Genetic architecture of the white matter connectome of the human brain

Zhiqiang Sha[1], Dick Schijven[1], Simon E. Fisher[1,2], Clyde Francks[1,2,3]*

[1] Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[3] Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands

Corresponding author:
Clyde Francks, DPhil. Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; E-mail: Clyde.Francks@mpi.nl

Manuscript information: 10,029 words (Introduction, Results, Discussion, Methods). 5 figures.
(Supplementary Information: 14 figures, 31 tables.)

Short title: Genetics of human brain structural connectivity

Abstract
White matter tracts form the structural basis of large-scale networks in the human brain. We applied brain-wide tractography to diffusion images from 30,810 adult participants (UK Biobank), and found significant heritability for 90 node-level network connectivity measures and 851 edge-level connectivity measures. Multivariate genome-wide association analyses identified 325 independently associated lead single-nucleotide polymorphisms (SNPs) across the genome, of which 80% had not been previously associated with human brain metrics. Enrichment analyses implicated neurodevelopmental processes including neurogenesis, neural differentiation, neural migration, neural projection guidance, and axon development, as well as prenatal brain expression especially in stem cells, astrocytes, microglia and neurons. We used the multivariate association profiles of lead SNPs to identify 31 genomic loci implicated in structural connectivity between core regions of the left-hemisphere language network. Polygenic scores for schizophrenia, bipolar disorder, autism spectrum disorder, attention-deficit hyperactivity disorder, left-handedness, Alzheimer's disease, amyotrophic lateral sclerosis, and epilepsy showed significant multivariate associations with structural connectivity, each implicating distinct sets of brain regions with trait-relevant functional profiles. This large-scale mapping study revealed common genetic contributions to the structural connectome of the human brain in the general adult population, highlighting links with polygenic disposition to brain disorders and behavioural traits.

One sentence summary: Variability in white matter connections of the human brain is associated with hundreds of newly discovered genomic loci that especially implicate stem, neural and glial cells during prenatal development, and is also associated with polygenic dispositions to various brain disorders and behavioural traits.

Keywords
Diffusion magnetic resonance imaging, white matter connectivity, brain fiber tracts, structural connectivity, multivariate association, genome-wide association scan, polygenic scores, UK Biobank.

Introduction

Cognitive functions and behaviours are supported by dynamic interactions of neural signals within large-scale brain networks(*1*). Neural signals propagate along white matter connections that link cortical, subcortical, and cerebellar regions, to form the structural connectome(*2, 3*). White matter connections also modulate neural signals and distribute trophic factors between brain regions(*4, 5*), helping to establish and maintain functional specialization of sub-networks. Various heritable psychiatric and neurological disorders can involve altered white matter structural connectivity, relating for example to cognitive deficits, clinical presentation or recovery(*6-10*). It is therefore of great interest to understand which DNA variants, genes and pathways affect white matter connections in the human brain, as they are likely to influence cognitive and behavioural variability in the population, as well as predisposition to brain disorders.

Diffusion tensor imaging (DTI) enables *in vivo* non-invasive study of white matter in the brain(*11, 12*). This technique characterizes the diffusion of water molecules, which occurs preferentially in parallel to nerve fibers due to constraints imposed by axonal membranes and myelin sheaths(*13, 14*). Metrics commonly derived from DTI, such as fractional anisotropy or mean diffusivity, reflect white matter microstructure and can index its integrity(*13-15*). In contrast, tractography involves defining white matter connections at the macro-anatomical scale, which permits the measurement of connectivity strengths by counting the streamlines that link each pair of regions. Streamlines are constructed to pass through multiple adjacent voxels in DTI data, when the principal diffusion tensor per voxel aligns well with some of its direct neighbors(*16*). Tractography therefore produces subject-specific measures of regional inter-connectivity that are ideally suited for brain network-level analysis.

Recently, genome-wide association studies (GWAS) have reported that a substantial proportion of inter-individual variability in white matter microstructural measures can be explained by common genetic variants, with single nucleotide polymorphism (SNP)-based heritabilities ranging from 22% to 66%(*17, 18*). These studies also identified specific genomic loci associated with white matter microstructural measures (*17, 18*). However, microstructural measures do not necessarily capture topological properties of macro-scale brain networks, such as the total amount of structural connectivity between distant pairs of brain regions. In principal, inter-individual variability in topological features of the white matter connectome may be influenced by genetic variants that are partly distinct from those that influence white matter microstructure. For example, genetic influences on axon outgrowth and guidance during the development of long-distance connections may be most detectable in terms of connection strengths as measured through tractography, without necessarily affecting the microstructural integrity of those connections. However, to our knowledge, nerve fiber tractography has not previously been used for large-scale genome-wide association analysis of brain structural networks, likely due to heavy computational requirements for running tractography in tens of thousands of individuals.

Here, we aimed to characterize the genetic architecture of white matter structural network connectivity in the human brain, using fiber tractography. DTI data from 30,810 participants of the UK Biobank adult population dataset were used to construct the brain-wide structural connectivity network of each individual. In combination with genome-wide genotype data, we then carried out a set of genetic analyses of tractography-derived metrics, in terms of the sum of white matter connectivity linking to each of 90 brain regions as network nodes, and 947 connectivity measures as network edges linking specific pairs of regions. The total connectivity of a node (brain region) likely relates to its global role in information transfer within multiple sub-networks, whereas individual connections between specific pairs of regions are more locally restricted measures. We anticipated that genetic influences on node-level and edge-level network measures might therefore be partly distinct, where some genetic effects are more relevant at larger scales whereas others could affect relatively specific circuit components.

Our genetic analyses included SNP-based heritability estimation, multivariate GWAS (mvGWAS), and biological annotation of associated loci. Then, to illustrate how multivariate gene-brain associations arose in the data, and how the brain-wide mvGWAS results could be queried in relation to any specific brain network of interest, we used the results to identify genomic loci that are associated with structural connections between core language-related regions of the left-hemisphere. Various aspects of language function – especially related to language

production - show strong hemispheric lateralization, with roughly 85% of people having left-hemisphere dominance(*19*).

Finally, we assessed how genetic disposition to brain disorders and other behavioural traits manifests in terms of white matter connectivity in the general population. To do so, we mapped multivariate associations of the brain-wide, white matter tractography metrics with polygenic scores for a variety of heritable brain disorders or behavioural traits: schizophrenia, bipolar disorder, autism, attention-deficit hyperactivity disorder, left-handedness, Alzheimer's disease, amyotrophic lateral sclerosis, and epilepsy. We annotated the resulting brain maps with cognitive functions, using large-scale meta-analyzed functional neuroimaging data, to describe aspects of brain function that may be affected by polygenic dispositions to different forms of neurodivergence in the general population.

Results

White matter connectomes of 30,810 adults

For each of 30,810 adult participants with diffusion MRI and genetic data after quality control, we performed deterministic fiber tractography(*16*) between each pair of regions defined in the Automated Anatomical Labeling atlas(*20*) (45 regions per hemisphere comprising cerebral cortical and subcortical structures) (Fig. 1; Methods). In the structural connectivity matrix of each individual, each region was considered a node, and each connection between a pair of regions was considered an edge. We excluded edges when more than 20% of individuals had no streamlines connecting a given pair of regions, resulting in 947 network edges. To quantify a given edge in each individual, the streamline count for that edge was divided by the individual-specific grey matter volume of the two regions being connected (as larger regions tend to have more streamlines connecting to them). These volume-adjusted network edge measures were also used to calculate the node-level connectivity of each region, i.e. the sum of all volume-weighted edge measures connecting with a given region, for each participant. The resulting node and edge measures were adjusted for demographic and technical covariates, and normalized across individuals (see Methods - Network construction and analysis), before being used for genetic analyses.

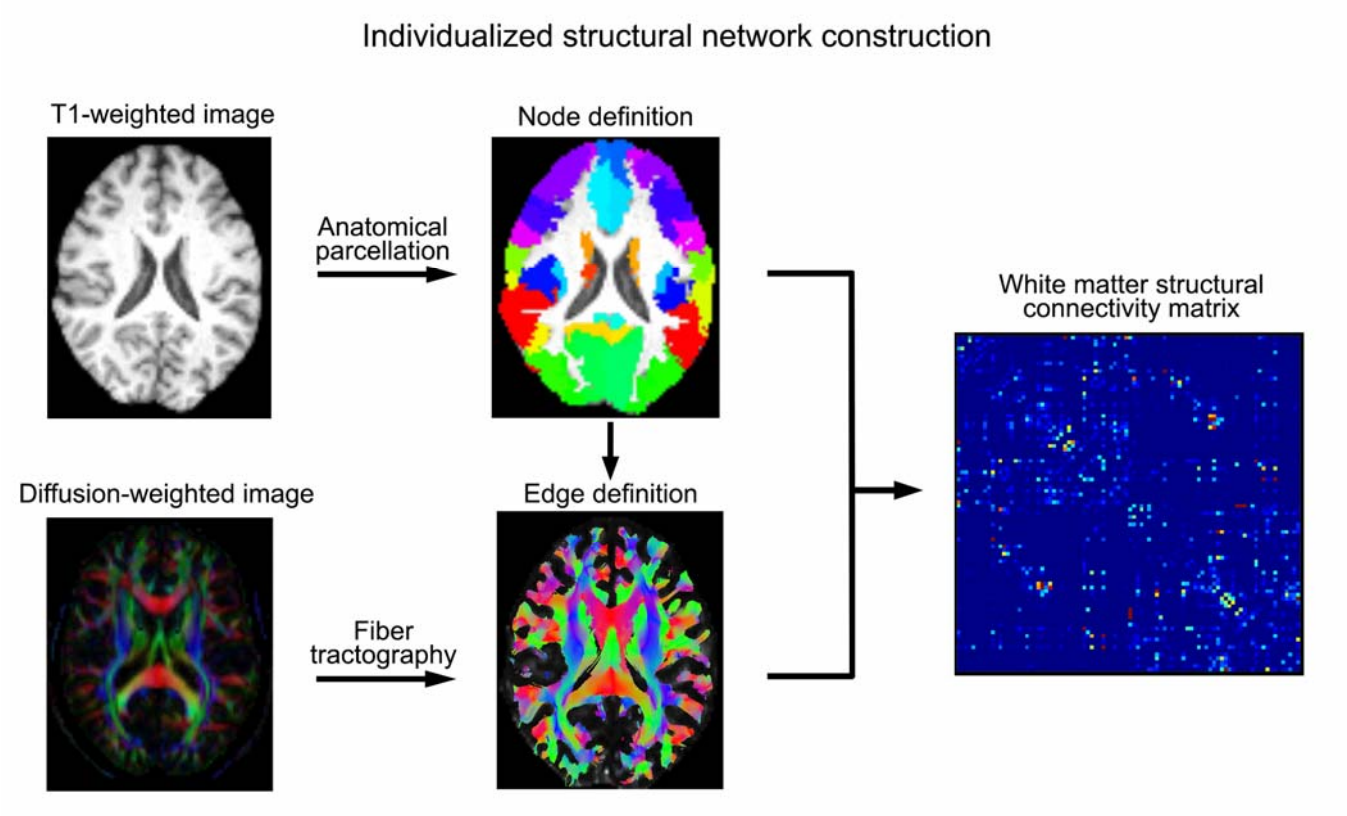

Figure 1. Schematic of white matter network construction within an individual brain. Network nodes were defined by mapping the Automated Anatomical Labeling atlas from common MNI space to individual space, with 45 regions per hemisphere (including cortical and subcortical structures). The edge between each pair of regions was defined as the number of streamlines constructed by tractography based on the corresponding diffusion tensor image, adjusted for the combined volume of the two connected regions. The process yielded a zero-diagonal symmetrical 90×90 undirected connectivity matrix for each of 30,810 participants (the upper triangles were then used for subsequent analyses).

Of the 947 network edges, 377 connected pairs of left-hemisphere regions, 355 connected pairs of right-hemisphere regions, and 215 involved interhemispheric connections. The top 10% of regions in terms of connectivity included the supplementary motor cortex, precuneus, medial superior frontal cortex, and subcortical regions bilaterally – caudate and thalamus (Supplementary Figure 1 and Supplementary Table 1). The latter observation is consistent with previous studies showing that subcortical regions connect widely with the cerebral cortex, to generate reciprocal cortical-subcortical interactions that together support many cognitive functions(*4, 21, 22*).

Heritabilities of connectivity measures

The GCTA software(*23*) was used to estimate the SNP-based heritability (h^2) for each network measure, that is, the extent to which variance in each connectivity measure was explained by common genetic variants across the autosomes (Methods). All of the 90 node-level (region-based) connectivity measures were significantly heritable (Bonferroni-corrected $p<0.05$ after testing 90 measures), ranging from 7.8% to 29.5% (mean h^2=18.5%; Fig. 2A, Supplementary Figure 2 and Supplementary Table 2). Most homologous nodes of the left and right hemispheres showed similar heritabilities, but some nodes showed prominent differences of heritability between hemispheres, such as the inferior parietal cortex (left: 27.0% vs. right: 19.42%), pars triangularis (left: 23.4% vs. right: 16.9%) and inferior occipital cortex (left: 8.0% vs. right: 15.7%; Fig. 2A and Supplementary Table 2). Eleven node-level connectivities showed h^2 estimates >25% (Supplementary Table 2), with the superior temporal cortex in the left hemisphere being highest (h^2=29.5%, $p<1\times10^{-20}$).

851 of 947 edge-level connectivities (i.e. connections between specific pairs of regions) showed significant heritability (Bonferroni-corrected $p<0.05$ after testing 947 measures, Supplementary Figure 2), ranging from 4.6% to 29.5% (mean 9.6%). Eleven edges had h^2>20%, primarily for connections linking frontal regions (e.g. superior and middle frontal cortex), supplementary motor and occipital cortex (e.g. cuneus and lingual). The mean h^2 was 9.9% for 351 edges within the left hemisphere, 10.0% for 333 edges within the right hemisphere, and 8.1% for 167 inter-hemispheric edges (Fig. 2B and Supplementary Tables 3-6). Across the 851 significantly heritable edges, heritability was not correlated with mean fiber length (rho=0.01, Supplementary Figure 3), suggesting that short-range and long-range white matter connections are similarly affected by genetic variation overall.

Reliability of heritable network measures was assessed using data from 1,005 of the 30,810 individuals who had undergone brain scans on two separate occasions (Methods): intraclass correlations (ICCs) based on the same processing pipeline applied to data from the first and second scanning visits had median 0.83 (range 0.45 to 0.93) for the 90 heritable node-level measures, and median 0.66 (range 0.28 to 0.94) for the 851 heritable edge-level measures (Supplementary Figure 4, Supplementary Tables 7 and 8). Measurement reliability was positively correlated with heritability: r=0.67 across 90 node-level connectivities, and r=0.60 across 851 heritable edge-level connectivities (Supplementary Figure 4). This is consistent with the fact that measurement error is assigned to the 'environmental' (i.e. non-genetic) component of trait variance in heritability analysis, so that less reliable measures tend to be less heritable. Regardless, we reasoned that all significantly heritable network measures had enough reliably measured variation to contribute to subsequent mvGWAS analysis. This is because detecting heritability relies on trait similarity being increased in pairs of individuals with higher genetic similarity, and therefore necessitates at least a small proportion of trait variance being reliably measured across and within individuals, while single genetic loci are only expected to explain very small amounts of that heritable variance.

The heritability data can be visualized interactively in a dynamic web-based interface (see Data availability statement).

Multivariate genome-wide association analyses of white matter connectivity

The MOSTest software(*24*) was used to perform two separate mvGWAS analyses, first for the 90 node-level connectivity measures in a single multivariate genome-wide screen, and then for the 851 edge-level connectivities in another single multivariate genome-wide screen, both times in relation to 9,803,735 SNPs spanning the genome. This analysis examined each SNP separately for its associations with multiple structural network measures, by simultaneously modelling the distributed nature of genetic influences across the brain (Methods). FUMA(*25*) was used to clump mvGWAS results on the basis of linkage disequilibrium (LD) and to identify independent lead SNPs at each associated genomic locus (Methods). At the $p=2.5\times10^{-8}$ significance level (i.e. the standard GWAS significance threshold of $p=5\times10^{-8}$ but Bonferroni-corrected for two mvGWAS), we identified 140 lead SNPs in 117 distinct genomic loci associated with node-level connectivities (Fig. 2C, Supplementary Figure 5 and Supplementary Table 9), and 211 lead SNPs in 166 distinct genomic loci associated with edge-level connectivities (Fig. 2C, Supplementary Figure 5 and Supplementary Table 10). Permutation analysis under the null hypothesis of no association indicated that MOSTest correctly controlled type I error (Methods; Supplementary Figures 6 and 7). Except for chromosome 21, each chromosome had at least one locus associated with either node-level or edge-level connectivity.

26 lead SNPs were found in common between the node-level mvGWAS and edge-level mvGWAS. While a degree of overlap was to be expected given that the node-level metrics were computed from the edge-level metrics (i.e. are not independent), the fact that the large majority of lead SNPs were detected for either node-level or edge-level connectivity, but not both, supports the importance of performing genetic association analyses at these different network levels.

For each lead SNP, MOSTest indicated the contribution of each brain metric to its multivariate association, by reporting a z-score derived from each metric's univariate association with that SNP (Methods; Supplementary Tables 11 and 12). In the node-level mvGWAS, regions with the greatest magnitude z-scores considered across all lead SNPs were the bilateral putamen (left mean $|z|=2.05$, right mean $|z|=1.86$), left pallidum (mean $|z|=2.02$), bilateral middle frontal cortex (left mean $|z|=1.98$, right mean $|z|=1.89$) and middle cingulate cortex (mean $|z|=1.82$; Supplementary Figure 8 and Supplementary Table 13). For example, the left putamen, which had the highest overall contribution across lead SNPs (mean $|z|=2.05$), was especially strongly associated with rs12146713 on 12q23.3 ($z=-10.48$), rs72748148 on 9q31.3 ($z=7.35$), rs798528 on 7p22.3 ($z=-6.74$), rs7935166 on 11p11.2 ($z=6.22$) and rs3795503 on 1q25.3 ($z=6.01$; Supplementary Table 13).

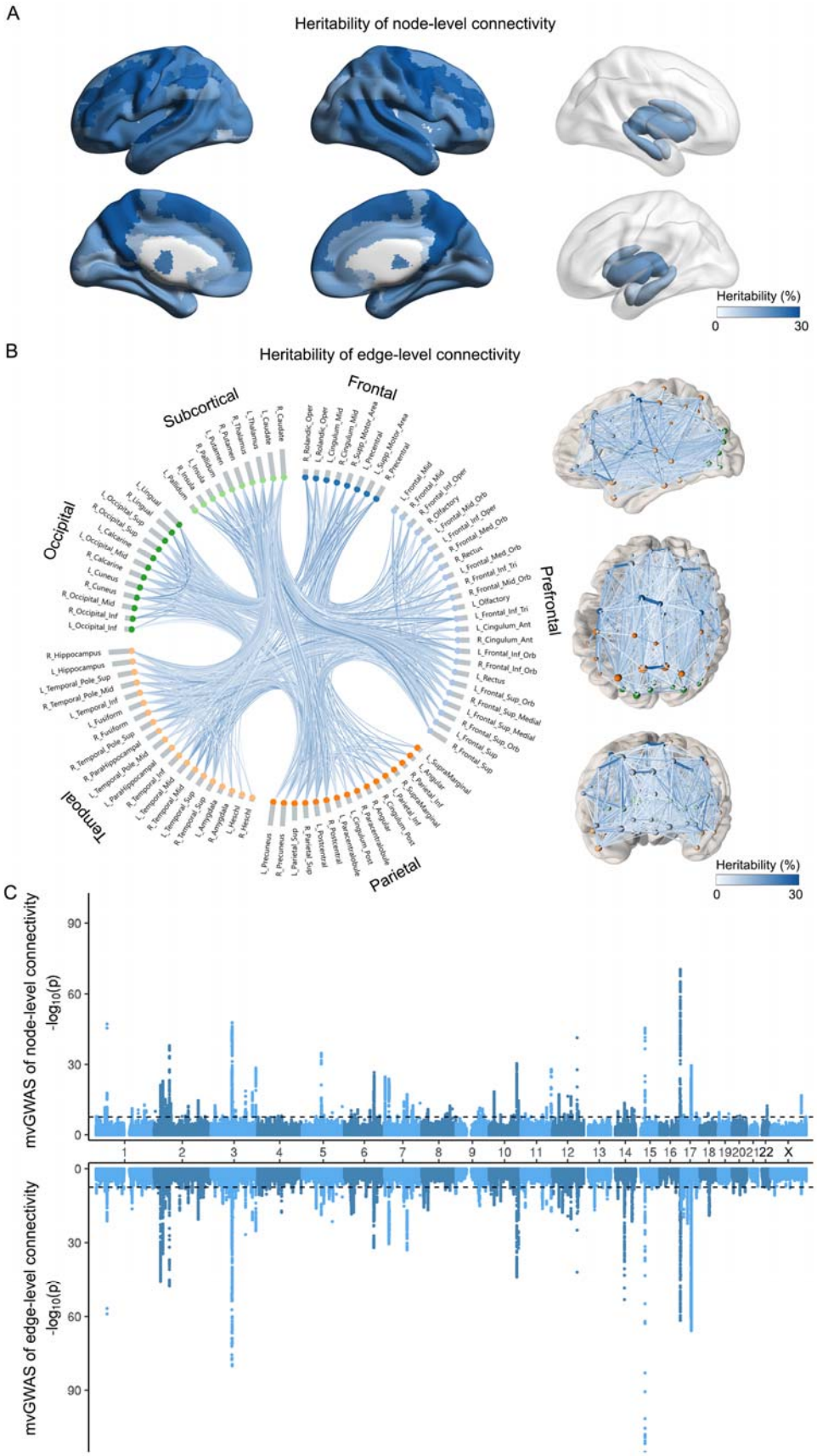

Figure 2. SNP-based heritability and multivariate GWAS analyses of node-level connectivity and edge-level connectivity in 30,810 participants. (A) All 90 node-level (i.e. regional) connectivities showed significant SNP-based heritability after Bonferroni correction, ranging from 7.8% to 29.5%. (B) 851 out of 947 edge-level connectivities showed significant SNP-based heritability after Bonferroni correction, ranging from 4.6% to 29.5%. Right panel: brain maps. Left panel: nodes grouped by frontal, prefrontal, parietal, temporal, and occipital cortical lobes, and subcortical structures. Heritabilities can be visualized interactively in a dynamic web-based interface (see Data availability statement). (C) Miami plot for multivariate GWAS of 90 node-level connectivities (upper) and 851 edge-level connectivities (lower). The black lines indicate the genome-wide significance threshold $p<2.5\times10^{-8}$ (Methods).

In the mvGWAS of edge-level connectivity, edges that showed high magnitude z-scores considered across all lead SNPs mainly connected the precuneus, calcarine, middle temporal and pre- and post-central cortex (Supplementary Table 14 and Supplementary Figure 9). The edge linking the left and right precuneus had the greatest contribution across lead SNPs (mean |z|=1.60), and was especially associated with the variants rs946711 on 10p12.31 (z=-5.58) and 3:190646282_TA_T on 3q28 (z=-5.53).

The majority of genomic loci associated with structural connectivity are novel

Together, our node-level connectivity mvGWAS and edge-level connectivity mvGWAS identified 325 lead SNPs, of which only 101 were previously associated with at least one trait in the NHGRI-EBI GWAS catalog(*26*) (Supplementary Tables 9 and 10). This indicates that the majority (68.92%) of loci implicated here in the structural connectome were not identified by previous studies. There were 65 SNPs in common with those reported in previous GWAS of brain measures(*18, 24, 27-29*). Specifically, 46 of our lead SNPs were previously associated with brain regional volumes(*27, 30*), 29 with regional cortical thicknesses(*24, 29*), 33 with regional cortical surface areas(*29, 31*), and 20 with white matter microstructure(*18, 32*). Apart from brain measures, 11 of our lead SNPs were associated with mental health traits (e.g. autism, schizophrenia, and anxiety)(*33, 34*), 11 of our lead SNPs with cognitive functions (e.g. cognitive ability and performance)(*35, 36*), 4 of our lead SNPs with neurological diseases (e.g. Alzheimer's disease and epilepsy)(*37, 38*), and 40 of our lead SNPs with non-brain physiological and physical variables (e.g. waist-hip ratio, cholesterol levels and lung function)(*39, 40*). In addition, we compared our results with those reported in a recent GWAS of white matter microstructure integrity for which the results have not been deposited in the GWAS Catalog(*17*): 32 of their lead SNPs overlapped with those from our mvGWAS analyses (Supplementary Table 15).

Functional annotations of genomic loci associated with the structural connectome

We used FUMA(*25*) to annotate SNPs to genes at significantly associated loci by three strategies: physical position, expression quantitative trait locus (eQTL) information and chromatin interactions (Methods). For the node-level connectivity mvGWAS, 879 unique genes were identified through these three strategies (Supplementary Table 16 and Supplementary Figure 10). 95 out of 140 lead SNPs had at least one eQTL or chromatin interaction annotation, indicating that these variants (or other variants in high LD with them) affect gene expression. For example, rs7935166 on 11p11.2 (multivariate z=5.71, $p=1.15\times10^{-8}$) is intronic to *CD82*, which has been reported to promote oligodendrocyte differentiation and myelination of white matter(*41*). This lead SNP is a brain eQTL(*42, 43*) of *CD82*, and also shows evidence for cross-locus chromatin interaction via the promoter of *CD82* in adult brain(*42*). Another example: rs35396874 on 6q21 (multivariate z=6.64, $p=3.17\times10^{-11}$) affects the expression of its surrounding gene *FOXO3*, a core element of the TLR/AKT/FoxO3 pathway that is important for repairing white matter injury mediated by oligodendrocyte progenitor cell differentiation(*44, 45*).

For the edge-level connectivity mvGWAS, functional annotation identified 1464 unique genes (Supplementary Table 17 and Supplementary Figure 10). 135 of 211 lead SNPs had at least one eQTL annotation or chromatin interaction. For example, rs13084442 on 3q26.31 (multivariate z=6.34, $p=2.32\times10^{-10}$) is an eQTL(*46*) of *TNIK*, a gene associated with neurogenesis and intellectual disability(*47, 48*). The same lead variant is also located in a region having a chromatin interaction with *TNIK* promoters in fetal and adult cortex(*49*). Similarly, the SNP rs28413051 on 4q31.23 (multivariate z=6.28, $p=3.47\times10^{-10}$) is an eQTL of *DCLK2* that is important for axon growth cone formation and neural migration(*50, 51*), and is also within a region interacting with the promoter of *DCLK2* in neural progenitor cells(*44*). A further example: allele C of rs13107325 on 4q24 (multivariate z=5.77, $p= 7.99\times10^{-9}$ in the node-level connectivity mvGWAS and multivariate z=8.74, $p=2.37\times10^{-18}$ in the edge-level connectivity mvGWAS) is a missense coding variant in the gene *SLC39A8* that showed a high combined annotation-dependent depletion (CADD) score of 23.1 (Methods), which indicates that this SNP is deleterious (its frequency was 7.01%). The same SNP has been associated with white matter microstructure integrity(*28*), schizophrenia(*52, 53*) and children's behavioural problems(*54*).

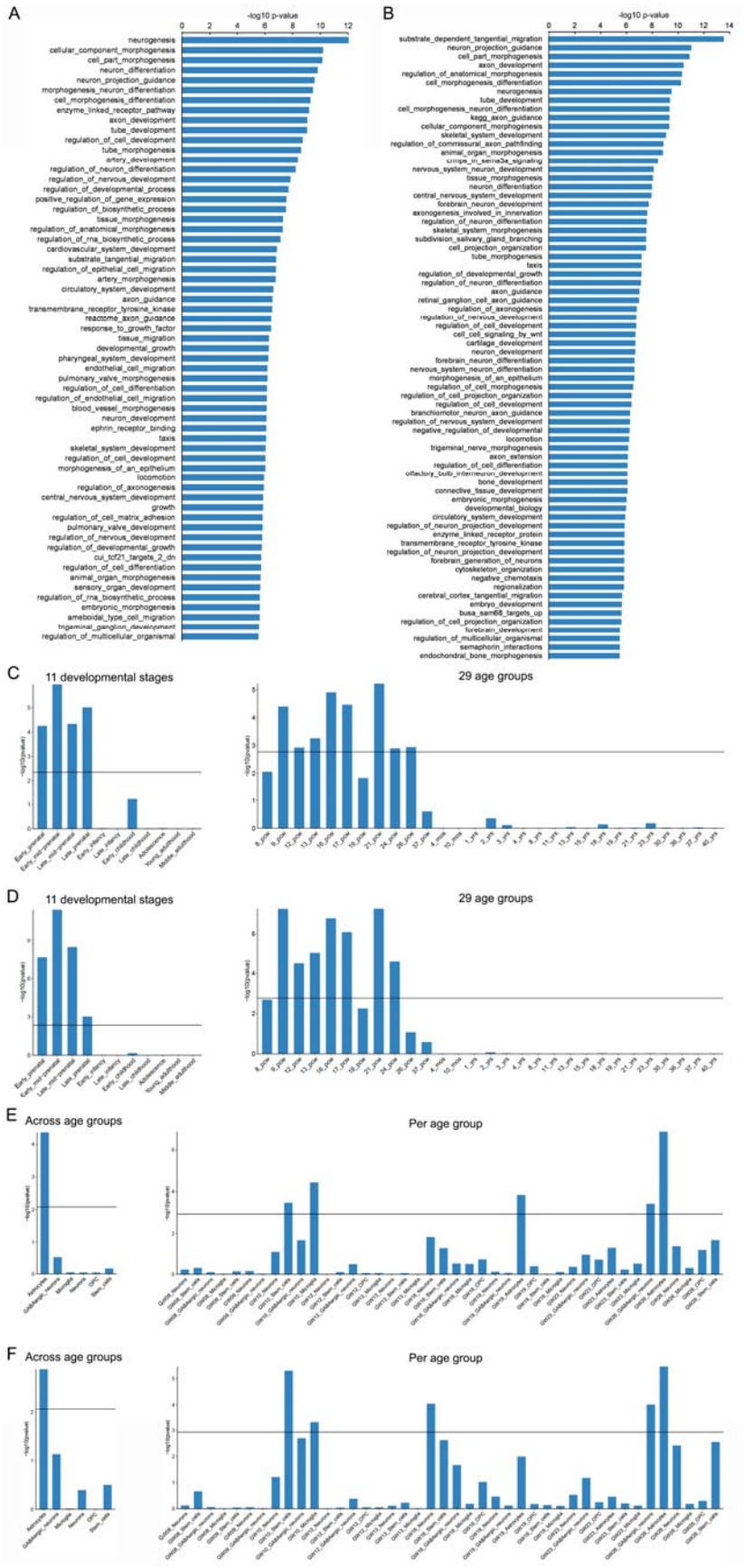

Figure 3. Genes associated with variation in the adult white matter connectome are enriched for specific neurodevelopmental roles. (A) Sixty-one functionally-defined gene sets showed significant enrichment of association with node-level connectivity. (B) Seventy-two functionally-defined gene sets showed significant enrichment of association with edge-level connectivity. (C-D) Based on BrainSpan data from 11 lifespan stages or 29 age groups, genes associated with variation in (C) adult node-level connectivity and (D) adult edge-level connectivity show upregulation in the human brain prenatally. (E-F) Based on single-cell gene expression data from the prenatal brain, genes associated with variation in (E) adult node-level connectivity show upregulation in astrocytes when considering all prenatal age groups combined, and in stem cells and microglia at 10 gestational weeks, astrocytes at 19 gestational weeks, and GABAergic neurons and astrocytes at 26 gestational weeks when breaking down by developmental stages, and similarly genes associated with variation in (F) adult edge-level connectivity show upregulation in astrocytes when considering all prenatal age groups combined, and in stem cells and microglia at 10 gestational weeks, neurons at 16 gestational weeks, and GABAergic neurons and astrocytes at 26 gestational weeks when breaking down by developmental stages. (C-F) Black lines indicate the significance threshold $p<0.05$ after Bonferroni correction within each analysis. PCW: post-conceptional weeks.

Gene-based association analysis and gene set enrichment analysis for the brain's structural connectome

We used MAGMA(*31*) to perform gene-based association analysis, which combines the mvGWAS evidence for association at each SNP within a given gene, while controlling for LD.

For node-level connectivities we identified 296 significant genes with $p<0.05$ (after Bonferroni correction for testing 20,146 genes and two sets of mvGWAS results) (Supplementary Table 18 and Supplementary Figure 11), 237 of which overlapped with those annotated by at least one of the three strategies used above (i.e. physical location, eQTL annotation or chromatin interaction). The gene-based p values were then used as input to perform gene-set enrichment analysis, in relation to 15,488 previously defined functional sets within the MSigDB database(*55*). Sixty-one gene sets showed significant enrichment (Bonferroni adjusted $p<0.05$ for testing 15,488 sets, Fig 3A and Supplementary Table 19) which mainly implicated neurodevelopmental processes, such as "go_neurogenesis" (beta=0.18, $p=5.53\times10^{-13}$; the most significant set), "go_neuron_differentiation" (beta=0.18, $p=1.55\times10^{-10}$), and "go_cell_morphogenesis_involved_in_neuron_differentiation" (beta=0.25, $p=3.39\times10^{-10}$).

For edge-level connectivities we identified 561 genes with significant gene-based association (Bonferroni-corrected p values <0.05 for testing 20,146 genes and two sets of mvGWAS results), 444 of which overlapped with genes mapped through physical location, eQTL annotation or chromatin interaction (Supplementary Table 20 and Supplementary Figure 11). Seventy-two gene sets were significant after Bonferroni correction for 15,488 sets (Fig. 3B and Supplementary Table 21), related especially to neural migration and the development of neural projections, such as "go_substrate_dependent_cerebral_cortex_tangential_migration" (beta=3.98, $p=2.61\times10^{-14}$; the most significant set), "go_neuron_projection_guidance" (beta=0.41, $p=8.59\times10^{-12}$), and "go_axon_development" (beta=0.29, $p=3.45\times10^{-11}$).

We tested our genome-wide, gene-based p values with respect to human brain gene expression data from the BrainSpan database(*56*), grouped according to 11 lifespan stages or 29 different age groups. Genes associated with node-level connectivity showed upregulation on average across much of the prenatal period, ranging from early (beta=0.04, $p=5.84\times10^{-5}$) to late (beta=0.08, $p=1.01\times10^{-5}$) prenatal stages, or from 9 (beta=0.002, $p=4.15\times10^{-5}$) to 26 (beta=0.003, $p=1.18\times10^{-3}$) post-conceptional weeks (Bonferroni-corrected p values <0.05; Fig. 3C and Supplementary Table 22). Similarly, genes associated with edge-level connectivities showed upregulation on average during early (beta=0.06, $p=2.35\times10^{-8}$) to late (beta=0.06, $p=1.01\times10^{-3}$) prenatal stages, or from 9 (beta=0.003, $p=5.92\times10^{-8}$) to 24 (beta=0.003, $p=2.67\times10^{-5}$) post-conceptional weeks (Bonferroni-corrected p values <0.05; Fig. 3D and Supplementary Table 23).

We also examined our genome-wide, gene-based association p values with respect to two independent single-cell gene expression datasets derived from human prefrontal cortex samples of different ages (GSE104276)(*57*). Combining across age groups, average upregulation was observed in astrocytes for genes associated with both node-level connectivity (beta=0.05, $p=4.34\times10^{-5}$) and edge-level connectivity (beta=0.04, $p=1.27\times10^{-3}$) (Bonferroni-corrected p values <0.05; Fig. 3E and 3F and Supplementary Tables 24 and 25). Breaking down by age, genes associated with node-level connectivity were upregulated on average in microglia (beta=0.02, $p=3.72\times10^{-5}$) and stem cells (beta=0.05, $p=3.51\times10^{-4}$) at 10 gestational weeks of age (GW), astrocytes at 19GW (beta=0.02, $p=1.48\times10^{-4}$) and 26GW (beta=0.05, $p=1.35\times10^{-7}$), and GABAergic neurons at 26GW (beta=0.04, $p=3.97\times10^{-4}$) (Fig. 3E and Supplementary Table 24). Similarly, genes associated with edge-level connectivities showed upregulation on average in microglia (beta=0.02, $p=5.07\times10^{-4}$) and stem cells (beta=0.06, $p=5.18\times10^{-6}$) at 10GW, neurons at 16GW (beta=0.06, $p=9.32\times10^{-5}$), and astrocytes (beta=0.05, $p=3.50\times10^{-6}$) and GABAergic neurons at 26GW (beta=0.05, $p=1.01\times10^{-4}$); Fig. 3F and Supplementary Table 25).

Genetics of left-hemisphere language network connectivity

To illustrate how the brain-wide mvGWAS results can be queried in relation to any specific brain network of interest, we selected four left-hemisphere regions that correspond to a network that is reliably activated by sentence-level language tasks in a left-lateralized manner in the majority of people and across languages(*58-61*), i.e. the opercular and triangular parts of inferior frontal cortex (Broca's region), and the superior and middle temporal cortex (including Wernicke's region, Fig. 4). These four nodes are linked by six edges with heritabilities

ranging from 7.3% to 17.1% (Supplementary Table 3), which together correspond well to the arcuate fasciculus and also probably include streamlines via the uncinate fasciculus – two prominent fiber tracts involved in language (Fig. 4)(*62, 63*). Of the 211 lead SNPs from our brain-wide mvGWAS of edge-level connectivity, 31 were significantly associated with at least one of these six edges according to the edge-specific z-scores derived from MOSTest (Bonferroni correction at 0.05; Supplementary Table 26). For example, rs12636275 on 3p11.1 is located within an intron of *EPHA3*, a gene that encodes an ephrin receptor subunit that regulates the formation of axon projection maps(*64*), and has also been associated with functional connectivity between language-related regions(*65*). Another example: rs7580864 on 2q33.1 is an eQTL of *PLCL1* that is implicated in autism(*66, 67*), a neurodevelopmental disorder that often affects language and social skills. Other positional candidate genes based on the 31 SNPs include *CRHR1*, encoding corticotropin releasing hormone receptor 1, and *CENPW* (centromere protein W) involved in chromosome maintenance and the cell cycle (Fig. 4 and Supplementary Table 26).

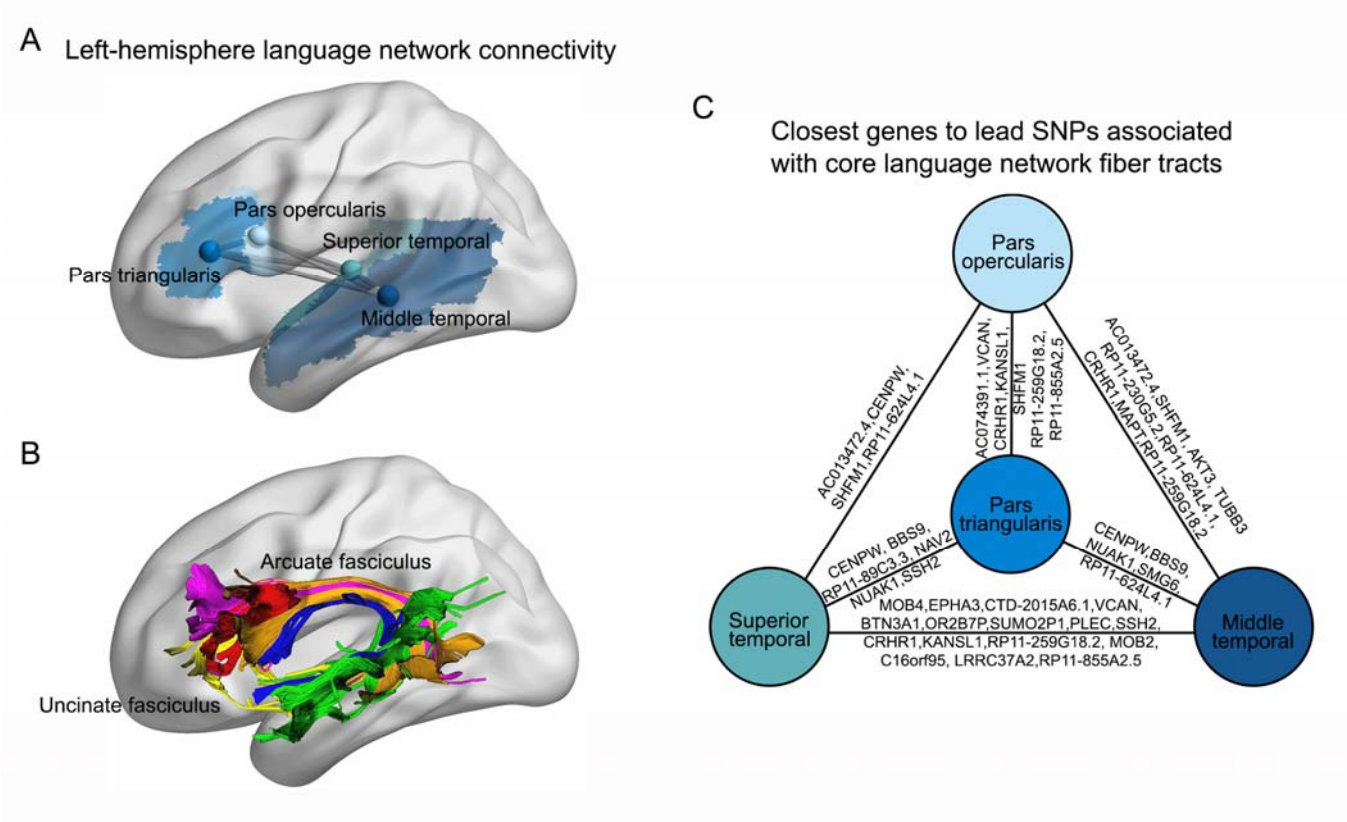

Figure 4. Genetics of left-hemisphere language network connectivity. A: Four regions with core functions in the left-hemisphere language network, encompassing the classically defined Broca's (frontal lobe) and Wernicke's (temporal lobe) areas. Also shown are the six edges connecting these four regions when considered as network nodes. B: Visualization of the six edges in an example individual, with red representing connections between the pars opercularis and pars triangularis, green connections between the middle temporal and superior temporal cortex, gold connections between the pars opercularis and middle temporal cortex, blue connections between the pars opercularis and superior temporal cortex, purple connections between the pars triangularis and middle temporal cortex, yellow connections between pars triangularis and superior temporal cortex. C: The closest genes to independent lead SNPs from the brain-wide multivariate GWAS of edge-level connectivity, that showed significant association with at least one of the six left-hemisphere language network edges (Bonferroni correction at 0.05; Supplementary Table 26).

Multivariate associations of the structural connectome with polygenic scores for brain disorders and behavioural traits

For each of the 30,810 individuals in our study sample we calculated polygenic scores(*68*) for various brain disorders or behavioural traits that have shown associations with white matter variation, using previously published GWAS summary statistics: schizophrenia(*17, 69-71*), bipolar disorder(*72-74*), autism(*17, 70, 75, 76*),

attention deficit/hyperactivity disorder(*77-79*), left-handedness(*80, 81*), Alzheimer's disease(*82-84*), amyotrophic lateral sclerosis(*85-87*), and epilepsy(*88-90*) (Methods). There were 18 significant partial correlations (i.e. adjusted for demographic and technical covariates – see Methods) between different pairs of these polygenic scores across individuals (Bonferroni-corrected $p<0.05$): 16 correlations were positive, with the highest between polygenic scores for schizophrenia and bipolar disorder ($r=0.36$, $p<1\times10^{-200}$), and between attention deficit/hyperactivity disorder and autism ($r=0.33$, $p<1\times10^{-200}$), while 2 were negative, between polygenic scores for amyotrophic lateral sclerosis and bipolar disorder ($r=-0.03$, $p=2.26\times10^{-6}$), and amyotrophic lateral sclerosis and autism ($r=-0.03$, $p=8.81\times10^{-6}$; Supplementary Table 27 and Supplementary Figure 12).

Separately for each of these polygenic scores, we used canonical correlation analysis to investigate their multivariate associations with the 90 heritable node-level connectivity measures across the 30,810 individuals. All canonical correlations were highly significant: schizophrenia $r=0.07$, $p=8.98\times10^{-34}$; bipolar disorder $r=0.07$, $p=1.53\times10^{-35}$; autism $r=0.06$, $p=7.87\times10^{-24}$; attention deficit/hyperactivity disorder $r=0.08$, $p=7.84\times10^{-44}$; left-handedness $r=0.07$, $p=1.74\times10^{-31}$; Alzheimer's disease $r=0.07$, $p=4.14\times10^{-33}$; amyotrophic lateral sclerosis $r=0.06$, $p=1.29\times10^{-25}$; epilepsy $r=0.05$, $p=1.49\times10^{-20}$. Therefore, polygenic dispositions to these various disorders or behavioural traits in the general population are partly reflected in the brain's white matter connectivity.

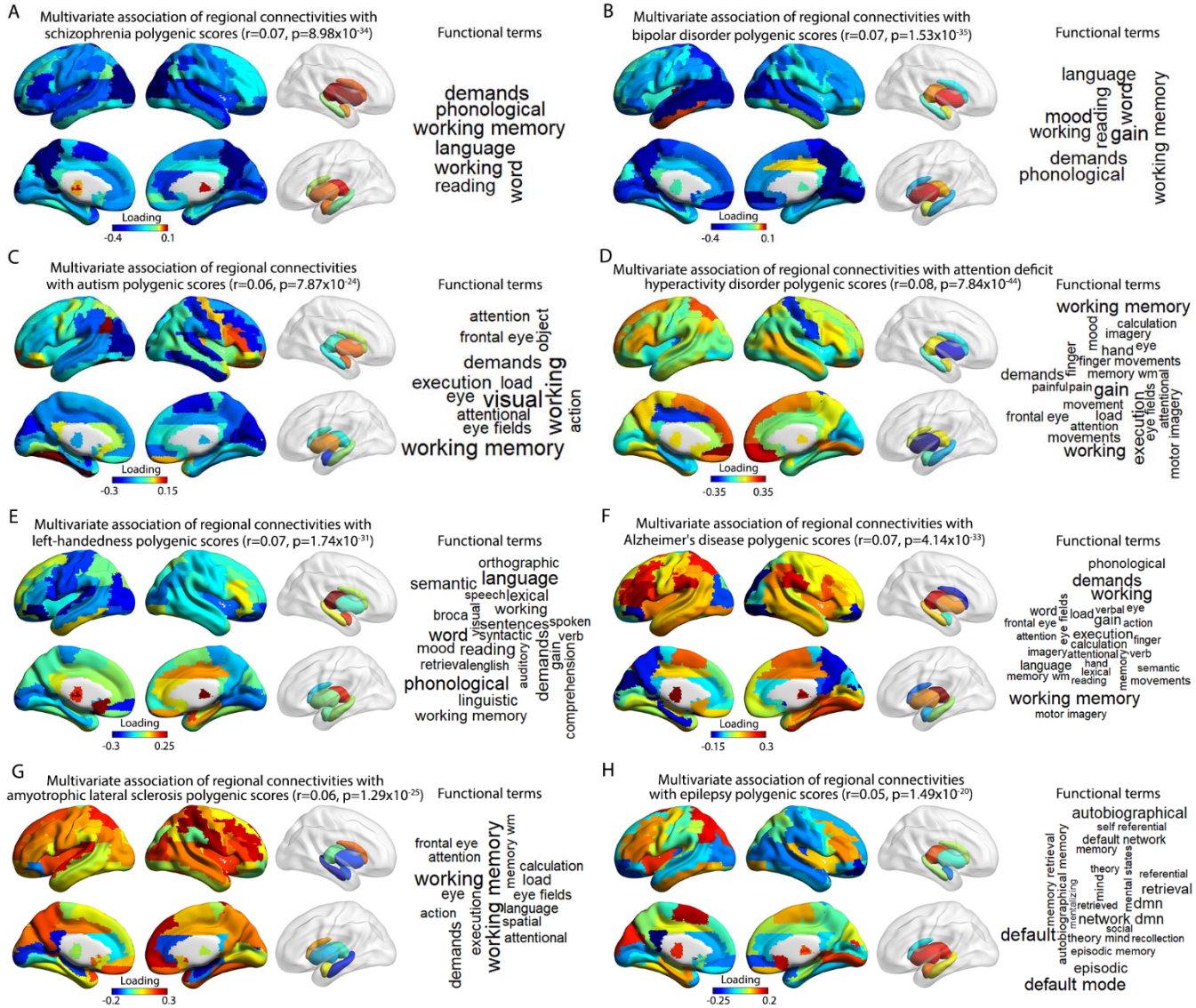

Figure 5. Polygenic dispositions to various brain-related disorders or behavioural traits show multivariate associations with regional (node-level) white matter connectivities in 30,810 participants. Loadings are shown from canonical correlation analyses that indicate the extent and direction to which each node-level connectivity is associated with polygenic scores for (A) schizophrenia, (B) bipolar disorder, (C) autism, (D) attention deficit/hyperactivity disorder, (E) left-handedness, (F) Alzheimer's disease, (G) amyotrophic lateral sclerosis and (H) epilepsy. A positive loading (red) indicates a higher node-level connectivity associated with increased polygenic disposition for a given disorder/behavioural trait, while a negative loading (blue) represents a lower node-level connectivity associated with increased polygenic disposition for a given disorder/behavioural trait. Word clouds represent functions associated with the map of regions (nodes) showing the strongest loadings (|r|>0.2) for each polygenic score. Functions were assigned using large-scale meta-analyzed functional neuroimaging data (Methods). The font sizes in the word clouds represent correlation magnitudes between the meta-analyzed functional maps for those terms and the co-activation map for the set of regions associated with each polygenic score. See Supplementary Table 30 for the correlation coefficients.

Canonical correlation analyses yielded loadings for each node-level connectivity measure, reflecting the extent and direction of each measure's association with polygenic disposition for a given disorder/behavioural trait. For psychiatric disorders the majority of loadings were negative, i.e. increased polygenic risk for these disorders was more often associated with reduced than increased connectivity across brain regions (Fig. 5, Supplementary Table 28). This was especially marked for polygenic risks for schizophrenia (85 regions with negative loadings, 5 regions with positive loadings), bipolar disorder (81 negative, 9 positive) and autism (64 negative, 26 positive). Polygenic disposition to left-handedness was also associated with more reduced node-level connectivities (62 negative loadings) than increased node-level connectivities (28 positive loadings). In contrast, increased polygenic risk for Alzheimer's disease was associated with increased white matter connectivity for a majority of brain regions (62 out of 90) in the UK Biobank data, even while some core regions of disorder pathology showed decreased connectivity, such as posterior cingulate and medial temporal cortex(*91-93*). (These results remained stable when excluding the *APOE* locus that is known to have a substantial individual effect on Alzheimer's disease risk – see Methods and Supplementary Table 29). Similar observations were made for polygenic risk for amyotrophic lateral sclerosis, where 74 out of 90 regions showed positive loadings (Fig. 5, Supplementary Table 28).

For each polygenic score we identified the specific node-level connectivities that showed the strongest loadings in canonical correlation analyses, i.e. brain regions with loadings >0.2 or <-0.2. These regions were used to create a single brain mask for each polygenic score, which was then used to query the Neurosynth database of 14,371 functional brain imaging studies(*94*). In this process, a brain-wide co-activation map was generated for each mask, based on all functional maps in the database, and these were then correlated with cognitive and behavioural term-specific maps derived from the studies included in the database(*94*).

For example, the mask for schizophrenia polygenic risk comprised 32 regions showing the strongest associations with white matter connectivity, distributed in the bilateral temporal, dorsoventral and posterior cingulate cortex (Fig. 5A and Supplementary Table 28), and there were 7 functional term-based correlations >0.2 with the corresponding co-activation map for these regions (Fig. 5A, Supplementary Figure 13 and Supplementary Table 30), including 'working memory' and 'language'. This suggests that polygenic disposition to schizophrenia influences the connectivity of brain regions especially involved in working memory and language (see Discussion). The mask for bipolar disorder polygenic risk comprised 30 regions, including temporal, medial frontal, superior parietal, and visual cortex, as well as hippocampus and caudate, and these regions together received functional annotations of 'mood', 'working memory' and language-related processes (Fig. 5B, Supplementary Figure 13 and Supplementary Tables 28 and 30). Polygenic risk for autism was mainly associated with white matter connectivity of the right dorsolateral prefrontal, right temporal, right sensorimotor and bilateral visual cortex, as well as the left amygdala, and these regions were annotated with visual, working memory, executive and attention functions (Fig. 5C, Supplementary Figure 13 and Supplementary Tables 28 and 30). Polygenic disposition to left-handedness was associated with node-level connectivity of Broca's area, left superior temporal cortex, left medial prefrontal and left visual cortex and right thalamus, functionally annotated with language-related cognitive functions (Fig. 5E, Supplementary Figure 13 and Supplementary Tables 28 and 30). See Fig. 5, Supplementary Figure 13 and Supplementary Tables 28 and 30 for the equivalent maps and functional annotations for all disorder/trait polygenic scores.

The polygenic risks for bipolar disorder and schizophrenia had the most similar brain maps, in terms of node-level structural connectivity associated with each of these polygenic risks (r=0.56 between the loadings for these two polygenic scores, across the 90 nodes; Supplementary Figure 14 and Supplementary Table 31).

Discussion

This large-scale mapping study employed white matter tractography and multivariate analysis to characterize the contributions of common genetic variants to individual differences in structural connectivity of the adult human brain. Multivariate associations between structural connectivity and polygenic dispositions to brain-related disorders or behavioural traits were also characterized, and described in terms of functional activations of the implicated brain regions. Together, these various analyses in over 30,000 individuals from the general population linked multiple levels of biological organization: from genes and cell types through developmental stages to adult brain structure and function, behaviour and individual differences, implicating hundreds of genomic loci that had not previously been associated with human brain measures.

Different brain regions are inter-connected through white matter nerve fibers - this fundamental property sub-serves functional networks involved in cognition and behaviour. In over 30,000 adults from the general population we found that inter-individual variation in white matter connectivity is especially influenced by genes that are i) active in the prenatal developing brain; ii) upregulated in stem cells, astrocytes, microglia, and neurons of the embryonic and fetal brain; and iii) involved in neurodevelopmental processes including neural migration, neural projection guidance and axon development. A likely neurodevelopmental origin of much inter-individual variation of adult white matter connectivity is consistent with findings from large-scale imaging genetics studies of other aspects of brain structural and functional variation(*17, 29, 76*). These statistical enrichment findings serve as a strong biological validation of our multivariate GWAS findings, as there was no reason for such clearly relevant functional enrichment to occur by chance in relation to brain white matter tracts.

Astrocytes are the largest class of brain glial cells with a range of known functions, including neuronal homeostasis and survival, regulation of synaptogenesis and synaptic transmission(*95*). Less well known is that during neurodevelopment, astrocytes can express positional guidance cues, such as semaphorin 3a, that are required for neuronal circuit formation, through mediating the attraction or repulsion of the growth cone at the axonal tip(*96*). In our gene-based association analysis, *SEMA3A* was the most significantly associated individual gene with edge-level connectivity in the whole genome. Taken together, our data suggest that the formation of fiber tracts in the developing human brain may be affected substantially by positional cues provided by astrocytes, in addition to neurons.

As regards microglia – these phagocytic cells are well known for immune functions but also help to remove dying neurons and prune synapses, as well as modulate neuronal activity(*97*). Less is known of their roles during development, but embryonic microglia are unevenly distributed in the brain and associate with developing axons, which again suggests roles in regulating axonal growth and positional guidance(*98*). Mouse brains without microglia, or with immune activated microglia, show abnormal dopaminergic axon outgrowth(*99*), while disruption of microglial function or depletion of microglia results in a failure of growing axons to adhere and form bundles in the corpus callosum – the largest fiber tract of the brain(*100*). Our data support such observations, through showing that genes up-regulated in microglia in the embryonic human brain are enriched for variants that associate with individual differences in adult white matter connectivity. Further research on the roles of astrocytes and microglia in fiber tract development is therefore warranted.

While our results point especially to genes involved in neurodevelopment, it is also likely that some genetic effects on white matter connectivity act later in life. For example, astrocytes and microglia may affect the maintenance and aging of brain fiber tracts during adulthood, with implications for brain disorders, and possibly suggesting therapeutic targets. We mapped the multivariate associations of polygenic scores for various brain-related disorders and behavioural traits with regional white matter connectivities, and annotated the resulting brain maps using meta-analyzed functional imaging data. Some maps and their annotations were consistent with the symptomatology of the traits in question – for example, polygenic disposition to bipolar disorder was associated with white matter connectivity of brain regions prominently involved in mood, while polygenic dispositions to

attention deficit/hyperactivity disorder or autism were associated with the connectivity of regions important for executive functions. Polygenic scores for left-handedness and for schizophrenia were associated with the connectivity of language-related regions – consistent with altered left-hemisphere functional dominance for language in both of these traits(*19, 101*), and a phenotypic association between them(*102*). Polygenic scores for left-handedness and schizophrenia have also been associated with altered structural asymmetry of grey matter in language-related regions(*70, 80*).

Regarding genetic risks for neurological disorders, polygenic scores for Alzheimer's disease and amyotrophic lateral sclerosis were associated with the connectivity of regions important for working memory, while polygenic scores for epilepsy were associated with connectivity of the default mode network – a set of brain regions involved in internally-initiated thoughts, and semantic and episodic memory(*103*). Previous analysis of white matter tracts in Alzheimer's disease has indicated a broad-based reduction of connectivity(*104*), so it was surprising that the majority of brain regions in the UK Biobank adult population dataset showed increased connectivity with higher polygenic risk for this disorder, even while some core regions of pathology showed decreased connectivity as expected. A similarly notable pattern was seen for amyotrophic lateral sclerosis, where increased polygenic risk was associated with increased structural connectivity for a majority of brain regions. It may be that increased connectivity of some regions occurs as a compensatory re-configuration in response to decreased connectivity of others, or at least that white matter connectivity is relatively spared while cortical grey matter is reduced during aging of those at higher polygenic risk (*105*). In addition, the UK Biobank volunteer sample is healthier than the general population(*106*). Those at higher polygenic risk who manifest a given disorder may tend not to participate, leaving an unusually healthy set of volunteers among those with high polygenic risk. Such recruitment bias may influence brain correlates of polygenic risk in the UK Biobank.

The brain-wide multivariate GWAS approach that we used provided high statistical power to detect relevant genomic loci, compared to a mass univariate approach(*24*). At the same time, the multivariate results could be queried *post hoc* to identify loci associated with particular edge-level connectivities of interest. We illustrated this by querying the results with respect to six connections linking four core regions of the left-hemisphere language network – together approximating to Broca's and Wernicke's classically defined functional areas(*107*). As expected, these connections together formed an overall feature that closely resembles the arcuate fasciculus plus some connections running through the uncinate fasciculus (Fig. 4) – the two major language-related tracts (*62, 63*). Thirty-one implicated loci included the *EPHA3* locus, encoding an ephrin receptor subunit that acts as a positional guidance cue for the formation of axon projection maps, and has also been associated with functional connectivity between regional components of the language network that are especially involved in semantics(*65*). This is therefore a concordant genetic finding with respect to both structural and functional connectivity of the human brain's language network.

In this study we used deterministic tractography which we found to be computationally feasible in more than 30,000 individuals (and which took roughly four months of processing on a cluster server). An alternative approach - probabilistic tractography – is generally more demanding in terms of run time and storage requirements, but can have advantages, especially as it permits modelling of multiple tract orientations per voxel and therefore allows for crossing fibers (*108, 109*). However, it has been reported that deterministic tractography tends to have a lower likelihood of generating false positive connections than probabilistic approaches(*110*). This is important because false positives can be more detrimental to the correct calculation of network measures than false negatives(*110*). Similarly, Sarwar et al. (*111*) assessed the performance across tractography models and reported that deterministic tractography yielded the most accurate connectome reconstructions, especially when omitting connections with the fewest number of streamlines. Accordingly, we only included a connection in our structural connectivity matrix when it was detected in at least 80% of participants. Applying this threshold removed 3058 weak or spurious connections from our study, leaving 947 for further analysis, of which 851 showed significant heritability and were taken forward into mvGWAS analysis. This threshold has been found to be suitable for white matter network property analysis, as it provides a balance between the elimination of false positive connections and creating false negative connections(*112*). Anatomical connectivity constructed by

deterministic tractography has been well confirmed by microdissection in the post mortem human brain(*113*), indicating robustness and reliability of this approach(*114*). In addition, deterministic tractography has been widely used to construct white matter connectivity patterns in previous diffusion MRI studies which investigated structural characteristics during neurodevelopment, aging, and in brain disorders(*8, 115, 116*).

It has been reported that tract specificity can be lost through bottlenecks such as the corpus callosum when applying DTI-based tractography(*117*), although other studies have reported that deterministic tractography can achieve successful reconstruction of interhemispheric connections via the corpus callosum(*118, 119*). In addition, interhemispheric connectivity via the corpus callosum that was reconstructed from conventional diffusion imaging data has been supported by high angular resolution diffusion imaging (HARDI), and also by post mortem examination of white matter anatomy(*120-122*). We found the average heritability of interhemispheric connections to be only slightly lower than for intrahemispheric connections, which suggests that the interhemispheric connections were measured reliably enough to contribute to the identification of genetic effects.

We measured structural connectivity linking pairs of regions defined according to parcellation at the macroscopic level, using the AAL atlas (26). This atlas was created by manual neuroanatomical delineation on the basis of high-resolution structural imaging. Previous studies have found that this atlas, in combination with DTI data, reliably indexes structural connectivity(*123-125*). In addition, the AAL atlas in combination with deterministic tractography has been applied before in a study of vascular burden and cognitive ability in the UK Biobank dataset(*126*), which showed that this protocol is suitable and practical given the large size and scanning resolution of this dataset. Furthermore, AAL atlas regions are defined in volume space and are therefore consistent to apply across cortical and subcortical structures (unlike cortical surface-based segmentations), which was a goal of the present study. Tractography using volume-based atlases, such as AAL, has also been shown to capture variation arising from cortical thickness or bundle shape better than surface-based atlases(*127, 128*). The AAL labeling system integrates anatomical features from sulcal and gyral geometry, while the relatively large regions help to overcome variability arising during spatial registration and normalization of brain images from different individuals. While it would be informative to apply different atlases in future studies, it is likely that the percentage of inter-regional connections not detected (i.e. with zero streamlines) would increase with a more fine-grained atlas having more, smaller parcellations.

This study had some limitations: i) We maximized our statistical power for GWAS by using the available data as one large discovery sample, but this did not permit a discovery-replication design(*129*). Nonetheless, ultimately the total combined analysis in the largest available sample is the most representative of the available evidence for association. As mentioned earlier in this section, the various enrichment analyses indicated biological validity of the GWAS findings. Indeed, it has been argued that discovery-replication designs have less utility in the current era of Biobank-scale genetic studies than they used to, and that other forms of validation such as biological enrichment should be given increased weight in interpretation(*130*). ii) This was a large-scale observational mapping study, which meant that many of the analyses were screen-based and descriptive. Science proceeds through a combination of observation and hypothesis testing – this study incorporated both to varying degrees. Some of the biological observations were striking and informative, for example the likely involvements of microglia and astrocytes in affecting white matter tracts during embryonic and fetal development, which should now be studied more extensively in animal models. iii) This study did not consider rare genetic variants (with population frequencies below 1%). Future analysis of the UK Biobank's exome and genome sequence data in relation to white matter connectivity may reveal further genes and suggest additional mechanisms, cell types and lifespan stages in affecting inter-individual variation.

In summary, we used large-scale analysis to chart the white matter connectivity of the human brain, its multivariate genetic architecture, and its associations with polygenic dispositions to brain-related disorders and behavioural traits. The analyses implicated specific genomic loci, genes, pathways, cell types, developmental

stages, brain regions, fiber tracts, and cognitive functions, thus integrating multiple levels of analysis, and suggesting a range of future research directions at each of these levels.

Materials and Methods

Sample quality control

This study was conducted under UK Biobank application 16066, with Clyde Francks as principal investigator. The UK Biobank received ethical approval from the National Research Ethics Service Committee North West-Haydock (reference 11/NW/0382), and all of their procedures were performed in accordance with the World Medical Association guidelines(*131*). Written informed consent was provided by all of the enrolled participants. We used the dMRI data released in February 2020, together with the genome-wide genotyping array data. For individuals with available dMRI and genotype data, we first excluded subjects with a mismatch of their self-reported and genetically inferred sex, with putative sex chromosome aneuploidies, or who were outliers according to heterozygosity (principle component corrected heterozygosity >0.19) and genotype missingness (missing rate >0.05) as computed by Bycroft et al(*132*). To ensure a high degree of genetic homogeneity, analysis was limited to participants with white British ancestry, which was defined by Bycroft et al.(*132*) using a combination of self-report and cluster analysis based on the first six principal components that capture genetic ancestry. We also randomly excluded one subject from each pair with a kinship coefficient >0.0442, as calculated by Bycroft et al.(*132*). All of these metrics are available within UK Biobank data categories 263 or 100313. Our inclusion procedure finally resulted in 30,810 participants, with a mean age of 63.84 years (range 45-81 years), 14,636 were male and 16,174 were female.

Genetic quality control

We downloaded imputed SNP and insertion/deletion genotype data from the UK Biobank (i.e. v3 imputed data released in March 2018; UK Biobank data category 263 and data field 22828). QCTOOL (v.2.0.6) and PLINK v2.0(*133*) were used to perform genotype quality control. Specifically, we excluded variants with minor allele frequency <1%, Hardy-Weinberg equilibrium test p value $<1\times10^{-7}$ and imputation INFO score <0.7 (a measure of genotype imputation confidence), followed by removing multi-allelic variants that cannot be handled by many programs used in genetic-related analyses. This pipeline finally yielded 9,803,735 bi-allelic variants.

Diffusion MRI-Based Tractography

Diffusion MRI data were acquired from Siemens Skyra 3T scanners running protocol VD13A SP4, with a standard Siemens 32-channel RF receive head coil(*134*). We downloaded the quality-controlled dMRI data which were preprocessed by the UK Biobank brain imaging team(*134, 135*) (UK Biobank data field 20250, first imaging visit). The preprocessing pipeline included corrections for eddy currents, head motion, outlier slices, and gradient distortion. We did not make use of imaging-derived phenotypes released by the UK Biobank team, such as FA and mean diffusivity (microstructural measures). Rather, we used the quality-controlled dMRI data to perform tractography in volume space in each individual, which generated three-dimensional curves that characterize white matter fiber tracts. Briefly, diffusion tensors were modeled to generate a FA image in native diffusion space, which was used for deterministic diffusion tensor tractography using MRtrix3(*136*). Streamlines were seeded on a 0.5 mm grid for every voxel with FA 0.15 and propagated in 0.5 mm steps using fourth-order Runge-Kutta integration. Tractography was terminated if the streamline length was <20 or >250 mm, if it turned an angle >45˚, or reached a voxel with an FA <0.15. These parameters were consistent with a previous study exploring the structural network correlates of cognitive performance using the UK Biobank dataset(*126*). Tens of thousands of streamlines were generated to reconstruct the white matter connectivity matrix of each individual on the basis of the Automated Anatomical Labelling atlas(*20*) comprising a total of 90 regions encompassing cortical and subcortical structures (45 regions per hemisphere). This deterministic tractography process took roughly 16 weeks on 6 cluster server nodes running in parallel.

From the streamline data we computed the mean lengths of all inter-regional connections reconstructed by the deterministic tractography, using the "scale_length" option of the "tck2connectome" function in the MRtrix3 toolbox(*137*).

Network construction and analysis

Describing the structural network of each participant requires the definition of network nodes and edges. In this study, the network nodes corresponded to the 90 regions of the Automated Anatomical Labeling atlas, a three-dimensional volume-based parcellation scheme(*20*). The labeling system integrates detailed anatomical features from sulcal and gyral geometry, reducing anatomical variability that can arise from spatial registration and normalization of brain images taken from different individuals(*20*). For each participant, the T1 images (UK Biobank data field 20252, first imaging visit) were nonlinearly transformed into the ICBM152 T1 template in the MNI space to generate the transformation matrix(*138*). Inverse transformation was used to warp the Automated Anatomical Labeling atlas(*20*) from the MNI volume space to native volume space. Discrete labeling values were preserved using a nearest-neighbor interpolation method(*138*). Two nodes were considered connected if they were joined by the endpoints of at least one reconstructed streamline. Separately for each individual in the dataset, network edges were computed by the number of streamlines connecting a given pair of regions, divided by the volume of the two regions, because regions with larger volumes tend to have more streamlines connecting to them. This is a common approach in studies of white matter networks(*139-141*). We only included edges that were detected in at least 80% of participants, which removed 3058 weak or spurious connections from our study. This yielded a zero-diagonal symmetrical 90×90 undirected connectivity matrix for each participant, in which 947 edges were retained. The node-level connectivity of a region was then defined as the sum of all existing volume-weighted edges between that node and all other nodes in the network, reflecting the total connectivity of that node within the overall network.

Rank-based inverse normalization across individuals was performed on each network measure, and regression on age (UK Biobank field: 21003), nonlinear age (i.e. $(\text{age-mean_age})^2$), assessment center (UK Biobank data field: 54), genotype measurement batch (data field: 22000) and sex (data field: 31). Residuals were then further regressed on the first ten genetic principal components that capture population genetic diversity (UK Biobank field: 22009)(*132*), followed by rank-based inverse normalization of the residuals once more, and visual inspection of their distributions to confirm normality. The normalized, transformed measures were used for subsequent genetic analyses.

SNP-based heritability

We constructed a genetic relationship matrix using 9,516,306 variants on the autosomes with minor allele frequencies >1%, INFO score >0.7 and Hardy-Weinberg equilibrium p value $>1\times10^{-7}$, using GCTA(*23*) (version 1.93.0beta). We further excluded one random participant from each pair having a kinship coefficient higher than 0.025 (as SNP-based heritability analysis is especially sensitive to participants with higher levels of relatedness), yielding 29,027 participants for this particular analysis. Genome-based restricted maximum likelihood analyses were then performed to estimate the SNP-based heritability for each normalized structural network measure, again using GCTA(*23*). Bonferroni correction was applied separately for each type of network measure to identify those that were significantly heritable at adjusted $p<0.05$: 90 node-level connectivities and 851 edge-level connectivities.

Reliability of heritable network measures

For our main analysis of the 30,810 individuals (above), we used data from the first scanning visit at a UK Biobank assessment center. 1,005 of these individuals had also undergone brain scans (T1-structural and DTI) on a subsequent, separate occasion, from 733 to 974 days after their first scan. To examine the reliability of significantly heritable brain network measures we re-ran deterministic tractography on the 'second scan' data from these 1,005 individuals, with the same set of parameters and quality filters as the primary analysis, and re-computed the same edge-wise connectivity metrics as in the primary analysis. Each edge measure was then linearly adjusted for the same covariates as the main analysis, and rank-based inverse normalization was applied as in the main analysis. The adjusted, normalized values from the first and second scans were then used to

compute the intraclass correlation coefficient (ICC) for each heritable network measure, to evaluate reliability(*142*). ICC was calculated by the following formula:

$$ICC = \frac{BMS - WMS}{BMS + (m-1)WMS}$$

where BMS represents the between-individual mean square, WMS represents the within-individual mean square, and m indicates the number of repeat measures (here m=2).

Multivariate genome-wide association analysis

A total of 9,803,735 bi-allelic variants were used for mvGWAS analysis, spanning all autosomes and chromosome X. The sample size for mvGWAS was 30,810 (see sample quality control, above). We applied the multivariate omnibus statistical test (MOSTest) toolbox(*24*) to perform mvGWAS analysis for the significantly heritable measures, separately for node-level connectivities and edge-level connectivities. MOSTest can leverage the distributed nature of genetic influences across hundreds of spatially distributed brain phenotypes, while accounting for their covariances, which can boost statistical power to detect variant-phenotype associations(*24*). Specifically, the multivariate correlation structure is determined on randomly permuted genotype data. MOSTest calculates the Mahalanobis norm as the sum of squared de-correlated z-values across univariate GWAS summary statistics, to integrate effects across measures into a multivariate z statistic for each genetic variant, and uses the gamma cumulative density function to fit an analytic form for the null distribution. This permits extrapolation of the null distribution below the $p=5\times10^{-8}$ significance threshold without performing an unfeasible number of permutations (5×10^{-8} is a widely used threshold for GWAS multiple test correction in European-descent populations(*143, 144*)). Close matching of the null p value distributions from the permuted and analytic forms indicate that the method correctly controls type 1 error – this was the case for all four of our mvGWAS analyses (Supplementary Figures 6 and 7). In this framework the signs (positive or negative) of univariate z scores indicate the corresponding directions of effects (with respect to increasing numbers of minor alleles at a given SNP), whereas multivariate z scores are always positive.

Identification of genomic loci and functional annotations

We used FUMA (version v1.4.0)(*25*) to identify distinct genomic loci showing significant multivariate associations with brain structural connectivity, and apply functional annotations, using default parameters. Linkage disequilibrium (LD) structure was applied according to the 1000 Genomes European reference panel(*145*). SNPs with genome-wide significant mvGWAS p values $<2.5\times10^{-8}$ that had LD $r^2<0.6$ with any others were identified. For each of these SNPs, other SNPs that had $r^2\geq0.6$ with them were included for further annotation (see below), and independent 'lead SNPs' were also defined among them as having low LD ($r^2<0.1$) with any others. If LD blocks of significant SNPs were located within 250 kb of each other, they were merged into one genomic locus. Therefore, some genomic loci could include one or more independent lead SNPs. The major histocompatibility complex region on chromosome 6 was excluded from this process by default, because of its especially complex and long-range LD structure.

Functional annotations were applied by matching chromosome location, base-pair position, reference and alternate alleles to databases containing known functional annotations, which were ANNOVAR(*146*) categories, Combined Annotation-Dependent Depletion(*147*) scores, RegulomeDB(*148*) scores and chromatin state(*149, 150*):

1. ANNOVAR catogorizes SNPs on the basis of their locations with respect to genes, e.g. exonic, intronic and intergenic, using Ensembl gene definitions.
2. Combined Annotation-Dependent Depletion scores predict deleteriousness, with scores higher than 12.37 suggesting potential pathogenicity(*151*).
3. RegulomeDB scores integrate regulatory information from eQTL and chromatin marks, and range from 1a to 7, with lower scores representing more importance for regulatory function.
4. Chromatin states show the accessibility of genomic regions, and were labelled by 15 categorical states on the basis of five chromatin marks for 127 epigenomes in the Roadmap Epigenomics Project(*150*), which

were H3K4me3, H3K4me1, H3K36me3, H3K27me3 and H3K9me3. For each SNP, FUMA calculated the minimum chromatin state across 127 tissue/cell-types in the Roadmap Epigenomics Project(*132*). Categories 1-7 are considered open chromatin states.

We also used FUMA to annotate significantly associated SNPs, and other candidate SNPs that had $r^2 \geq 0.6$ with them, according to previously reported phenotype associations ($p<5 \times 10^{-5}$) in the National Human Genome Research Institute-European Bioinformatics Institute catalogue(*26*).

Multivariate association profiles of independently associated lead SNPs

For each SNP, MOSTest derives a z-score for each brain measure, calculated from the p value of the univariate association of that SNP with each individual measure. The z-scores give an indication of which measures contribute most to the multivariate association for a given SNP(*24*). We used the z-scores from the mvGWAS of fiber tracts to identify lead SNPs that were significantly associated with at least one from a set of six left-hemisphere language-related fiber tracts (see main text). To determine significance in this context, a threshold z-score with unsigned magnitude >3.56 was applied, corresponding to a p value of 2.37×10^{-4} (i.e. $p<0.05$ after Bonferroni correction for all 211 lead SNPs from the mvGWAS of fiber tracts, and considering six fiber tracts).

To determine which structural connectivity measures contributed most to the multivariate associations as considered across lead SNPs, we summed the unsigned univariate z-scores separately for each measure across all lead SNPs (separately for the mvGWAS analyses of node-level connectivities and fiber tracts).

SNP-to-gene mapping

Independent lead SNPs, and candidate SNPs having LD $r^2>0.6$ with a lead SNP, were mapped to genes in FUMA using the following three strategies.

1. Positional mapping was used to map SNPs to protein-coding genes based on physical distance (within 10 kb) in the human reference assembly (GRCh37/hg19).
2. eQTL mapping was used to annotate SNPs to genes up to 1 Mb away based on a significant eQTL association, i.e. where the expression of a gene is associated with the allelic variation, according to information from four brain-expression data repositories, including PsychENCORE(*46*), CommonMind Consortium(*42*), BRAINEAC(*152*) and GTEx v8 Brain(*43*). FUMA applied a FDR of 0.05 within each analysis to define significant eQTL associations.
3. Chromatin interaction mapping was used to map SNPs to genes on the basis of seven brain-related Hi-C chromatin conformation capture datasets: PsychENCORE EP link (one way)(*46*), PsychENCORE promoter anchored loops(*42*), HiC adult cortex(*49*), HiC fetal cortex(*49*), HiC (GSE87112) dorsolateral prefrontal cortex(*44*), HiC (GSE87112) hippocampus(*44*) and HiC (GSE87112) neural progenitor cells(*44*). We further selected only those genes for which one or both regions involved in the chromatin interaction overlapped with a predicted enhancer or promoter region (250 bp upstream and 500 bp downstream of the transcription start site) in any of the brain-related repositories from the Roadmap Epigenomics Project(*150*), that is; E053 (neurospheres) cortex, E054 (neurospheres) ganglion eminence, E067 (brain) angular gyrus, E068 (brain) anterior caudate, E069 (brain) cingulate gyrus, E070 (brain) germinal matrix, E071 (brain) hippocampus middle, E072 (brain) inferior temporal lobe, E073 (brain) dorsolateral prefrontal cortex, E074 (brain) substantia nigra, E081 (brain) fetal brain male, E082 (brain) fetal brain female, E003 embryonic stem (ES) H1 cells, E008 ES H9 cells, E007 (ES-derived) H1 derived neuronal progenitor cultured cells, E009 (ES-derived) H9 derived neuronal progenitor cultured cells and E010 (ES-derived) H9 derived neuron cultured cells. FUMA applied a FDR of 1×10^{-6} to identify significant chromatin interactions (default parameter), separately for each analysis.

Gene-based association analysis

MAGMA (v1.08)(*31*), with default parameters as implemented in FUMA (SNP-wise mean model), was used to test the joint association arising from all SNPs within a given gene (including 50 kb upstream to 50 kb downstream), while accounting for LD between SNPs. SNPs were mapped to 20,146 protein-coding genes on the

basis of National Center for Biotechnology Information build 37.3 gene definitions, and each gene was represented by at least one SNP. Bonferroni correction was applied for the number of genes ($p<0.025/20,146$), separately for each mvGWAS.

Gene-set enrichment analysis

MAGMA (v1.08), with default settings as implemented in FUMA, was used to examine the enrichment of association for predefined gene sets. This process tests whether gene-based p values among all 20,146 genes are lower for those genes within pre-defined functional sets than the rest of the genes in the genome, while correcting for other gene properties such as the number of SNPs. A total of 15,488 gene sets from the MSigDB database version 7.0(*55*) (5500 curated gene sets, 7343 gene ontology (GO) biological processes, 1644 GO molecular functions, and 1001 GO cellular components) were tested. Bonferroni correction was applied to correct for the number of gene sets ($p<0.05/15,488$), separately for each mvGWAS.

Cell-type-specific expression analysis in developing human cortex

Based on a linear regression model, the CELL TYPE function of FUMA was used to test whether gene-based association z-scores were positively associated with higher expression levels in certain cell types, based on single-cell RNA sequencing data from the developing human prefrontal cortex (GSE104276)(*57*). This dataset comprised 1) expression per cell type per age group, ranging from 8 to 26 postconceptional weeks, and 2) expression profiles per cell type, averaged over all ages combined. Results were considered significant if the association p values were smaller than the relevant Bonferroni threshold for the number of cell types/age groups. Analysis was performed separately for each mvGWAS.

Developmental stage analysis

We used MAGMA (default settings as implemented in FUMA) to examine whether lower gene-based association p values tended to be found for genes showing relatively higher expression in BrainSpan gene expression data(*56*) from any particular lifespan stage when contrasted with all other stages, separately for 29 different age groups ranging from 8 postconceptional weeks to 40 years old, and 11 defined lifespan stages from early prenatal to middle adulthood. A FDR threshold of 0.05 was applied separately for each analysis. Positive beta coefficients for this test indicate that genes showing more evidence for association are relatively upregulated on average at a given lifespan stage.

The Brainspan study originally collected and assigned human brain postmortem tissue samples to one of 31 developmental/lifespan stages(*56*), but FUMA's implementation excluded two age groups that had less than three samples each (i.e. 25 postconceptional weeks and 35 postconceptional weeks), resulting in 29 age groups being specified for this analysis by FUMA.

Polygenic disposition to brain-related disorders or behavioural traits

We used PRS-CS(*68*) to compute polygenic scores for 30,810 UK Biobank individuals (see Sample quality control) for each of the following brain-related disorders or behavioural traits, using GWAS summary statistics from previously published, large-scale studies: schizophrenia (n=82,315)(*69*), bipolar disorder (n=51,710)(*72*), autism (n=46,350)(*75*), attention deficit/hyperactivity disorder (n=55,374)(*77*), left-handedness (n=306,377)(*80*), Alzheimer's disease (n=63,926)(*82*), amyotrophic lateral sclerosis (n=152,268)(*85*) and epilepsy (n=44,889)(*88*). None of these previous studies used UK Biobank data, except for the GWAS of left-handedness(*80*) – however the individuals in that GWAS were selected to be non-overlapping and unrelated to those with brain image data from the February 2020 data release, so that none of the 30,810 UK Biobank individuals from the present study had been included in that GWAS. This ensured that training and target sets for polygenic score calculation were independent. PRS-CS infers posterior effect sizes of autosomal SNPs on the basis of genome-wide association summary statistics, within a high-dimensional Bayesian regression framework. We used default parameters and the recommended global effect size shrinkage parameter $\varphi=0.01$, together with linkage disequilibrium information based on the 1000 Genomes Project phase 3 European-descent reference panel(*153*). Polygenic scores were calculated using 1,097,390 SNPs for schizophrenia, 1,098,372 SNPs for bipolar disorder, 1,092,080 SNPs for

autism, 1,042,054 SNPs for attention deficit/hyperactivity disorder, 1,103,632 SNPs for left-handedness, 1,105,067 SNPs for Alzheimer's disease, 1,085,071 SNPs for amyotrophic lateral sclerosis, and 852,343 SNPs for epilepsy (these numbers came from 3-way overlaps between UK Biobank data, GWAS results, and 1000 Genomes data). PRS-CS has been shown to perform in a highly similar manner to other established polygenic risk algorithms, with noticeably better out-of-sample prediction than an approach based on p value thresholds and LD clumping(*154, 155*).

Polygenic scores were linearly adjusted for the effects of age, nonlinear age (i.e. $(\text{age-mean_age})^2$), assessment centre, genotype measurement batch, sex, and the first ten genetic principal components that capture population genetic diversity, before performing rank-based inverse normalization (i.e. the same set of covariate effects that the brain metrics were adjusted for - see Methods: Network construction and analysis) and visual inspection of their distributions to confirm normality. The adjusted and normalized polygenic scores were used as input for subsequent analyses.

Separately for polygenic scores for each disorder or behavioural trait, canonical correlation analysis across 30,810 participants ('*canoncorr*' function in MATLAB) was used to test multivariate association with the 90 heritable node-level connectivity measures (which had also been adjusted for covariates and normalized - see Network construction and analysis). This multivariate analysis identified a linear combination of the 90 node-level connectivity measures (i.e. a canonical variable) that maximally correlated with the polygenic score for a particular disorder or behavioural trait across participants. Separately for the polygenic score of each disorder or behavioural trait, the cross-participant Pearson correlation between each node-level connectivity and the canonical variable was used as a loading, reflecting the extent and direction of the contribution that a node-level connectivity made to a particular multivariate association.

We also assessed the pairwise correlations across individuals between adjusted and normalized polygenic scores for the different disorders and behavioural traits.

As the *APOE* locus is known to have a substantial effect on the risk of Alzheimer's disease, we also re-calculated polygenic scores for this disease after excluding a region from Chr19:45,116,911 to Chr19:46,318,605 (GRCh37)(*156*) around this locus, and repeated the residualization, normalization and canonical correlation analyses to check that the results stably reflected the polygenic contribution to risk.

Functional annotation of brain regions associated most strongly with polygenic scores

From each separate canonical correlation analysis of polygenic scores and node-level connectivity, we identified the regions showing loadings of >0.2 or <-0.2, which were then used to define a single mask in standard brain space (Montreal Neurological Institute space 152) (i.e. one mask for each polygenic score). Each mask was analyzed using the 'decoder' function of the Neurosynth database (http://neurosynth.org), a platform for large-scale synthesis of functional MRI data(*94*). This database defines brain-wide activation maps corresponding to specific cognitive or behavioural task terms using meta-analyzed functional activation maps. The database included 1,334 term-specific activation maps corresponding to cognitive or behavioural terms from 14,371 studies. Each mask that we created was used separately as input to define a brain-wide co-activation map based on all studies in the database. The resulting co-activation maps were then correlated with each of the 1,334 term-specific activation maps(*94*). We report only terms with correlation coefficients r>0.2, while excluding anatomical terms, nonspecific terms (e.g., 'Tasks'), and one from each pair of virtually duplicated terms (such as 'Words' and 'Word'). This method does not employ inferential statistical testing, but rather ranks terms based on the correlations between their activation maps and that of the input mask.

Data Availability

The primary data used in this study are available via the UK Biobank website www.ukbiobank.ac.uk. Specific UK Biobank data field codes are given in the Methods section. Other publicly available data sources and applications are cited in the Methods section. We have made our multivariate GWAS summary statistics available online within the GWAS catalog https://www.ebi.ac.uk/gwas/. Our heritability data can be visualized interactively using a dynamic web-based interface: https://immersive.erc.monash.edu/neuromarvl/?save=4563dcf5-d11e-4f89-9075-0c933e321de3_192.87.79.51

Code availability

This study used openly available software and codes, specifically GCTA (https://cnsgenomics.com/software/gcta/#GREML), MRtrix3 (https://www.mrtrix.org/), MOSTest (https://github.com/precimed/mostest), FUMA (https://fuma.ctglab.nl/), MAGMA (https://ctg.cncr.nl/software/magma, also implemented in FUMA), PRS-CS (https://github.com/getian107/PRScs) and Neurosynth (https://www.neurosynth.org/). In addition, we have made our pipeline codes available at the following URL: https://hdl.handle.net/1839/8d718e44-a1b7-4d6d-994e-4b0c83f63ec4

Disclosures

The authors declare no competing interests.

Acknowledgements

This research was funded by the Max Planck Society (Germany) and the Netherlands Organization for Scientific Research (Language in Interaction consortium: Gravitation grant number 024.001.006). The study was conducted using the UK Biobank resource under application no. 16066 with C.F. as the principal applicant. Our study made use of quality-controlled brain images generated by an image-processing pipeline developed and run on behalf of UK Biobank. The funders had no role in study design, data collection and analysis, the decision to publish or preparation of the manuscript.

References

1. D. S. Bassett, O. Sporns, Network neuroscience. *Nat Neurosci* **20**, 353-364 (2017).
2. H. Duffau, Stimulation mapping of white matter tracts to study brain functional connectivity. *Nat Rev Neurol* **11**, 255-265 (2015).
3. M. Thiebaut de Schotten, C. Foulon, P. Nachev, Brain disconnections link structural connectivity with function and behaviour. *Nat Commun* **11**, 5094 (2020).
4. T. E. Behrens *et al.*, Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat Neurosci* **6**, 750-757 (2003).
5. R. E. Passingham, K. E. Stephan, R. Kotter, The anatomical basis of functional localization in the cortex. *Nat Rev Neurosci* **3**, 606-616 (2002).
6. M. P. van den Heuvel, O. Sporns, A cross-disorder connectome landscape of brain dysconnectivity. *Nat Rev Neurosci* **20**, 435-446 (2019).
7. N. A. Crossley *et al.*, The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382-2395 (2014).
8. S. C. de Lange *et al.*, Shared vulnerability for connectome alterations across psychiatric and neurological brain disorders. *Nat Hum Behav* **3**, 988-998 (2019).
9. R. Schulz *et al.*, Assessing the integrity of corticospinal pathways from primary and secondary cortical motor areas after stroke. *Stroke* **43**, 2248-2251 (2012).
10. C. Grefkes, G. R. Fink, Connectivity-based approaches in stroke and recovery of function. *Lancet Neurol* **13**, 206-216 (2014).
11. S. Pajevic, C. Pierpaoli, Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: application to white matter fiber tract mapping in the human brain. *Magn Reson Med* **42**, 526-540 (1999).
12. A. L. Alexander, J. E. Lee, M. Lazar, A. S. Field, Diffusion tensor imaging of the brain. *Neurotherapeutics* **4**, 316-329 (2007).
13. P. J. Basser, J. Mattiello, D. LeBihan, MR diffusion tensor spectroscopy and imaging. *Biophys J* **66**, 259-267 (1994).
14. P. J. Basser, J. Mattiello, D. LeBihan, Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B* **103**, 247-254 (1994).
15. C. Beaulieu, P. S. Allen, Determinants of anisotropic water diffusion in nerves. *Magn Reson Med* **31**, 394-400 (1994).
16. P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, A. Aldroubi, In vivo fiber tractography using DT-MRI data. *Magn Reson Med* **44**, 625-632 (2000).
17. B. Zhao *et al.*, Common genetic variation influencing human white matter microstructure. *Science* **372**, (2021).
18. B. Zhao *et al.*, Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Mol Psychiatry* **26**, 3943-3955 (2021).
19. B. Mazoyer *et al.*, Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness. *PLoS One* **9**, e101165 (2014).
20. N. Tzourio-Mazoyer *et al.*, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273-289 (2002).
21. J. McFadyen, R. J. Dolan, M. I. Garrido, The influence of subcortical shortcuts on disordered sensory and cognitive processing. *Nat Rev Neurosci* **21**, 264-276 (2020).
22. M. P. van den Heuvel, O. Sporns, Network hubs in the human brain. *Trends Cogn Sci* **17**, 683-696 (2013).
23. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).

24. D. van der Meer *et al.*, Understanding the genetic determinants of the brain with MOSTest. *Nat Commun* **11**, 3512 (2020).
25. K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
26. J. MacArthur *et al.*, The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901 (2017).
27. B. Zhao *et al.*, Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet* **51**, 1637-1644 (2019).
28. L. T. Elliott *et al.*, Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216 (2018).
29. K. L. Grasby *et al.*, The genetic architecture of the human cerebral cortex. *Science* **367**, (2020).
30. D. van der Meer *et al.*, Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. *Mol Psychiatry* **25**, 3053-3065 (2020).
31. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
32. E. Persyn *et al.*, Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat Commun* **11**, 2175 (2020).
33. J. G. Thorp *et al.*, Symptom-level modelling unravels the shared genetic architecture of anxiety and depression. *Nat Hum Behav* **5**, 1432-1442 (2021).
34. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* **8**, 21 (2017).
35. J. E. Savage *et al.*, Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* **50**, 912-919 (2018).
36. M. Lam *et al.*, Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific Neural Expression and Potential Nootropic Drug Targets. *Cell Rep* **21**, 2597-2613 (2017).
37. G. W. Beecham *et al.*, Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet* **10**, e1004606 (2014).
38. EPICURE Consortium *et al.*, Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. *Hum Mol Genet* **21**, 5359-5372 (2012).
39. T. G. Richardson *et al.*, Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS medicine* **17**, e1003062 (2020).
40. S. L. Pulit *et al.*, Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* **28**, 166-174 (2019).
41. A. Mela, J. E. Goldman, The tetraspanin KAI1/CD82 is expressed by late-lineage oligodendrocyte precursors and may function to restrict precursor migration and promote oligodendrocyte differentiation and myelination. *Journal of Neuroscience* **29**, 11172-11181 (2009).
42. M. Fromer *et al.*, Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* **19**, 1442-1453 (2016).
43. The GTEX Consortium *et al.*, The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).
44. A. D. Schmitt *et al.*, A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep* **17**, 2042-2059 (2016).
45. T. Srivastava *et al.*, A TLR/AKT/FoxO3 immune tolerance-like pathway disrupts the repair capacity of oligodendrocyte progenitors. *J Clin Invest* **128**, 2025-2041 (2018).

46. D. Wang *et al.*, Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, (2018).
47. M. P. Coba *et al.*, TNiK is required for postsynaptic and nuclear signaling pathways and cognitive function. *Journal of Neuroscience* **32**, 13987-13999 (2012).
48. S. Anazi *et al.*, A null mutation in TNIK defines a novel locus for intellectual disability. *Hum Genet* **135**, 773-778 (2016).
49. P. Giusti-Rodríguez *et al.*, Using three-dimensional regulatory chromatin interactions from adult and fetal cortex to interpret genetic results for psychiatric disorders and cognitive traits. *BioRxiv*, 406330 (2019).
50. H. Nawabi *et al.*, Doublecortin-Like Kinases Promote Neuronal Survival and Induce Growth Cone Reformation via Distinct Mechanisms. *Neuron* **88**, 704-719 (2015).
51. T. Pramparo, Y. H. Youn, J. Yingling, S. Hirotsune, A. Wynshaw-Boris, Novel embryonic neuronal migration and proliferation defects in Dcx mutant mice are exacerbated by Lis1 reduction. *J Neurosci* **30**, 3002-3012 (2010).
52. R. G. Mealer *et al.*, Glycobiology and schizophrenia: a biological hypothesis emerging from genomic research. *Mol Psychiatry* **25**, 3129-3139 (2020).
53. Q. Luo *et al.*, Association of a Schizophrenia-Risk Nonsynonymous Variant With Putamen Volume in Adolescents: A Voxelwise and Genome-Wide Association Study. *JAMA Psychiatry* **76**, 435-445 (2019).
54. K. E. Wahlberg *et al.*, Polymorphisms in Manganese Transporters SLC30A10 and SLC39A8 Are Associated With Children's Neurodevelopment by Influencing Manganese Homeostasis. *Front Genet* **9**, 664 (2018).
55. A. Liberzon *et al.*, Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011).
56. J. A. Miller *et al.*, Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206 (2014).
57. S. Zhong *et al.*, A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524-528 (2018).
58. A. D. Friederici, S. M. Gierhan, The language network. *Curr Opin Neurobiol* **23**, 250-254 (2013).
59. B. Lipkin *et al.*, Probabilistic atlas for the language network based on precision fMRI data from> 800 individuals. *Scientific Data* **9,** 529 (2022).
60. S. Malik-Moraleda *et al.*, An investigation across 45 languages and 12 language families reveals a universal language network. *Nat Neurosci*, **25**, 1014-1019 (2022).
61. L. Labache *et al.*, A SENtence Supramodal Areas AtlaS (SENSAAS) based on multiple task-induced activation mapping and graph analysis of intrinsic connectivity in 144 healthy right-handers. *Brain Struct Funct* **224**, 859-882 (2019).
62. A. S. Dick, P. Tremblay, Beyond the arcuate fasciculus: consensus and controversy in the connectional anatomy of language. *Brain* **135**, 3529-3550 (2012).
63. H. Duffau, P. Gatignol, S. Moritz-Gasser, E. Mandonnet, Is the left uncinate fasciculus essential for language? *Journal of neurology* **256**, 382-389 (2009).
64. J. G. Flanagan, P. Vanderhaeghen, The ephrins and Eph receptors in neural development. *Annu Rev Neurosci* **21**, 309-345 (1998).
65. Y. Mekki *et al.*, The genetic architecture of language functional connectivity. *Neuroimage* **249**, 118795 (2022).
66. F. Zheng *et al.*, GABA Signaling Pathway-associated Gene PLCL1 Rare Variants May be Associated with Autism Spectrum Disorders. *Neurosci Bull* **37**, 1240-1245 (2021).
67. S. A. Graham, S. E. Fisher, Understanding Language from a Genomic Perspective. *Annu Rev Genet* **49**, 131-160 (2015).
68. T. Ge, C. Y. Chen, Y. Ni, Y. A. Feng, J. W. Smoller, Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* **10**, 1776 (2019).
69. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).

70. Z. Sha, D. Schijven, C. Francks, Patterns of brain asymmetry associated with polygenic risks for autism and schizophrenia implicate language and executive functions but not brain masculinization. *Mol Psychiatry* **26**, 7652–7660 (2021).
71. E.-M. Stauffer *et al.*, Grey and white matter microstructure is associated with polygenic risk for schizophrenia. *Molecular Psychiatry*, 1-10 (2021).
72. E. A. Stahl *et al.*, Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**, 793-803 (2019).
73. C. M. Torgerson *et al.*, DTI tractography and white matter fiber tract characteristics in euthymic bipolar I patients and healthy control subjects. *Brain Imaging Behav* **7**, 129-139 (2013).
74. E. Ji *et al.*, Increased and Decreased Superficial White Matter Structural Connectivity in Schizophrenia and Bipolar Disorder. *Schizophr Bull* **45**, 1367-1378 (2019).
75. J. Grove *et al.*, Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444 (2019).
76. Z. Sha *et al.*, The genetic architecture of structural left-right asymmetry of the human brain. *Nat Hum Behav* **5**, 1226-1239 (2021).
77. D. Demontis *et al.*, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75 (2019).
78. M. N. Pavuluri *et al.*, Diffusion tensor imaging study of white matter fiber tracts in pediatric bipolar disorder and attention-deficit/hyperactivity disorder. *Biol Psychiatry* **65**, 586-593 (2009).
79. R. Beare *et al.*, Altered structural connectivity in ADHD: a network based analysis. *Brain Imaging Behav* **11**, 846-858 (2017).
80. Z. Sha *et al.*, Handedness and its genetic influences are associated with structural asymmetries of the cerebral cortex in 31,864 individuals. *Proc Natl Acad Sci U S A* **118**, (2021).
81. A. Wiberg *et al.*, Handedness, language areas and neuropsychiatric diseases: insights from brain imaging and genetics. *Brain* **142**, 2938-2947 (2019).
82. B. W. Kunkle *et al.*, Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
83. C. Y. Lo *et al.*, Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease. *J Neurosci* **30**, 16876-16885 (2010).
84. D. J. Phillips *et al.*, Graph theoretic analysis of structural connectivity across the spectrum of Alzheimer's disease: The importance of graph creation methods. *Neuroimage Clin* **7**, 377-390 (2015).
85. W. van Rheenen *et al.*, Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat Genet* **53**, 1636-1648 (2021).
86. S. Rose *et al.*, Direct evidence of intra- and interhemispheric corticomotor network degeneration in amyotrophic lateral sclerosis: an automated MRI structural connectivity study. *Neuroimage* **59**, 2661-2669 (2012).
87. G. Douaud, N. Filippini, S. Knight, K. Talbot, M. R. Turner, Integration of structural and functional magnetic resonance imaging in amyotrophic lateral sclerosis. *Brain* **134**, 3470-3479 (2011).
88. E. International League Against Epilepsy Consortium on Complex, Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat Commun* **9**, 5269 (2018).
89. J. Engel, Jr. *et al.*, Connectomics and epilepsy. *Curr Opin Neurol* **26**, 186-194 (2013).
90. P. Besson *et al.*, Structural connectivity differences in left and right temporal lobe epilepsy. *Neuroimage* **100**, 135-144 (2014).
91. W. W. Seeley, R. K. Crawford, J. Zhou, B. L. Miller, M. D. Greicius, Neurodegenerative diseases target large-scale human brain networks. *Neuron* **62**, 42-52 (2009).

92. P. L. Lee *et al.*, Posterior Cingulate Cortex Network Predicts Alzheimer's Disease Progression. *Front Aging Neurosci* **12**, 608667 (2020).
93. D. Berron, D. van Westen, R. Ossenkoppele, O. Strandberg, O. Hansson, Medial temporal lobe connectivity and its associations with cognition in early Alzheimer's disease. *Brain* **143**, 1233-1248 (2020).
94. T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, T. D. Wager, Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* **8**, 665-670 (2011).
95. L. Ben Haim, D. H. Rowitch, Functional diversity of astrocytes in neural circuit regulation. *Nat Rev Neurosci* **18**, 31-41 (2017).
96. A. V. Molofsky *et al.*, Astrocyte-encoded positional cues maintain sensorimotor circuit integrity. *Nature* **509**, 189-194 (2014).
97. A. Badimon *et al.*, Negative feedback control of neuronal activity by microglia. *Nature* **586**, 417-423 (2020).
98. K. Reemst, S. C. Noctor, P. J. Lucassen, E. M. Hol, The Indispensable Roles of Microglia and Astrocytes during Brain Development. *Front Hum Neurosci* **10**, 566 (2016).
99. P. Squarzoni *et al.*, Microglia modulate wiring of the embryonic forebrain. *Cell Rep* **8**, 1271-1279 (2014).
100. L. Pont-Lezica *et al.*, Microglia shape corpus callosum axon tract fasciculation: functional impact of prenatal inflammation. *Eur J Neurosci* **39**, 1551-1557 (2014).
101. I. Sommer, N. Ramsey, R. Kahn, A. Aleman, A. Bouma, Handedness, language lateralisation and anatomical asymmetry in schizophrenia: meta-analysis. *Br J Psychiatry* **178**, 344-351 (2001).
102. M. Hirnstein, K. Hugdahl, Excess of non-right-handedness in schizophrenia: meta-analysis of gender effects and potential biases in handedness assessment. *Br J Psychiatry* **205**, 260-267 (2014).
103. D. Bendetowicz *et al.*, Two critical brain networks for generation and combination of remote associations. *Brain* **141**, 217-233 (2018).
104. M. Daianu *et al.*, Breakdown of brain connectivity between normal aging and Alzheimer's disease: a structural k-core network analysis. *Brain connectivity* **3**, 407-422 (2013).
105. F. G. Hillary, J. H. Grafman, Injured Brains and Adaptive Networks: The Benefits and Costs of Hyperconnectivity. *Trends Cogn Sci* **21**, 385-401 (2017).
106. A. Fry *et al.*, Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* **186**, 1026-1034 (2017).
107. P. Hagoort, Nodes and networks in the neural architecture for language: Broca's region and beyond. *Curr Opin Neurobiol* **28**, 136-141 (2014).
108. T. E. Behrens *et al.*, Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn Reson Med* **50**, 1077-1088 (2003).
109. T. E. Behrens, H. J. Berg, S. Jbabdi, M. F. Rushworth, M. W. Woolrich, Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage* **34**, 144-155 (2007).
110. A. Zalesky *et al.*, Connectome sensitivity or specificity: which is more important? *Neuroimage* **142**, 407-420 (2016).
111. T. Sarwar, K. Ramamohanarao, A. Zalesky, Mapping connectomes with diffusion MRI: deterministic or probabilistic tractography? *Magn Reson Med* **81**, 1368-1384 (2019).
112. M. A. de Reus, M. P. van den Heuvel, Estimating false positives and negatives in brain networks. *Neuroimage* **70**, 402-409 (2013).
113. L. Petit *et al.*, The structural connectivity of the human angular gyrus as revealed by microdissection and diffusion tractography. *Brain Struct Funct*, (2022).
114. G. Theaud *et al.*, TractoFlow: A robust, efficient and reproducible diffusion MRI pipeline leveraging Nextflow & Singularity. *Neuroimage* **218**, 116889 (2020).
115. T. Zhao *et al.*, Age-related changes in the topological organization of the white matter structural connectome across the human lifespan. *Hum Brain Mapp* **36**, 3777-3792 (2015).

116. A. Griffa, P. S. Baumann, J. P. Thiran, P. Hagmann, Structural connectomics in brain diseases. *Neuroimage* **80**, 515-526 (2013).
117. K. H. Maier-Hein *et al.*, The challenge of mapping the human connectome based on diffusion tractography. *Nat Commun* **8**, 1349 (2017).
118. R. Krupnik, Y. Yovel, Y. Assaf, Inner Hemispheric and Interhemispheric Connectivity Balance in the Human Brain. *J Neurosci* **41**, 8351-8361 (2021).
119. V. R. Karolis, M. Corbetta, M. Thiebaut de Schotten, The architecture of functional lateralisation and its relationship to callosal connectivity in the human brain. *Nat Commun* **10**, 1417 (2019).
120. K. Jarbo, T. Verstynen, W. Schneider, In vivo quantification of global connectivity in the human corpus callosum. *Neuroimage* **59**, 1988-1996 (2012).
121. B. Y. Lee, X. H. Zhu, X. Li, W. Chen, High-resolution imaging of distinct human corpus callosum microstructure and topography of structural connectivity to cortices at high field. *Brain Struct Funct* **224**, 949-960 (2019).
122. J. D. Schmahmann, J. Schmahmann, D. Pandya, *Fiber pathways of the brain*. (OUP USA, 2009).
123. F. Bai *et al.*, Topologically convergent and divergent structural connectivity patterns between patients with remitted geriatric depression and amnestic mild cognitive impairment. *J Neurosci* **32**, 4307-4318 (2012).
124. H. Huang *et al.*, Development of human brain structural networks through infancy and childhood. *Cereb Cortex* **25**, 1389-1404 (2015).
125. D. Joel *et al.*, Sex beyond the genitalia: The human brain mosaic. *Proc Natl Acad Sci U S A* **112**, 15468-15473 (2015).
126. J. Shen, D. J. Tozer, H. S. Markus, J. Tay, Network Efficiency Mediates the Relationship Between Vascular Burden and Cognitive Impairment: A Diffusion Tensor Imaging Study in UK Biobank. *Stroke* **51**, 1682-1689 (2020).
127. E. St-Onge, A. Daducci, G. Girard, M. Descoteaux, Surface-enhanced tractography (SET). *Neuroimage* **169**, 524-539 (2018).
128. D. Shastin *et al.*, Surface-based tracking for short association fibre tractography. *Neuroimage* **260**, 119423 (2022).
129. S. M. Smith *et al.*, An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci* **24**, 737-745 (2021).
130. J. E. Huffman, Examining the current standards for genetic discovery and replication in the era of mega-biobanks. *Nat Commun* **9**, 5054 (2018).
131. C. Sudlow *et al.*, UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
132. C. Bycroft *et al.*, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
133. S. Purcell *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).
134. F. Alfaro-Almagro *et al.*, Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400-424 (2018).
135. K. L. Miller *et al.*, Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* **19**, 1523-1536 (2016).
136. J. D. Tournier, F. Calamante, A. Connelly, MRtrix: diffusion tractography in crossing fiber regions. *International journal of imaging systems and technology* **22**, 53-66 (2012).
137. P. Hagmann *et al.*, Mapping human whole-brain structural networks with diffusion MRI. *PLoS One* **2**, e597 (2007).
138. S. M. Smith *et al.*, Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23 Suppl 1**, S208-219 (2004).

139. Y. Sun, J. Li, J. Suckling, L. Feng, Asymmetry of Hemispheric Network Topology Reveals Dissociable Processes between Functional and Structural Brain Connectome in Community-Living Elders. *Front Aging Neurosci* **9**, 361 (2017).
140. E. Calabrese, A. Badea, G. Cofer, Y. Qi, G. A. Johnson, A Diffusion MRI Tractography Connectome of the Mouse Brain and Comparison with Neuronal Tracer Data. *Cereb Cortex* **25**, 4628-4637 (2015).
141. N. Ratnarajah *et al.*, Structural connectivity asymmetry in the neonatal brain. *Neuroimage* **75**, 187-194 (2013).
142. P. E. Shrout, J. L. Fleiss, Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* **86**, 420-428 (1979).
143. C. J. Hoggart, T. G. Clark, M. De Iorio, J. C. Whittaker, D. J. Balding, Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol* **32**, 179-185 (2008).
144. O. A. Panagiotou, J. P. Ioannidis, Genome-Wide Significance Project, What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *International journal of epidemiology* **41**, 273-286 (2012).
145. W. Zhang *et al.*, Comparing genetic variants detected in the 1000 genomes project with SNPs determined by the International HapMap Consortium. *J Genet* **94**, 731-740 (2015).
146. K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
147. P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894 (2019).
148. A. P. Boyle *et al.*, Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-1797 (2012).
149. J. Ernst, M. Kellis, ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216 (2012).
150. Roadmap Epigenomics Consortium *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
151. M. Kircher *et al.*, A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-315 (2014).
152. A. Ramasamy *et al.*, Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* **17**, 1418-1428 (2014).
153. The 1000 Genomes Project Consortium *et al.*, A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
154. G. Ni *et al.*, A comparison of ten polygenic score methods for psychiatric disorders applied across multiple cohorts. *Biological psychiatry* **90**, 611-620 (2021).
155. A. B. Zheutlin *et al.*, Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. *Am J Psychiatry* **176**, 846-855 (2019).
156. H. Foo *et al.*, Associations between Alzheimer's disease polygenic risk scores and hippocampal subfield volumes in 17,161 UK Biobank participants. *Neurobiology of Aging* **98**, 108-115 (2021).