative to the full-term group.

<sup>b</sup>Perinatal data were collected by chart review. Birth weight SD score was calculated as each infant's weight relative to that expected for sex and gestational age using a British Growth Reference dataset. Small for gestational age was defined as more than two SD below the mean birthweight SD score. Bronchopulmonary dysplasia was defined as oxygen requirement at 36 weeks. Cystic periventricular leukomalacia and intraventricular haemorrhage were diagnosed from routine cranial ultrasound prior to term-equivalent age. Intraventricular haemorrhage was graded according to Papile et al (30). Infection was defined as sepsis and/or proven necrotising enterocolitis.

<sup>c</sup>*n*=1 with missing data.

<sup>d</sup>IQ was assessed using the Wechsler Abbreviated Scale of Intelligence.

<sup>e</sup>*n*=7 with missing data

$f_{n=3}$  with missing data.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript