Supplementary Information for: Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome

Authors: Kirstie J Whitaker*, Petra E Vértes*, Rafael Romero-Garcia, František Váša, Michael Moutoussis, Gita Prabhu, Nikolaus Weiskopf, Martina F Callaghan, Konrad S Wagstyl, Timothy Rittman, Roger Tait, Cinly Ooi, John Suckling, Becky Inkster, Peter Fonagy, Raymond J Dolan, Peter B Jones, Ian M Goodyer, the NSPN Consortium, Edward T Bullmore.

*These authors contributed equally to this work.

Contents

Supplementary Methods

Participant recruitment

2135 healthy young people in the age range 14-24 years were recruited from schools, colleges, NHS primary care services and direct advertisements in north London and Cambridgeshire. This primary cohort was stratified into 5 contiguous age-related strata: 14-15 years inclusive, 16-17 years, 18-19 years, 20-21 years, and 22-24 years. Recruitment within each stratum was evenly balanced for sex and ethnicity with the proportion of subjects ascribing their ethnicity as 'white' matched to within 10% of that in the local populations. All primary cohort participants satisfied the following eligibility criteria: aged between 14 and 24 years inclusive; able to understand written and spoken English; willing and able to give informed consent for recruitment into the study cohort and consent to be re-contacted directly for possible participation in future studies within the consortium. They were excluded if they were currently, or had recently (within the last 12 months) participated in a clinical trial of an investigational medical product. Participants completed questionnaire measures of socio-demographic status, family and educational or occupational environments, and sub-clinical psychopathology.

For the MRI study, 300 participants were sampled from the primary cohort, comprising N=60 participants (30 male and 30 female) in each of the 5 age strata defined for the primary cohort. All participants in this MRI cohort satisfied the following additional eligibility criteria: willing and able to give informed consent for participation in the study and to attend a full-day assessments at either the University of Cambridge or University College London study sites; have normal or corrected-to-normal vision. Participants were excluded if they were currently being treated for a psychiatric disorder or for drug or alcohol dependence; had a current or past history of neurological disorders or trauma including epilepsy, or head injury causing loss of consciousness; had a learning disability requiring specialist educational support and/or medical treatment; or had a safety contraindication on the MRI scanning checklist. Secondary cohort participants attended one of the study sites (University of Cambridge or University College London) for a day of clinical, cognitive and MRI assessments.

Multi-parametric mapping

The MPM sequence comprised three multi-echo 3D FLASH (fast low angle shot) scans, one radiofrequency (RF) transmit field map and one static magnetic (B0) field map scan (1–7). The 3 multi-echo FLASH scans were acquired with predominant T1-, proton density- (PD), and MT-weighting by appropriate choice of the repetition time (TR) and the flip angle α: TR/α = 18.7 ms/20° for the T1w scan and 23.7 ms/6° for the PDw and the MTw scans. MT-weighting was achieved by applying an off-resonance Gaussian-shaped RF pulse (4 ms duration, 220° nominal flip angle, 2 kHz frequency offset from water resonance) prior to the excitation. Multiple gradient echoes were acquired with alternating readout polarity at six equidistant echo times (TE) between 2.2 and 14.7 ms for the T1w and MTw acquisitions and at 8 equidistant TE between 2.2 ms and 19.7 ms for the PDw acquisition. Other acquisition parameters were: 1 mm

isotropic resolution, 176 sagittal partitions, field of view (FOV) = 256 × 240 mm, matrix = 256 × 240 ×176, parallel imaging using GRAPPA factor 2 in phase-encoding (PE) direction (AP), 6/8 partial Fourier in partition direction, non-selective RF excitation, readout bandwidth BW = 425 Hz/pixel, RF spoiling phase increment = 50°. The total acquisition time was ~25 min. Participants were given standard ear protection and instructed to lie still and rest during the scan.

Maps of the local RF transmit field were measured and estimated from a 3D EPI acquisition of spin and stimulated echoes (SE and STE) with different refocusing flip angles (6, 7). Imaging parameters were: 4 mm isotropic resolution, matrix = 64 × 48 × 48 and FOV = 256 mm × 192 mm × 192 mm along readout × PE × partition direction, parallel imaging using GRAPPA factor 2 × 2 in PE and partition direction, TE_{SE}/TE_{STE}/TM (mixing time)/TR = 37.06/37.06/31.2/500 ms, acquisition time 3 min. The flip angles of the SE/STE refocusing pulses were decreased from 230°/115° to 130°/65° and in steps of 10°/5°, acquisition time 3 min.

To correct the 3D EPI RF transmit field maps for geometric distortion and off-resonance effects, a map of the static magnetic field (B0) was acquired with the following parameters (6, 7): 2D double-echo FLASH sequence with 64 axial slices, slice thickness = 2 mm, inter-slice gap = 1 mm, TR = 1020 ms, TE1/TE2 = 10/12.46 ms, α = 90°, matrix = 64 × 64, FOV = 192 × 192 mm, right-left PE direction, BW = 260 Hz/pixel, flow compensation, acquisition time ~2 min.

MRI parameter estimation

The effective transverse relaxation rate $R2^*$ was estimated from the logarithm of the signal intensities of the 8 proton density weighted echoes using a linear regression against echo time. Next, arithmetic mean T1w, PDw and MTw images were calculated from the first 6 echoes of the respective multi-echo acquisitions, in order to increase the SNR (4). The resulting three mean images were used to calculate the parameter maps of the MT saturation, the apparent longitudinal relaxation rate R1 and the effective proton density PD* using previously developed models describing the image intensity of FLASH scans (1–3). Here we used only the R1 and MT quantitative maps.

Quantitative R1 maps were determined from the apparent R1 maps by correcting for local RF transmit field inhomogeneities and imperfect RF spoiling using the approach described by Preibsch and Deichmann (8), which was adapted to the FLASH acquisition parameters used here. RF transmit field maps were calculated from the 3D EPI acquisition and corrected for off-resonance effects as by Lutti et al (6).

The semi-quantitative MT saturation parameter is relatively robust against differences in relaxation times and RF transmit inhomogeneities—unlike the conventional MT ratio, which is affected by R1 and RF transmit field variations (3, 9). Additionally, small residual higher order dependencies of the MT saturation on the local RF transmit field were corrected using a semi-empirical approach, resulting in a corrected MT saturation value used in the further

processing: $MT_{corrected} = MT_{uncorrected} \cdot (1-0.4)(1-0.4 \cdot RF_{local})$. $MT_{uncorrected}$ is the original MT value and RF_{local} the relative local flip angle compared to the nominal flip angle.

Cortical depth estimations of magnetisation transfer

Cortical thickness varies from around 1mm-4mm while our imaging data was sampled at 1mm isotropic resolution. Ten intra-cortical fractional depth projections therefore samples the cortex every 0.1-0.4mm. Surface-based interpolation of voxel intensities, coupled with surface based smoothing techniques have been demonstrated to be sensitive to sub-voxel resolution information (10) but nevertheless there are limitations. To ensure sensitivity to changes in thicker regions of cortex, with the constraints of 1mm isotropic voxels, we chose to have 10 cortical samples.

Our intra-cortical samples were chosen as fractions of cortical depth to adjust for differing cortical thicknesses between regions, while infra-cortical locations are chosen at absolute depths to ensure consistency across regions. 0.4mm and 0.8mm reflect two samples that are approximately equal to the spacing used for intra-cortical samples in the thickest cortical regions. We did not sample further than 1mm into white matter as the projections may end up sampling nearby cortex (particularly when measured from the gyral walls) or possibly subcortical gray matter (when measured from the bottom of a sulcus).

Gene expression sample collections and processing

The Allen Human Brain Atlas is a publicly available online resource of microarray-based gene expression profiles for an anatomically comprehensive set of brain regions (11). The atlas is based on post-mortem tissue from 6 donors with no known history of neuropathological or neuropsychiatric disease, who also passed a set of serology, toxicology and RNA quality screens. The donors were a 24-year-old African American male (H0351.2001), a 39-year-old African American male (H0351.2002), a 57-year old Caucasian male (H0351.1009), a 31-year old Caucasian male (H0351.1012), a 49-year old Hispanic female (H0351.1015) and a 55-year old Caucasian male (H0351.1016).

Each brain was cut into (0.5-1.0 cm thick) and frozen. Slabs were sectioned to allow expert neuroanatomic annotation, delineation and sampling. RNA was isolated and microarray data were generated for about 500 samples per hemisphere, representing all anatomical structures in approximate proportion to their volume. Each sample is then associated with expression levels for about 60,000 gene probes with 93% of known genes represented by at least 2 probes (see technical white paper at http://human.brain-map.org).

MRI and AIBS data matching

We used the Maybrain package (https://github.com/rittman/maybrain) to match the centroids of the MRI regions to the closest Allen Brain Institute for Brain Science (AIBS) sample in any of the 6 donors. We then determined the anatomical structure that this sample was located in. Expression data were averaged across all samples from all donors in the matching anatomical

structure, across both hemispheres. This is necessary as only two of the four donors have bilateral data (due to similarity of gene expression between hemispheres). Prior to averaging across subjects, two MRI regions were excluded as both the mean and the range of gene expression values in these regions were outliers compared to the other cortical regions of interest. Expression data for each probe were then normalised by taking the z-score for that probe across the remaining 306 regions in each subject.

The data were also averaged across probes corresponding to the same gene. We excluded probes that were not matched to gene symbols in the AIBS data. The final output of Maybrain is X, a 306x20737 matrix of whole-genome gene expression data for the 306 remaining MRI regions of interest. These data are made available in the **DATA.zip** compressed directory within the project's figshare repository as *PLS_gene_predictor_vars.csv* (see supplementary note on accessing data, below).

Gene ontology enrichment analysis

We used the online tool GOrilla (http://cbl-gorilla.cs.technion.ac.il, version 28 Nov 2015) to search for enriched GO terms that appear densely at the top of the ranked list of genes in the PLS components. For these analyses we uncheck the "Run GOrilla in fast mode" option and use the "P-value threshold 10-4" in the advanced parameter settings in order to best approximate FDR correction with $\alpha = 0.05$.

We used REViGO (http://revigo.irb.hr) to summarise the list of significant GO terms by identifying redundant terms (for example discarding "ion transmembrane transport" when we had kept the more specific term "potassium ion transmembrane transport").

The list of significant enrichment terms for both the most positively weighted (up-regulated) genes and the most negatively weighted (down-regulated) genes can be found in the supplementary file *WhitakerVertes_PLSEnrichmentGeneList.xlsx* (see supplementary note on accessing data, below). The enrichment of down-regulated genes was obtained by providing the inverse ranking of genes to the GOrilla software tool. Redundant terms are shaded in grey in the supplementary file, to highlight the most meaningful GO annotations. We also shaded out terms annotated to over 1000 genes given their generality (for example "cell communication").

Although the main text reports PLS and enrichment results based on the full cohort (297 subjects with brain scans) we also examined the robustness of the PLS and enrichment analysis by repeating it for the discovery and validation cohorts (N=100 and N=197) separately. As expected, we find that the significantly enriched terms reported for the complete dataset are replicated in the sub-samples. However, these analyses (especially in the smaller discovery cohort) also yield additional enrichment terms which do not replicate across cohorts and which we therefore do not report. The full PLS and enrichment results for all three cohorts are provided in the supplementary file *WhitakerVertes_PLSEnrichmentGeneList.xlsx*. In addition, we provide

a detailed visualization of all significantly enriched biological processes embedded in the hierarchical tree of GO terms for PLS2 (see supplementary note on accessing data, below).

Structural covariance and network analyses

We created association matrices for cortical thickness by correlating all participants' values pairwise, yielding a 308 x 308 matrix. Graphs were then constructed using NetworkX (12) and a minimum spanning tree was used as the backbone to the graphs in order to ensure no nodes would be disconnected from the network. Graphs were thresholded at densities between 5% and 30% at 5% intervals, maintaining only the strongest connections, and the remaining connections were set to have equal weightings.

The following nodal topological measures were calculated:

- Degree: In binarised graphs, the degree of a node is the number of edges connecting it to the rest of the network. Higher degree values indicate that a node is more highly connected to other nodes within the graph.

- Closeness: The closeness centrality for each node *u* is the reciprocal of the sum of the shortest path distances from u to all other nodes. Since the sum of distances depends on the number of nodes in the graph (*n*), closeness is normalised by the sum of minimum possible distances n−1. Higher values of closeness indicate higher centrality.

- Connection distance: The average connection distance for each node was calculated as the mean Euclidean distance in standard space for all non-zero edges connecting to the node.

The following global topological measures were calculated:

- Average clustering: The nodal clustering coefficient is the fraction of edges linking the neighbours of a node out of all possible edges between them. The overall clustering coefficient C(G) is defined as the average clustering coefficient of all nodes in graph G.

- Global efficiency: The shortest path length L_{uv} between a pair of nodes *u* and *v* is defined as the minimum number of edges that need to be traversed to get from node *u* to node *v*. The average path length, L(G), is the sum of these values for each node, and global efficiency, E(G), is the average inverse path length for each node.

- Assortativity: The assortativity coefficient, A(G), is the Pearson correlation coefficient of degree between pairs of linked nodes. Positive values of assortativity indicate that nodes are likely to connect to nodes with similar number of connections

(degree) to themselves and the graph is said to be assortative, while negative values indicate that nodes are likely to connect to nodes with very different number of connections to themselves and the graph is said to be disassortative.

- Modularity: Many complex networks have a modular community structure, whereby they contain subsets of highly interconnected nodes called modules. The modularity, Q(G), of a graph quantifies the quality of a suggested partition of the network into modules by measuring the fraction of the network's edges that fall inside modules compared with the expected value of this fraction if edges were distributed at random (13). The maximum value, M(G), of the modularity found for any partition of a given graph into modules yields a measure of the degree of modularity of the network, compared with random networks. Community detection was conducted by implementing the louvain method described by Blondel and colleagues (14) through the community python module (http://perso.crans.org/aynaud/communities/).

- Small-world coefficient: The small-world coefficient is defined as the ratio of the clustering and path length of a given network to an equivalent random network with same degree distribution (15). Networks with a high small world coefficient will have high levels of local clustering among nodes of a network and short paths that globally link all nodes of the network meaning all nodes are linked through relatively few intermediate steps, despite the fact that most nodes maintain only a few direct connections. Random networks are not small-world as they have low local clustering, while lattice networks are also not small-world due to their long path length.

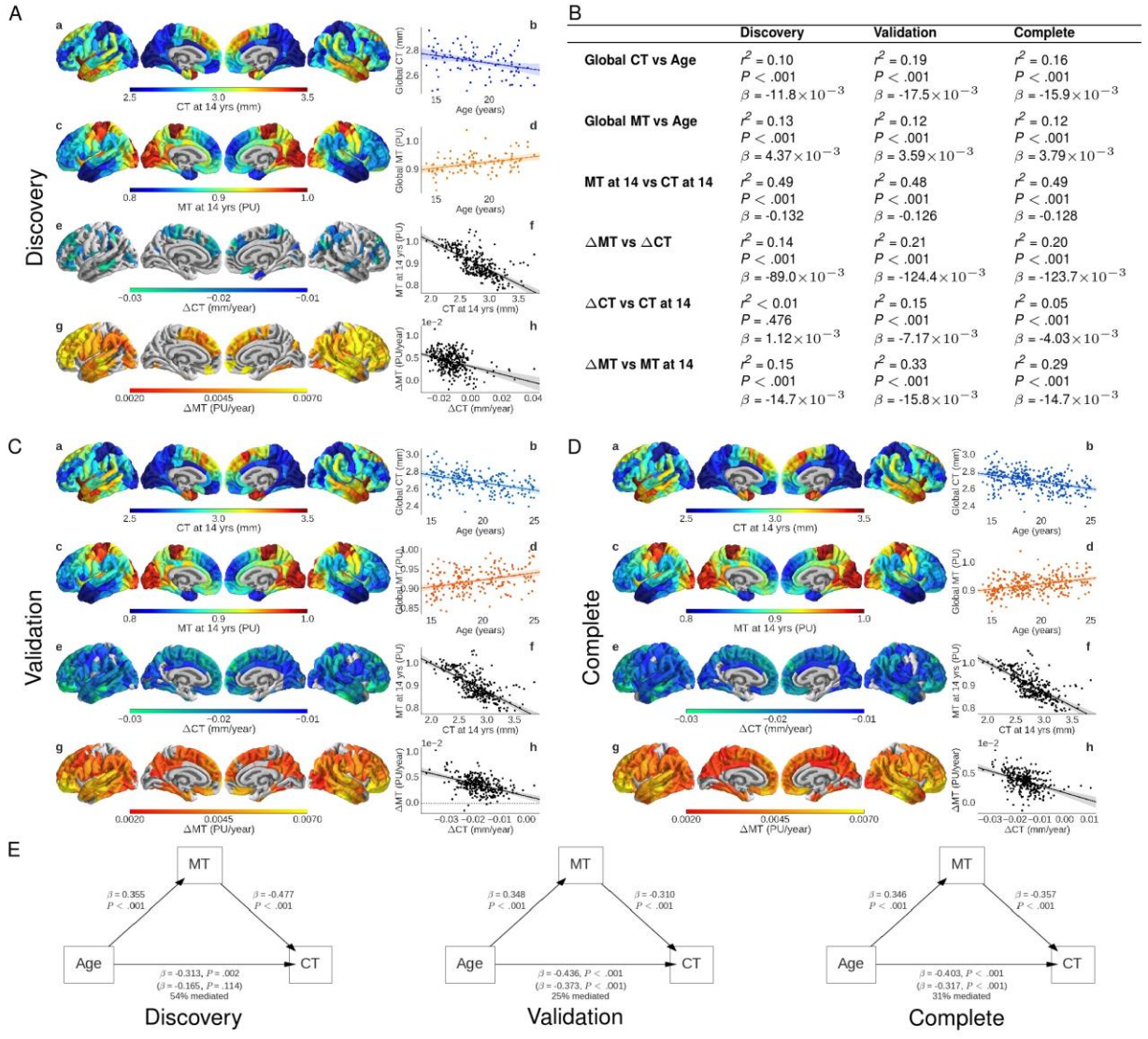
Finally, the rich club coefficient was also calculated for each graph. The rich club coefficient is the ratio, for every degree *k*, of the number of actual to the number of potential edges for nodes with degree greater than k. We chose to define "rich" nodes as those which had degree greater than the point at which the rich club coefficient passed 0.8.

All topological measures were also calculated for 1000 random graphs with the same number of nodes and edges and preserved degree distribution.

Supplementary Figures

Figure S1

Independent replication of adolescent cortical shrinkage and intra-cortical myelination, associated statistics, and mediation analyses. Panels **A**, **C** and **D** are replications of **Fig. 1** for the discovery, validation and complete cohorts. These confirm the anti-correlated regional patterns of baseline CT and MT as well as adolescent shrinkage (ΔCT) and intra-cortical myelination (ΔMT). Panel **B** shows effect sizes, explained variance and significance levels for each of the scatter plots. **E**: Global measures of magnetisation transfer at 70% cortical depth partially mediate the correlations between with age and global cortical thickness for the discovery, validation and complete cohorts.



	Discovery	Validation	Complete
Global CT vs Age	r^2 = 0.10 P < .001 β = -11.8×10^{-3}	r^2 = 0.19 P < .001 β = -17.5×10^{-3}	r^2 = 0.16 P < .001 β = -15.9×10^{-3}
Global MT vs Age	r^2 = 0.13 P < .001 β = 4.37×10^{-3}	r^2 = 0.12 P < .001 β = 3.59×10^{-3}	r^2 = 0.12 P < .001 β = 3.79×10^{-3}
MT at 14 vs CT at 14	r^2 = 0.49 P < .001 β = -0.132	r^2 = 0.48 P < .001 β = -0.126	r^2 = 0.49 P < .001 β = -0.128
ΔMT vs ΔCT	r^2 = 0.14 P < .001 β = -89.0×10^{-3}	r^2 = 0.21 P < .001 β = -124.4×10^{-3}	r^2 = 0.20 P < .001 β = -123.7×10^{-3}
ΔCT vs CT at 14	r^2 < 0.01 P = .476 β = 1.12×10^{-3}	r^2 = 0.15 P < .001 β = -7.17×10^{-3}	r^2 = 0.05 P < .001 β = -4.03×10^{-3}
ΔMT vs MT at 14	r^2 = 0.15 P < .001 β = -14.7×10^{-3}	r^2 = 0.33 P < .001 β = -15.8×10^{-3}	r^2 = 0.29 P < .001 β = -14.7×10^{-3}

Figure S2

Cortical parcellation and independent replication of intra-cortical magnetisation transfer mapping. A: Anatomical location of the 308 cortical regions generated by parcellating the cortical surface within anatomical regions of the Desikan-Killiany atlas, ensuring that all regions have the same surface area. Panels **B**, **C** and **D** are replications of **Fig. 2b-e** for the discovery, validation and complete cohorts. These confirm that baseline MT increases monotonically with distance from the pial surface while ΔMT is greatest at 70% cortical depth. The correlation between baseline MT and myelin basic protein (*MBP*) is also greatest at this depth. Boxes represent the median and interquartile range over all 308 regions at each depth.

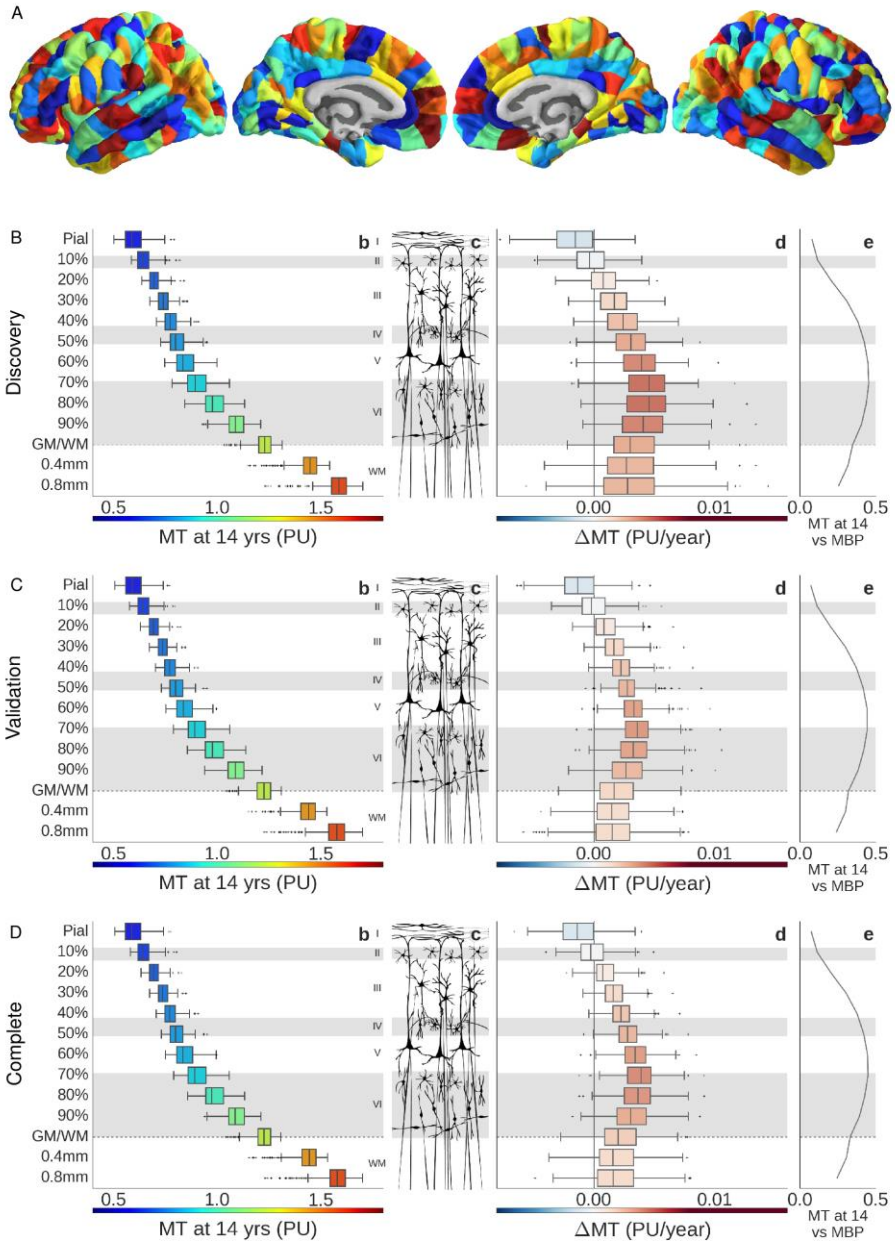
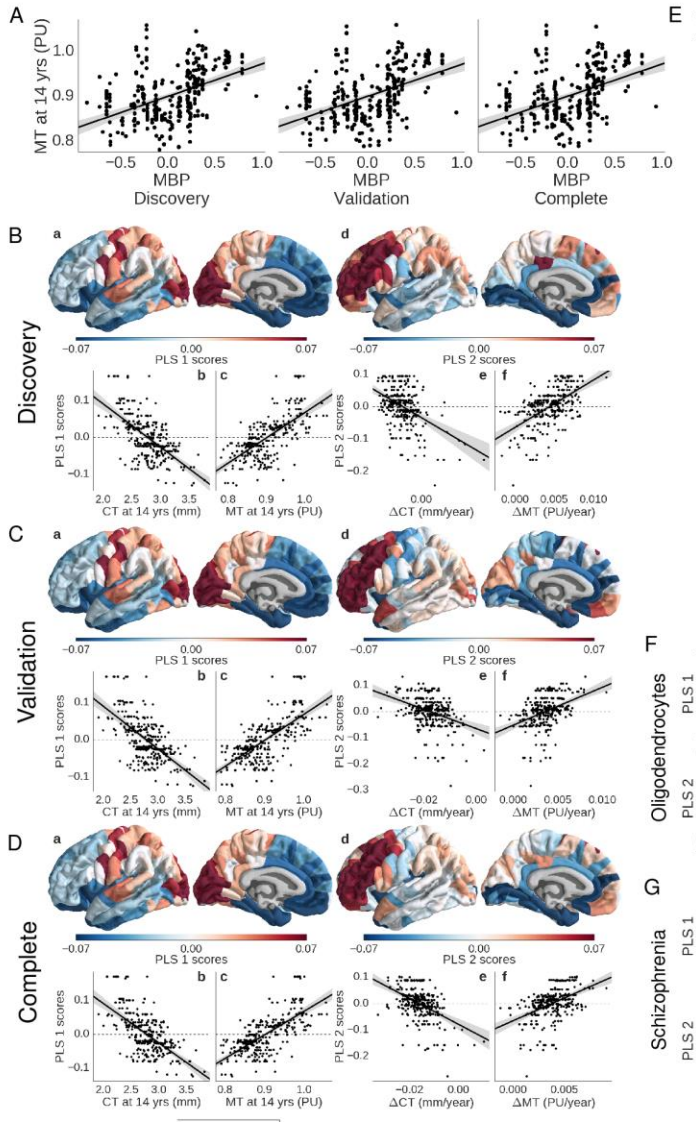


Figure S3

Independent replication of association between gene expression profiles and cortical thickness and magnetisation transfer and associated statistics. A: Baseline MT at 70% cortical depth is correlated with myelin basic protein expression. Panels **B**, **C** and **D** are replications of **Fig. 3a-f** for the discovery, validation and complete cohorts. Panel **E** shows effect sizes, explained variance and significance levels for correlations between regional scores on gene-based PLS components 1 and 2 and MRI measures. Oligodendrocyte-related genes are over expressed in PLS2 (**F**, bottom row) and schizophrenia risk genes are significantly over- or under-expressed (**G**, bottom row) compared to 10,000 randomly drawn gene sets of the same size in each of the discovery, validation and complete cohorts. Blue line indicates the 95th percentile value of the random gene sets.



E

	Discovery	Validation	Complete
PLS1 vs CT at 14	r^2 = 0.39 P < .001 β = -0.118	r^2 = 0.40 P < .001 β = -0.114	r^2 = 0.40 P < .001 β = -0.116
PLS1 vs ΔCT	r^2 = 0.03 P = .006 β = -1.04	r^2 = 0.07 P < .001 β = 2.71	r^2 = 0.02 P = .020 β = 1.38
PLS1 vs MT at 14	r^2 = 0.46 P < .001 β = 0.683	r^2 = 0.48 P < .001 β = 0.694	r^2 = 0.48 P < .001 β = 0.697
PLS1 vs ΔMT	r^2 = 0.01 P = .046 β = -3.1	r^2 = 0.12 P < .001 β = -12.8	r^2 = 0.09 P < .001 β = -11.3
PLS2 vs CT at 14	r^2 < 0.01 P = .146 β = 0.016	r^2 < 0.01 P = .442 β = 0.008	r^2 < 0.01 P = .472 β = 0.008
PLS2 vs ΔCT	r^2 = 0.20 P < .001 β = -2.82	r^2 = 0.14 P < .001 β = -3.68	r^2 = 0.23 P < .001 β = -4.90
PLS2 vs MT at 14	r^2 = 0.03 P < .001 β = -0.180	r^2 = 0.06 P < .001 β = -0.251	r^2 = 0.04 P < .001 β = -0.200
PLS2 vs ΔMT	r^2 = 0.30 P < .001 β = 14.7	r^2 = 0.16 P < .001 β = 14.5	r^2 = 0.22 P < .001 β = 17.2

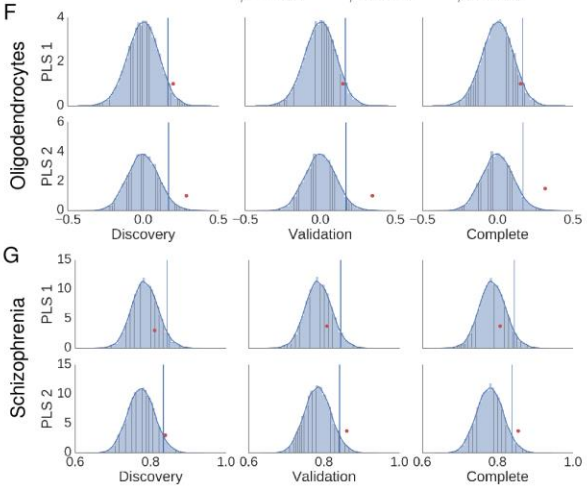
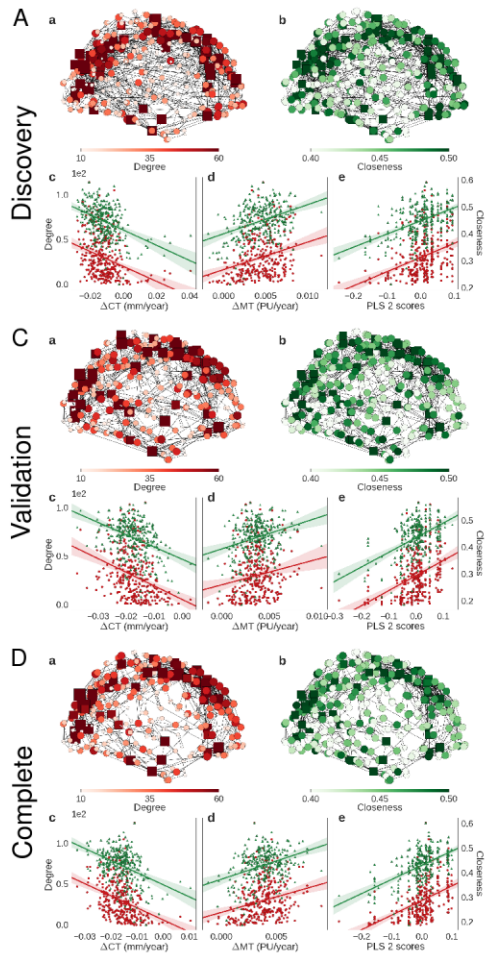


Figure S4

Structural covariance network and independent replication of association between hubs of the human brain connectome and MRI changes as well as gene expression profiles of adolescent cortical consolidation. A, **C** and **D**: Replication of **Fig 4a-f** for discovery, validation and complete cohorts, confirming that adolescent brain changes occur predominantly in high-degree and closeness hubs of the human brain connectome with associated statistics (**B**).



B

	Discovery	Validation	Complete
CT at 14 vs Degree	$r^2 < 0.01$ $P = .574$ $\beta = -2.64$	$r^2 < 0.01$ $P = .480$ $\beta = 3.02$	$r^2 < 0.01$ $P = .892$ $\beta = -0.48$
CT at 14 vs Closeness	$r^2 < 0.01$ $P = .542$ $\beta = -6.75 \times 10^{-3}$	$r^2 < 0.01$ $P = .128$ $\beta = 18.70 \times 10^{-3}$	$r^2 < 0.01$ $P = .732$ $\beta = 4.48 \times 10^{-3}$
△CT vs Degree	$r^2 = 0.11$ $P < .001$ $\beta = -0.83 \times 10^{3}$	$r^2 = 0.13$ $P < .001$ $\beta = -1.40 \times 10^{3}$	$r^2 = 0.14$ $P < .001$ $\beta = -1.48 \times 10^{3}$
△CT vs Closeness	$r^2 = 0.18$ $P < .001$ $\beta = -2.83$	$r^2 = 0.16$ $P < .001$ $\beta = -4.56$	$r^2 = 0.18$ $P < .001$ $\beta = -4.91$
MT at 14 vs Degree	$r^2 < 0.01$ $P = .788$ $\beta = 4.15$	$r^2 < 0.01$ $P = .236$ $\beta = -26.55$	$r^2 < 0.01$ $P = .516$ $\beta = -11.88$
MT at 14 vs Closeness	$r^2 < 0.01$ $P = .404$ $\beta = -46.3 \times 10^{-3}$	$r^2 = 0.01$ $P = .036$ $\beta = -134.9 \times 10^{-3}$	$r^2 < 0.01$ $P = .160$ $\beta = -94.9 \times 10^{-3}$
△MT vs Degree	$r^2 = 0.09$ $P < .001$ $\beta = 3.08 \times 10^{3}$	$r^2 = 0.03$ $P = .002$ $\beta = 2.64 \times 10^{3}$	$r^2 = 0.07$ $P < .001$ $\beta = 3.81 \times 10^{3}$
△MT vs Closeness	$r^2 = 0.15$ $P < .001$ $\beta = 10.8$	$r^2 = 0.07$ $P < .001$ $\beta = 11.6$	$r^2 = 0.13$ $P < .001$ $\beta = 14.8$
PLS1 vs Degree	$r^2 < 0.01$ $P = .446$ $\beta = 16.89$	$r^2 < 0.01$ $P = .612$ $\beta = -12.96$	$r^2 < 0.01$ $P = .922$ $\beta = -2.36$
PLS1 vs Closeness	$r^2 < 0.01$ $P = .288$ $\beta = 65.6 \times 10^{-3}$	$r^2 < 0.01$ $P = .592$ $\beta = -35.4 \times 10^{-3}$	$r^2 < 0.01$ $P = .894$ $\beta = 9.4 \times 10^{-3}$
PLS2 vs Degree	$r^2 = 0.16$ $P < .001$ $\beta = 158.3$	$r^2 = 0.15$ $P < .001$ $\beta = 154.1$	$r^2 = 0.21$ $P < .001$ $\beta = 178.6$
PLS2 vs Closeness	$r^2 = 0.22$ $P < .001$ $\beta = 487.2 \times 10^{-3}$	$r^2 = 0.22$ $P < .001$ $\beta = 543.1 \times 10^{-3}$	$r^2 = 0.26$ $P < .001$ $\beta = 572.7 \times 10^{-3}$

Figure S5

MRI measures and gene expression patterns show rostral-caudal gradients. Panels **A**, **B** and **C** show correlations between MRI measures of baseline CT and MT, change in CT and MT, and gene expression patterns PLS1 and PLS2 with each region's x, y and z coordinate for the discovery, validation and complete cohorts.

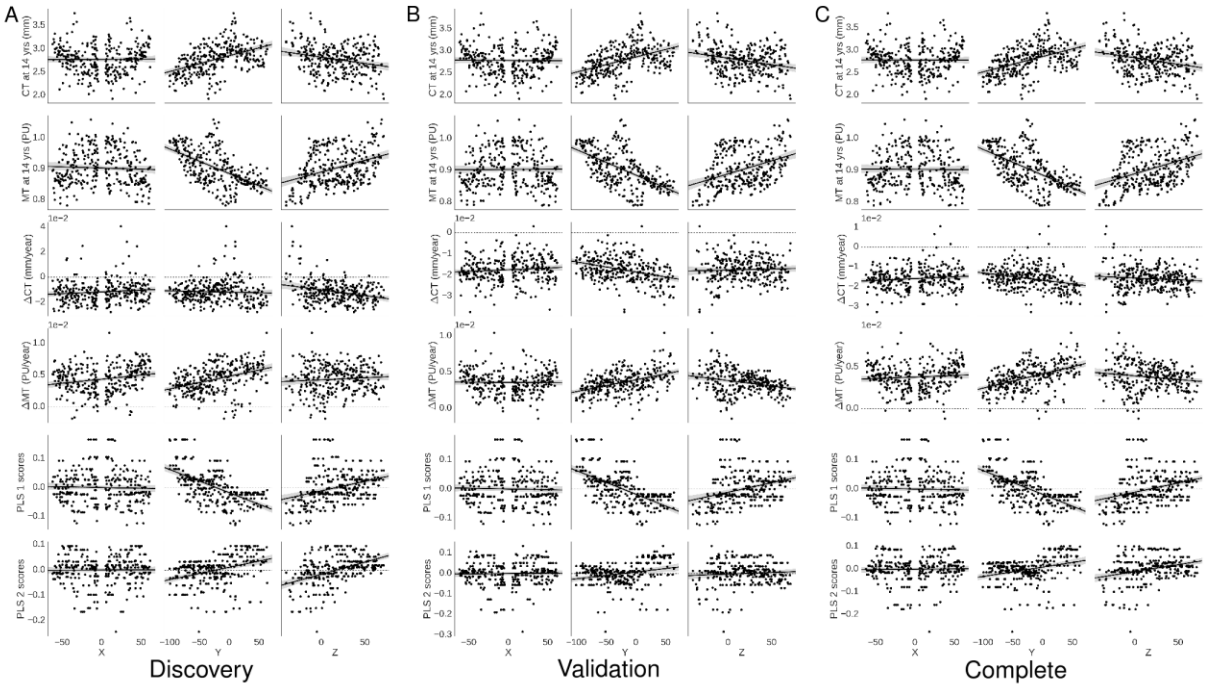
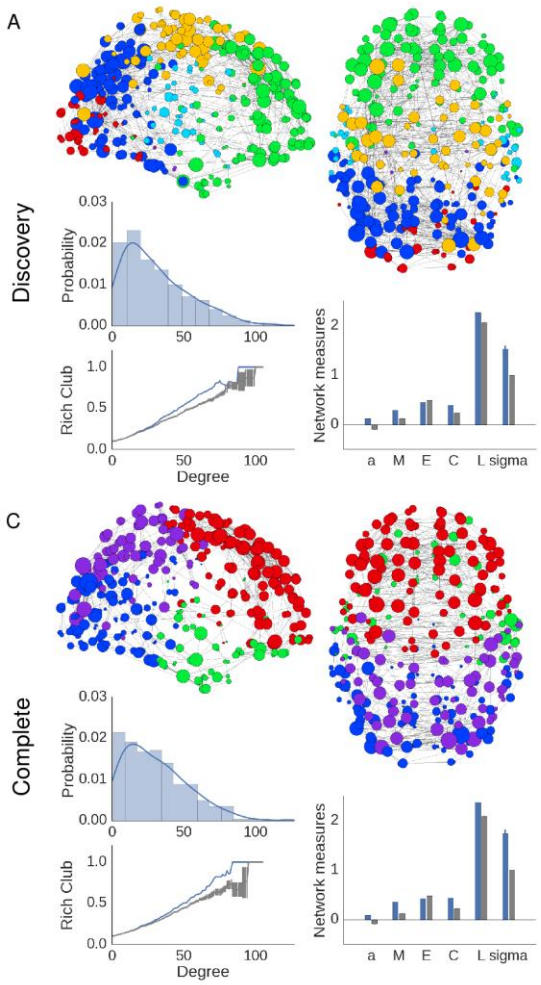


Figure S6

Structural covariance network and associated global measures. Panels **A**, **B** and **C** show that the structural covariance network has modular structure, heavy tailed degree distribution and rich club curve for the discovery, validation and complete cohorts. **D**: Global network measures for the structural covariance network compared to random networks that preserve degree distribution.



D

	Discovery	Validation	Complete
Assortativity	0.13 rand: -0.09 [-0.11, -0.07]	0.07 rand: -0.08 [-0.10, -0.06]	0.09 rand: -0.08 [-0.10, -0.06]
Clustering	0.40 rand: 0.24 [0.23, 0.25]	0.40 rand: 0.24 [0.23, 0.25]	0.45 rand: 0.23 [0.22, 0.24]
Efficiency	0.45 rand: 0.49 [0.49, 0.49]	0.44 rand: 0.48 [0.48, 0.48]	0.43 rand: 0.49 [0.49, 0.49]
Modularity	0.30 rand: 0.13 [0.12, 0.13]	0.34 rand: 0.12 [0.12, 0.13]	0.37 rand: 0.13 [0.12, 0.13]
Shortest Path	2.26 rand: 2.07 [2.06, 2.08]	2.35 rand: 2.10 [2.10, 2.11]	2.38 rand: 2.09 [2.08, 2.10]
Small World	1.53 [1.48, 1.58] rand: 1.00	1.50 [1.44, 1.57] rand: 1.00	1.74 [1.67, 1.80] rand: 1.00

Supplementary Table Legends

Table S1

Demographic information for the discovery and validation cohorts, along with the 2135 members of the Neuroscience in Psychiatry Network who were not invited for an MRI scan. For continuous measures age, intelligence quotient (IQ), handedness and index of multiple deprivation (IMD), median and interquartile range (IQR) measures are provided.

Participant Demographics

	Discovery	**Validation**	**Remaining 2K**
Number of participants	100	200	2135
Location	61.0% Cambridge	95.0% Cambridge	53.4% Cambridge
Gender	50.0% male	50.0% male	45.0% male
Age (years)	18.5 (IQR: 16.6-21.3)	18.6 (IQR: 16.9-21.6)	18.8 (IQR: 16.8-21.6)
IQ	111.5 (IQR: 103.8-121.0)	111.0 (IQR: 102.0-119.0)	NA
Handedness	76.2 (IQR: 51.5-100.0)	80.0 (IQR: nan-90.0)	NA
IMD	10.9 (IQR: 5.4-21.2)	8.2 (IQR: 4.4-11.8)	13.2 (IQR: 6.6-24.4)
Ethnicity	83.0% White 5.0% Asian 2.0% Black 6.0% Mixed 4.0% Other 0.0% Declined to state	83.5% White 4.5% Asian 2.5% Black 8.5% Mixed 0.0% Other 0.5% Declined to state	77.5% White 9.6% Asian 4.4% Black 5.8% Mixed 1.5% Other 0.4% Declined to state

Supplementary Movie Legends

Movie S1

Baseline magnetisation transfer (MT at age 14) for each region and at each fractional depth through cortex, then at five 0.2mm steps into the white matter underlying cortex. MT increases monotonically. This movie corresponds to **Fig. 2b** in the main manuscript and represents data from the discovery cohort. Movies that correspond to the validation and complete cohorts are available in the **VALIDATION.zip** and **COMPLETE.zip** compressed directories in the project's figshare respository, within the **PYSURFER_IMAGES/COVARS_ones/PNGS/MOVIES** subdirectory for each cohort.

Movie S2

Change in magnetisation transfer with age (ΔMT) for each region and at each fractional depth through cortex, then at five 0.2mm steps into the white matter underlying cortex. Change is greatest and most widespread at 70% cortical depth. This movie corresponds to **Fig. 2d** in the main manuscript and represents data from the discovery cohort. Movies that correspond to the validation and complete cohorts are available in the **VALIDATION.zip** and **COMPLETE.zip** compressed directories in the project's figshare respository, within the **PYSURFER_IMAGES/COVARS_ones/PNGS/MOVIES** subdirectory for each cohort.

Supplementary Note: Accessing data, analysis code and supplementary files

Data

All data to replicate the analyses presented here are available for download from the **DATA.zip** compressed directory contained within the project's figshare repository: NSPN: Adolescent consolidation of human connectome hubs.

DATA.zip download link: https://dx.doi.org/10.6084/m9.figshare.2057796

Specifically, **DATA.zip** contains:

- *DemographicData.csv*: Age, gender, study location (Cambridge or UCL), IQ (where available), handedness (where available), ethnicity and socioeconomic status data for 2436 participants in the NSPN 2k cohort.

- *[COHORT]/PARC_500aparc_[MEASURE]_behavmerge.csv*: MRI data for each of the three cohorts (discovery, validation and complete) for each of the 308 regions in the 500mm^2 parcellation. Separate files are provided for mean and standard deviation values of MT at each measurement depth. Fractional depths are not represented as measured from the pial surface as in the manuscript rather they are named, as originally calculated, from the grey/white matter boundary. Therefore, for example, the values corresponding to MT at 70% cortical depth for all participants in the discovery cohort are in the file *DISCOVERY/PARC_500aparc_MT_mean_projfrac+030_mean_behavmerge.csv*.

- *PLS_gene_predictor_vars.csv*: Gene expression matrix containing gene-expression for each of 20,737 genes at each of the 306 (of 308) cortical regions that had usable data from the AIBS data set.

- *Candidate_genes_oligo.csv*: The list of genes from the oligodendrocyte candidate gene set along with their index in the whole-genome gene expression matrix.

- *Candidate_genes_schizophrenia.csv*: The list of genes from the schizophrenia-risk candidate gene set along with their index in the whole-genome gene expression matrix.

The following three files are included because they have been drawn offline and are necessary to create the main manuscript figures using the analysis and figure-generation code provided.

  - *CorticalLayers_schematic_cells.jpg*: This image is represented in **Fig. 2c**.

  - *CorticalLayers_schematic_methods.jpg*: This image is represented in **Fig. 2a**.

  - *Fig3_Enrich_withColourBar.png*: This image is represented in **Fig. 3g**.

Analysis Code

All analysis code necessary to replicate the analyses presented in the paper is available for download from the **SCRIPTS.zip** compressed directory contained within the project's figshare repository. **NSPN_CorticalMyelination_AnalysisWrapper.py** calls all other functions (stored in the **SCRIPTS** folder) and gives a clear explanation of the analysis steps, including the automated creation of figures and tables.

SCRIPTS.zip download link: https://dx.doi.org/10.6084/m9.figshare.2057805

NSPN_CorticalMyelination_AnalysisWrapper.py download link: https://dx.doi.org/10.6084/m9.figshare.2057808

The analysis code is dependent on the following software packages:

- FSL http://fsl.fmrib.ox.ac.uk/fsl/fslwiki version: 5.0.6
- Freesurfer http://freesurfer.net version: freesurfer-Linux-centos4_x86_64-stable-pub-v5.3.0
- Pysurfer http://pysurfer.github.io version: 0.6
- Nibabel http://nipy.org/nibabel version: 1.2.0
- Anaconda http://docs.continuum.io/anaconda/index version: 2.3.0 (64-bit)
- Python via Anaconda (above) version 2.7.10
- Conda via Anaconda (above) version: 3.15.1
- Seaborn http://stanford.edu/~mwaskom/software/seaborn version: 0.6.0
- Networkx https://networkx.github.io version: 1.9.1
- Community http://perso.crans.org/aynaud/communities/index.html version: 0.3
- Statsmodels http://statsmodels.sourceforge.net version: 0.6.1
- Matlab http://uk.mathworks.com release: 2012b
- Matlab statistical and machine learning toolbox http://uk.mathworks.com/products/statistics release: 2012b
- MiKTeX http://miktex.org/ version 2.9

Results

The results folder is not necessary and will be created by the analysis code if it does not already exist. However, many of the results presented here rely on permutation tests and therefore results may not be identical across different runs of the same analysis code. We encourage readers to replicate the analyses themselves (as we have done many times) but provide in the project's figshare repository the specific output files that were used in the creation of this manuscript.

The results folder is very large and exceeds figshare's file size limits. Therefore, it has been split into four parts within the project's figshare repository. The discovery, validation and complete analyses are provided separately (**DISCOVERY.zip**, **VALIDATION.zip** and **COMPLETE.zip**), with a compressed folder containing the remaining directories that integrate those results as a fourth file **CT_MT_ANALYSES.zip**.

The **DISCOVERY**, **VALIDATION** and **COMPLETE** directories should be unzipped so they are inside the **CT_MT_ANALYSES** directory. Combined, these folders contain all the output used to create this manuscript including the tables, high (and low) resolution figures and movies.

CT_MT_ANALYSES.zip download link: https://dx.doi.org/10.6084/m9.figshare.1618815

DISCOVERY.zip download link: https://dx.doi.org/10.6084/m9.figshare.2057811

VALIDATION.zip download link: https://dx.doi.org/10.6084/m9.figshare.2057814

COMPLETE.zip download link: https://dx.doi.org/10.6084/m9.figshare.2057820

Supplementary Files
All supplemental files have been uploaded to the project's figshare repository as **SUPPLEMENTAL_FILES.zip**.

SUPPLEMENTAL_FILES.zip download link: https://dx.doi.org/10.6084/m9.figshare.1618810

Specifically, **SUPPLEMENTAL_FILES.zip** contains:

- *WhitakerVertes_PLSEnrichmentGeneList.xlsx*: This file contains the list of significant enrichment terms for both the most positively weighted (up-regulated) genes and the most negatively weighted (down-regulated) genes in the second PLS component (PLS2) separately for the discovery, validation and complete cohorts. The enrichment of down-regulated genes was obtained by providing the inverse ranking of genes to the GOrilla software tool. Redundant terms, as determined by the REVIGO online software tool, are shaded in grey in the supplementary file, to highlight the most meaningful GO annotations. We also shaded out terms annotated to over 1000 genes given their generality (for example "cell communication").

- *GO_[COHORT]_PLS2_[DIR].png*: These six figures are very large and provide a detailed visualization of all significantly enriched biological processes embedded in the hierarchical tree of GO terms for PLS2. For example, *GO_complete_PLS2_pos.png* is the original tree (including process labels) presented in **Fig. 3g** and represents up-regulated biological processes in PLS2 in the complete cohort. *GO_complete_PLS2_neg.png* represents down-regulated biological processes in PLS2 in the complete cohort. The colour-coding of boxes represents the degree of significance of each term based on its

uncorrected P-values: white ($P > 10^{-3}$), yellow ($10^{-3} < P < 10^{-5}$), pale orange ($10^{-5} < P < 10^{-7}$), orange ($10^{-7} < P < 10^{-9}$), red ($P < 10^{-9}$). Note that the colourbars for the corresponding figure in the main text have been adapted to represent FDR corrected P values and the file *WhitakerVertes_PLSEnrichmentGeneList.xlsx* reports both FDR corrected and uncorrected P values.

- *RegionalMeasures_[COHORT].pdf*: These three tables present summary measures of baseline CT, baseline MT, ΔCT with age, ΔMT with age, PLS2 weightings, degree and closeness for the discovery, validation and complete cohorts. The *P* values are not corrected for multiple comparisons but presented in order to assess the patterns in the data. All MT values in this table are sampled at 70% cortical depth. Values were calculated based on 308 regions but for readability we present median values from all sub-regions within each of the 34 Desikan-Killiany atlas regions. Tables separated by hemisphere (68 regions) and for all 308 regions are available in the **DISCOVERY.zip**, **VALIDATION.zip** and **COMPLETE.zip** compressed directories in the project's figshare respository, within the **TABLES** subdirectory for each cohort.

Ongoing support

On acceptance of this manuscript, a GitHub repository was created in order to facilitate ongoing work with the project and support additional users. Information is available at the repository home page, wiki and associated website: http://kirstiejane.github.io/NSPN_WhitakerVertes_PNAS2016.

Supplementary Note: Adherence to Transparency and Openness Promotion (TOP) Guidelines

In 2015 the Center for Open Science published guidelines that journals could adopt to ensure best practice in research reporting (16). Here we outline these guidelines and to which level of adherence the current study best fits.

Citation standards: level 3

All processed data, program code and other methods are appropriately cited in the supplementary information.

Data transparency: level 3

- All MRI and gene expression data after parcellation and the behavioural and demographic information for the 297 participants featured in this manuscript are available in the **DATA.zip** compressed folder in the project's figshare respository. Additionally, the candidate gene lists for oligodendrocyte specific genes (derived from ref 17) and Schizophrenia risk genes (derived from ref 18) are available in the same location. These files were made available upon first submission of this manuscript.

Analytic methods (code) transparency: level 3

- All code to replicate the analyses is provided in the **SCRIPTS.zip** compressed compressed folder in the project's figshare respository along with a wrapper script (**NSPN_CorticalMyelination_AnalysisWrapper.py**; https://dx.doi.org/10.6084/m9.figshare.2057808) that completes all statistical analyses and reporting.

Research Materials: level 2

Edinburgh Handedness Index (EHI) is available in the original publication (19).

Index of Multiple Deprivation (IMD) values are available from the Office of National Statistics at https://data.gov.uk/dataset/index_of_multiple_deprivation_imd_2007.

The Weschler Abbreviated Scale of Intelligence (WASI) may be acquired from http://www.pearsonclinical.com/education/products/100000593/wechsler-abbreviated-scale-of-intelligence-wasi.html.

Inquiries for multi-parametric mapping (MPM) acquisition parameters for Siemens Trio TIM scanners and/or raw data processing code in Matlab may be sent to Dr Nikolaus Weiskopf at n.weiskopf@ucl.ac.uk.

Design and analysis transparency: level 3

Our materials and methods section, along with the shared data and code described above, permit us to complete, where appropriate, the checklists recommended by Poldrack et al (20).- the closest match found in the Enhancing the quality and transparency of health Research (Equator) Network (www.equator-network.org) - and Ridgway et al (21).

Preregistration of studies and analysis plans: level 1

Neither this study nor the statistical analysis plan were pre-registered in an institutional registry.

Replication: NA

The TOP guidelines outline suggestions for journals to accept independent peer-reviewed replication studies. This guideline is not relevant to individual publications, but we would like to note that all analyses in the current manuscript replicated in an independent (validation) cohort, albeit not by independent investigators.

Supplementary Note: Author Contributions

KJW and ETB designed MRI analyses

PEV, TR and ETB designed genetic analyses

KJW and PEV conducted statistical analyses and network analyses

KJW, PEV, FV, RRG, KSW and RT processed and quality controlled MRI data

MFC and NW provided advice on creation, interpretation and analysis of quantitative maps from MPM data

CO and JS coordinated data management

BI, MM, GP coordinated data collection

PF, RJD, PBJ, IMG, ETB designed the study

KJW, PEV and ETB wrote the manuscript

All authors critically appraised the manuscript

Supplementary Note: Neuroscience in Psychiatry Network (NSPN) Consortium author list

Chief investigator:

Ian Goodyer

Principal investigators:

Edward Bullmore
Raymond Dolan
Peter Fonagy
Peter Jones

Associated faculty:

Peter Dayan
Paul Fletcher
John Suckling
Nikolaus Weiskopf
Pasco Fearon

Project managers:

Becky Inkster
Gita Prabhu

Postdoctoral research associates and associated research fellows:

Eran Eldar
Tobias Hauser
Konstantinos Ioannidis
Gemma Lewis
Alda Mita
Michael Moutoussis
Sharon Neufeld
Ela Polek-MacDaeid
Rafael Romero-Garcia
Michelle St Clair
Jan Stochl
Roger Tait
Beata Tick
Umar Toseeb
Anne-Laura van Harmelen
Petra Vértes
Kirstie Whitaker
Geert-Jan Will
Gabriel Ziegler

Jorge Zimbron

PhD & MSc students:

Joost Haarsma
Sian Davies
Juliet Griffin
Michael Hart
Jakob Seidlitz
Max Shinn
František Váša
Konrad Wagstyl

Data managers:

Cinly Ooi
Barry Widmer

Research assistants:

Ayesha Alrumaithi
Sarah Birt
Kalia Cleridou
Hina Dadabhoy
Ashlyn Firkins
Sian Granville
Elizabeth Harding
Alexandra Hopkins
Daniel Isaacs
Janchai King
Clare Knight
Danae Kokorikou
Christina Maurice
Cleo McIntosh
Jessica Memarzia
Harriet Mills
Ciara O'Donnell
Sara Pantaleone
Jennifer Scott
Alison Stribling

Administration team:

Junaid Bhatti
Neil Hubbard

Natalia Ilicheva
Michael Kentell
Ben Wallis
Laura Villis

Supplementary References

1. Weiskopf N, et al. (2011) Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *Neuroimage* 54(3):2116–24.
2. Helms G, Dathe H, Dechent P (2008) Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. *Magn Reson Med* 59(3):667–72.
3. Helms G, Dathe H, Kallenberg K, Dechent P (2008) High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. *Magn Reson Med* 60(6):1396–407.
4. Helms G, Dechent P (2009) Increased SNR and reduced distortions by averaging multiple gradient echo signals in 3D FLASH imaging of the human brain at 3T. *J Magn Reson Imaging* 29(1):198–204.
5. Helms G, Dathe H, Weiskopf N, Dechent P (2011) Identification of signal bias in the variable flip angle method by linear display of the algebraic Ernst equation. *Magn Reson Med* 66(3):669–77.
6. Lutti A, et al. (2012) Robust and fast whole brain mapping of the RF transmit field B1 at 7T. *PLoS One* 7(3):e32379.
7. Lutti A, Hutton C, Finsterbusch J, Helms G, Weiskopf N (2010) Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T. *Magn Reson Med* 64(1):229–38.
8. Preibisch C, Deichmann R (2009) Influence of RF spoiling on the stability and accuracy of T1 mapping based on spoiled FLASH with varying flip angles. *Magn Reson Med* 61(1):125–35.
9. Helms G, Dathe H, Dechent P (2010) Modeling the influence of TR and excitation flip angle on the magnetization transfer ratio (MTR) in human brain obtained from 3D spoiled gradient echo MRI. *Magn Reson Med* 64(1):177–85.
10. Rosas HD, et al. (2002) Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology* 58(5):695–701.
11. Hawrylycz MJ, et al. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489(7416):391–399.
12. Hagberg A, Swart P, Chult D (2008) Exploring network structure, dynamics, and function using NetworkX. (SciPy):11–15.
13. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 69(2 Pt 2):026113.
14. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):P10008.
15. Humphries MD, Gurney K (2008) Network "small-world-ness": A quantitative method for determining canonical network equivalence. *PLoS One* 3(4). doi:10.1371/journal.pone.0002051.
16. Nosek BA, et al. (2015) Promoting an open research culture. *Science (80-)* 348(6242):1422–5.
17. Cahoy JD, et al. (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 28(1):264–78.
18. Ripke S, et al. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511(7510):421–427.
19. Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9(1):97–113.
20. Poldrack RA, et al. (2008) Guidelines for reporting an fMRI study. *Neuroimage* 40(2):409–414.
21. Ridgway GR, et al. (2008) Ten simple rules for reporting voxel-based morphometry studies. *Neuroimage* 40(4):1429–35.