

Genetic architecture of the structural connectome

Received: 13 September 2022

Accepted: 12 February 2024

Published online: 04 March 2024

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Myelinated axons form long-range connections that enable rapid communication between distant brain regions, but how genetics governs the strength and organization of these connections remains unclear. We perform genome-wide association studies of 206 structural connectivity measures derived from diffusion magnetic resonance imaging tractography of 26,333 UK Biobank participants, each representing the density of myelinated connections within or between a pair of cortical networks, subcortical structures or cortical hemispheres. We identify 30 independent genome-wide significant variants after Bonferroni correction for the number of measures studied (126 variants at nominal genome-wide significance) implicating genes involved in myelination (*SEMA3A*), neurite elongation and guidance (*NUAK1*, *STRN*, *DPYSL2*, *EPHA3*, *SEMA3A*, *HGF*, *SHTNI*), neural cell proliferation and differentiation (*GMNC*, *CEL4*, *HGF*), neuronal migration (*CCDC88C*), cytoskeletal organization (*CTTNBP2*, *MAPT*, *DAAMI*, *MYO16*, *PLEC*), and brain metal transport (*SLC39A8*). These variants have four broad patterns of spatial association with structural connectivity: some have disproportionately strong associations with corticothalamic connectivity, interhemispheric connectivity, or both, while others are more spatially diffuse. Structural connectivity measures are highly polygenic, with a median of 9.1 percent of common variants estimated to have non-zero effects on each measure, and exhibited signatures of negative selection. Structural connectivity measures have significant genetic correlations with a variety of neuropsychiatric and cognitive traits, indicating that connectivity-altering variants tend to influence brain health and cognitive function. Heritability is enriched in regions with increased chromatin accessibility in adult oligodendrocytes (as well as microglia, inhibitory neurons and astrocytes) and multiple fetal cell types, suggesting that genetic control of structural connectivity is partially mediated by effects on myelination and early brain development. Our results indicate pervasive, pleiotropic, and spatially structured genetic control of white-matter structural connectivity via diverse neurodevelopmental pathways, and support the relevance of this genetic control to healthy brain function.

The human brain's volume is about equal parts gray matter and white matter. Unlike gray matter, which mainly contains densely packed cell bodies of neurons and glia, white matter is almost exclusively made up of bundles of myelinated axons¹. The primary purpose of myelin is to

considerably speed up action potential conduction along axons, from 0.5–10 m/s to ~150 m/s², enabling axons (nerve fibers) traveling through the white matter to efficiently transmit information between distant brain regions. White matter fibers are key constituents of the

✉ e-mail: m.wainberg@utoronto.ca; colin.hawco@camh.ca; shreejoy.tripathy@camh.ca

brain's structural connectome, the complete set of anatomical connections between brain cells³. Structural connectivity is a fundamental organizational property of the brain^{3,4}.

Microstructural properties of the white matter can be quantified non-invasively in vivo via diffusion magnetic resonance imaging (dMRI). This is because the myelin around an axon, as well as the axonal membranes themselves and the coherence or density (organization) of axons within fiber bundles, constitute physical barriers that restrict the diffusion of water molecules, so that they diffuse more readily along the axon than perpendicular to it, resulting in a distortion of the MRI signal when a directional magnetic field gradient is applied^{5–7}.

Efforts such as The Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA) Consortium⁸ and the UK Biobank⁹ have gathered large-scale cohorts with both brain MRI and genetic data. Genome-wide association studies (GWAS) of MRI phenotypes^{10–26} based on these efforts have elucidated the genetic architecture of brain structure and function with unprecedented detail. Most white-matter GWAS to date^{13,14,18,19} have relied upon tract-based spatial statistics (TBSS) (13), averaging white matter microstructural properties across entire tracts, such as the corpus callosum. TBSS effectively indexes general white-matter health²⁷, but does not explicitly consider how brain regions connect to each other.

Structural connectivity can be evaluated using a technique known as fiber tracking or tractography^{28–30}, which constructs a model of the paths of white matter fibers based on the directionality of water diffusion through the white matter. By overlaying a brain atlas of one's choice, the connectivity between pairs of brain regions can be quantified. Thus, unlike TBSS, tractography has the ability to quantify the structural connectome, providing rich detail on the whole-brain organizational properties relating to complex processes such as cognition³¹ and functional connectivity³². Tractography has implicated connectomic disturbances in diverse neuropsychiatric disorders³³, including schizophrenia^{34–46}, bipolar disorder^{41,44,47,48}, major depressive disorder⁴⁹, attention-deficit hyperactivity disorder^{50–52}, autism^{53–55}, Alzheimer's disease^{56–58}, and multiple sclerosis^{59–66}.

Here, we expand upon prior white matter GWAS to examine 206 tractography-based structural connectivity measures representing the rich organizational structure of the brain. Each measure quantifies the density of white matter fibers within or between a pair of well-established large-scale cortical brain networks⁶⁷, subcortical structures, or hemispheres. Our primary analysis (summarized in Fig. 1) encompasses 26,333 participants from the UK Biobank with diffusion MRI scans – two orders of magnitude larger than the only prior GWAS of structural connectivity⁶⁸, which studied 366 participants.

Results

Tractography across 26,333 UK Biobank participants

We established a tractography pipeline to infer white-matter structural connectomes from 26,333 UK Biobank participants (2400 with replicate scans, which were not used for the GWAS) with structural and diffusion MRI scans passing quality control (see “Cohort selection and MRI quality control”, Methods). Participants were 53% female and aged 40–70 (median 55) at the time of their first scan. dMRI data were extensively pre processed according to the UKB pipelines prior to use (biobank.ctsu.ox.ac.uk/crystal/ukb/docs/brain_mri.pdf). To obtain connectomes for each participant, we used the MRtrix3 diffusion MRI software package⁶⁹ to: (1) estimate a fiber orientation distribution via multi-shell multi-tissue constrained spherical deconvolution of the diffusion MRI scan⁷⁰; (2) perform anatomically constrained probabilistic tractography via second-order integration⁷¹, selecting 1 million streamlines seeded from the gray/white matter interface; (3) weight each streamline via the SIFT2 algorithm⁷² so that the density of reconstructed connections better reflects the density of the underlying white matter fibers; and (4) generate a symmetric 214 × 214

connectome matrix (Fig. 2A) denoting the density of streamlines connecting each pair of parcels in the 200-parcel Schaefer cortical atlas⁷³ plus 14 subcortical parcels from the Harvard-Oxford atlas (Supplementary Data 4), scaled by the average length of the connecting streamlines and divided by the product of the two parcel volumes.

Rather than performing 22,791 GWAS, one for each pair of parcels, we opted to reduce the dimensionality of each participant's connectome matrix in a biologically interpretable way. We derived three types of measures, for a total of 206 measures (Fig. 2B): (1) hemisphere-level cortical-to-cortical connectivity (3 measures: left intra-hemisphere, right intra-hemisphere and inter-hemisphere), (2) network-level cortical-to-cortical connectivity within and between each of the 14 hemisphere-specific “Yeo 7” networks⁶⁷ (105 measures, e.g. left-hemisphere visual, “LH Vis”, to right-hemisphere somato-motor, “RH SomMot”), and (3) cortical-to-subcortical connectivity between each of these 14 “Yeo 7” networks and 7 subcortical structures: thalamus, caudate, putamen, pallidum, hippocampus, amygdala, and accumbens (98 measures). These 206 measures were moderately replicable (median scan-rescan intraclass correlation coefficient = 0.67) across the 2400 participants with replicate scans, taken 1.0–5.3 years (median 2.2) after the original scan (Fig. 2C).

Variants associated with structural connectivity

We performed genome-wide association studies on these 206 connectivity measures using the regenie genetic analysis toolkit⁷⁴. We analyzed 9,423,516 variants present in the UK Biobank's imputed genotypes (imputed to a combination of the Haplotype Reference Consortium⁷⁵, UK10K⁷⁶ and 1000 Genomes Phase 3⁷⁷ cohorts) which had at least 1% minor allele frequency and passed quality control (see Methods). We covaried for age, sex, age × sex, age², age² × sex, genotyping array, scanner site ($n = 22$ sites), total intracranial volume, and the top 10 genotype principal components. We identified 30 genome-wide significant loci (Table 1, Fig. 1, Fig. 3) associated with at least one of the 206 measures after Bonferroni correction for the number of measures tested (i.e. at $p < 5 \times 10^{-8} / 206 \approx 2.4 \times 10^{-10}$). We identified 126 loci (Fig. 1, Fig. 3, Supplementary Data 1) at the less stringent threshold of nominal genome-wide significance ($p < 5 \times 10^{-8}$). In a replication analysis of these 126 loci in 665 participants of non-European genetic ancestry, 75 of the 126 lead variants had at least 1% frequency and passed quality control, and these variants were 2.7 times more likely to have the same direction of effect than expected by chance (Fisher $p = 0.038$; Supplementary Data 2). The one association from the prior structural connectivity GWAS mentioned in the introduction⁶⁸, the intronic variant rs2618516 in *SPONI*, failed to replicate in the European cohort (uncorrected $p > 0.006$ for all 206 measures).

Two broad patterns of spatial association with structural connectivity

We next explored the spatial pattern of association of each of the 30 lead variants across the 206 measures (Fig. 4, Supplementary Fig. 1). We annotated associations passing Bonferroni-corrected genome-wide significance ($p < 5 \times 10^{-8} / 206 \approx 2.4 \times 10^{-10}$) with asterisks (*). We characterized the pleiotropic effects of these 30 loci on the 206 measures, adopting a Bonferroni correction for just the 30 variants times the number of measures ($p < 0.05 / 30 / 206 \approx 8.1 \times 10^{-6}$), indicated with a dot (·) in Fig. 4. At this relaxed threshold, we observed substantial pleiotropy of our 30 lead variants across the 206 measures. The median lead variant was associated with 7.5 of the 206 measures, and the most pleiotropic variant, rs2796245 (near *CD34*), was associated with 53.

We observed four broad patterns of spatial association with structural connectivity (Fig. 1). The most significant of the 30 lead variants, near *CCDC88C*, was almost exclusively associated with corticothalamic connectivity, i.e. connectivity between various cortical

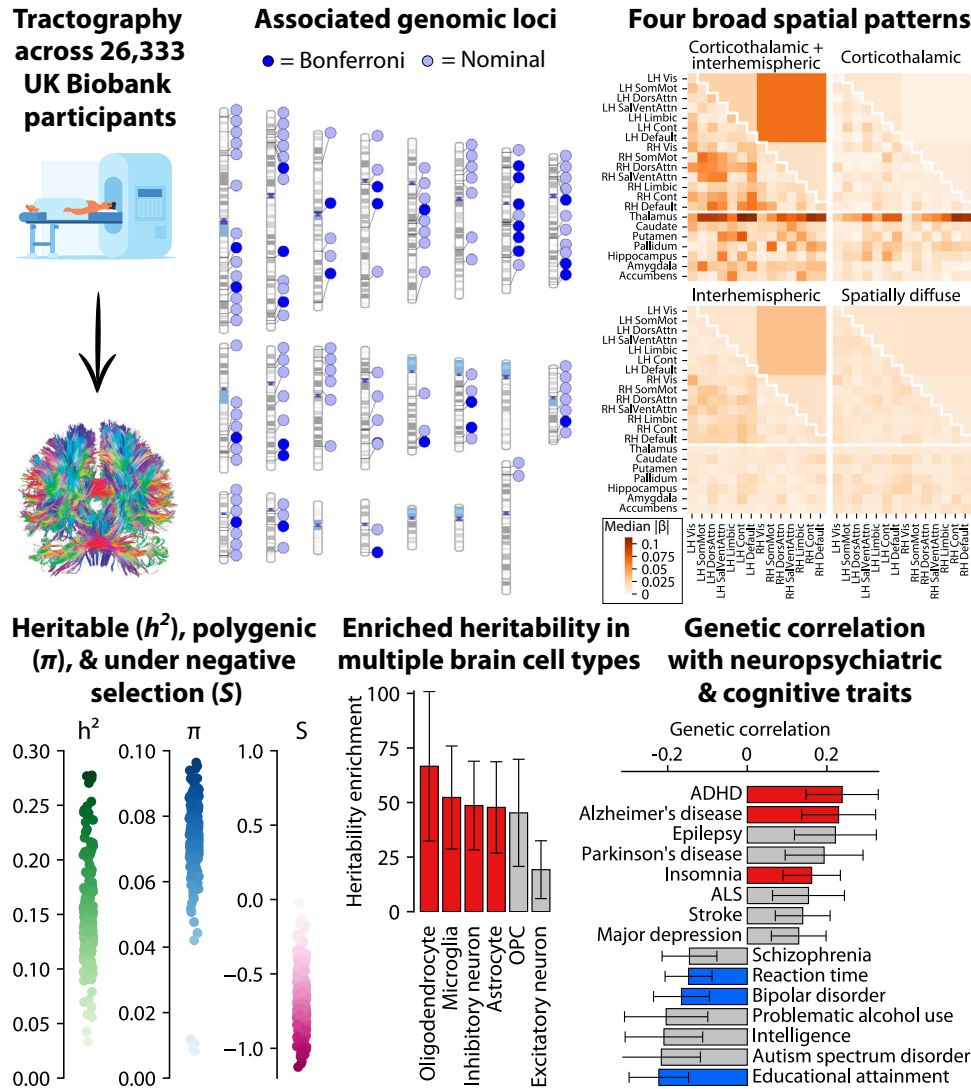


Fig. 1 | Study overview. Top left: Measurement of tractograms from participants with brain imaging data in the UK Biobank. Top middle: genomic locations of common genetic variants associated with structural connectivity, with chromosomes in ascending order (first row: chr1-8; second row: chr9-16; third row: chr17-22 and chrX). 30 independent variants were genome-wide significant after Bonferroni correction for the 206 structural connectivity measures studied (dark blue), and 96 more reached nominal genome-wide significance (light blue). Top right: the 30 variants cluster into four broad patterns of spatial association: corticothalamic and interhemispheric (3 variants), corticothalamic only (1 variant), interhemispheric only (8 variants), and spatially diffuse (18 variants). Heatmaps denote the median

effect size magnitude ($|\beta|$) of variants with each pattern on each of the 206 structural connectivity measures. Bottom left: heritability (h^2), polygenicity (π), and selection parameter (S) for each of the 206 measures. Bottom middle and right: heritability enrichments for regions with increased chromatin accessibility in each of 6 adult brain cell types, and genetic correlations with 15 brain-related traits. Bars show maximums across the 206 measures; red/blue bars are significant after Bonferroni correction. OPC oligodendrocyte progenitor cell, ADHD attention-deficit/hyperactivity disorder, ALS amyotrophic lateral sclerosis, disord. disorder. Error bars represent 95% confidence intervals.

networks and the thalamus. Three others, near *GMNC*, *NUAK1*, and *INPP5D*, were primarily associated with inter-hemisphere cortical connections as well as corticothalamic ones (even though in the case of *INPP5D*, no inter-hemisphere connections reached significance). This inter-hemisphere pattern could exist in part because inter-hemisphere connections are more reliably detected, as reflected by their higher replicability (Fig. 2C). Eight others, near *EEF1AKMT2*, *STRN*, *CTTNBP2*, *MYO16*, *PLEC*, *SLC4A10*, *SLC45A4*, and *ENSG00000282278*, showed the inter-hemisphere pattern but not the corticothalamic one. The remaining 18 variants tended to have spatially diffuse patterns of association.

Other, more subtle patterns were also present. For instance, the *SLC39A8* variant rs13107325, uniquely among the 30 variants, was markedly associated with cortico-hippocampal, and to a lesser extent cortico-amygdalar, connectivity relative to all other measures.

SLC39A8's unique pattern of association may reflect its unique mechanism as a brain metal transporter (see below).

Associated variants implicate genes with neurodevelopmental roles

17 of the 30 lead variants had nearest genes with direct biological relevance to neurodevelopment. The most significant association was with rs941760 (chr14:91,881,751), an intronic variant in *CCDC88C*. *CCDC88C* loss-of-function variants disrupt Wnt signaling, a core developmental signaling pathway⁷⁸, and cause autosomal recessive hydrocephalus, sometimes accompanied by periventricular neuronal heterotopias (brain malformations resulting from abnormal neuronal migration), developmental delay and seizures⁷⁹⁻⁸¹. (Another of the 30 lead variants has *WNT16*, a member of the Wnt family, as the nearest gene, but the

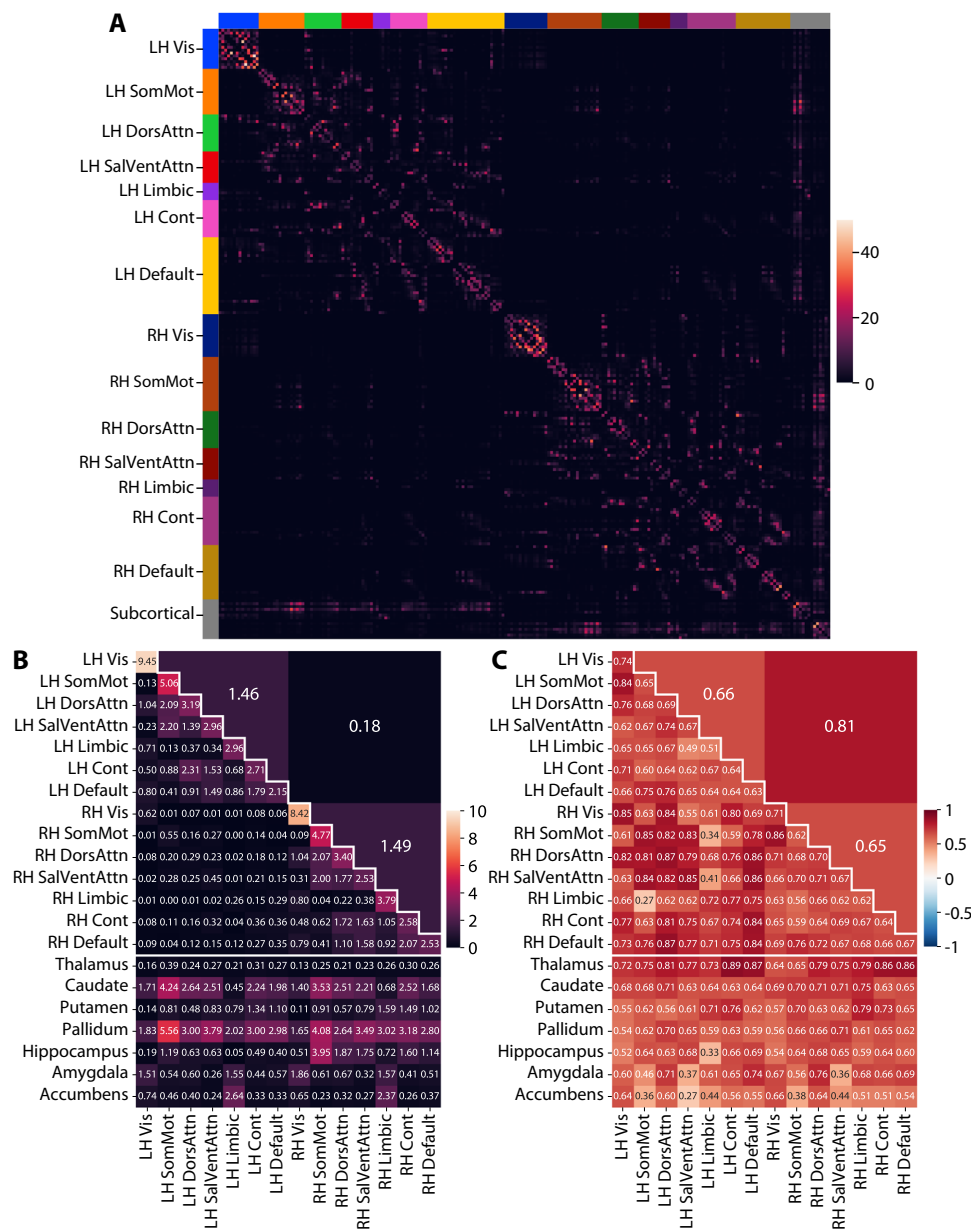


Fig. 2 | Tractography across 26,333 UK Biobank participants. A Participant-averaged connectivity between each pair of parcels in the 200-parcel Schaefer cortical atlas + 14 subcortical parcels from the Harvard-Oxford atlas (in order: left thalamus, left caudate, left putamen, left pallidum, left hippocampus, left amygdala, left accumbens, right thalamus, right caudate, right putamen, right pallidum, right hippocampus, right amygdala, right accumbens). Heatmap entries indicate the weighted number of streamlines connecting each pair of parcels, averaged across participants (see “Tractography pipeline”, “Methods”). Color bands along

the top and left indicate Yeo 7 networks. LH left-hemisphere, RH right-hemisphere, Vis visual, SomMot somatomotor, DorsAttn dorsal attention, SalVentAttn salience/ventral attention, Limbic limbic, Cont control, Default default mode. **B** Participant averages of the 206 connectivity measures. The three large tiles above the diagonal represent hemisphere-level measures. **C** Inter-replicate type 3 intraclass correlation coefficients (ICCs) across 2400 participants with replicate MRI scans, calculated using the *intraclass_corr* function from the *pingouin* Python package²⁰⁰.

relevance of *WNT16* to neurodevelopment has not been conclusively shown.)

The second-most significant association was with an intergenic variant, rs905124 (chr3:190,657,360), with *GMNC* as the nearest protein-coding gene. *GMNC*, also known as *GEMC1*, is a master regulator of multiciliated cell differentiation⁸², and in particular of the multiciliated ependymal cells that line the brain’s ventricles. *GMNC* is necessary for radial glial cells in the sub-ventricular zone (one of the two main brain regions where adult neurogenesis occurs) to differentiate into multiciliated ependymal cells; when *GMNC* is downregulated, these radial glial cells tend to instead differentiate into neural stem cells, promoting

neurogenesis^{83,84}. *GMNC* is associated with congenital hydrocephalus (buildup of cerebrospinal fluid in the ventricles) in both humans and mice⁸⁴. rs905124 has been previously associated with brain volume and cortical surface area according to the GWAS Catalog⁸⁵.

The third-most significant variant, rs13107325 (chr4:103,188,709), is a missense variant in *SLC39A8* (A391T). *SLC39A8* encodes a transmembrane transporter of zinc, cadmium, iron and manganese ions sometimes referred to as ZIP8⁸⁶. *SLC38A8*/*ZIP8* deficiency impairs manganese uptake and compromises the function of manganese-dependent enzymes, including those required for glycosylation, leading to severe neurodevelopmental defects and other

Table 1 | The 30 genome-wide significant loci after Bonferroni correction for the number of measures studied

Variant	Location (hg19)	A1	A2	A1 freq.	# measures	Most significantly associated measure	Effect size	p value	Nearest gene(s)	Candidate causal gene(s)
rs941760	chr14:91,881,751	T	C	44%	8	RH Cont to thalamus	-0.10	5.9×10^{-42}	CCDC88C	CCDC88C
rs905124	chr3:190,657,360	A	T	38%	21	RH Cont to thalamus	0.10	3.5×10^{-39}	GMNC	GMNC
rs13107325	chr4:103,188,709	T	C	7%	7	RH Limbic to hippocampus	-0.20	2.5×10^{-35}	SLC39A8	SLC39A8
rs11245366	chr10:126,482,849	T	C	56%	2	LH Vis to RH Vis	-0.08	1.3×10^{-23}	EEF1AKMT2	--
rs12146713	chr12:106,476,805	C	T	10%	12	LH Cont to thalamus	-0.11	1.2×10^{-19}	NUAK1	NUAK1
rs35050623	chr2:37,063,240	CT	C	54%	5	LH Vis to RH Vis	-0.06	5.7×10^{-17}	STRN	STRN
rs4843550	chr16:87,236,383	C	G	57%	4	RH Limbic to putamen	-0.06	2.0×10^{-16}	C16orf95	--
rs4799450	chr18:35,028,901	C	T	64%	2	RH Vis to pallidum	0.07	5.7×10^{-16}	CELF4	CELF4
rs199790004	chr8:26,427,757	A	AG	62%	1	LH Default to putamen	0.06	3.1×10^{-14}	DPYSL2	DPYSL2
rs35124509	chr3:89,521,693	C	T	39%	3	LH Vis to LH Cont	-0.06	2.0×10^{-13}	EPHA3	EPHA3
rs150346963	chr7:117,625,599	T	C	41%	1	LH Default to RH Limbic	0.06	2.5×10^{-13}	CTTNBP2	CTTNBP2
rs118087478	chr17:44,051,589	G	T	22%	1	LH Limbic to caudate	-0.07	3.9×10^{-13}	MAPT	MAPT
rs142005327	chr7:120,969,969	GCT	G	26%	2	LH SalVentAttn to LH Cont	0.06	6.9×10^{-13}	WNT16	--
rs79814107	chr14:59,637,503	G	A	12%	1	LH Cont to RH Vis	0.08	1.6×10^{-12}	DAAM1	DAAM1
rs34844662	chr13:109,687,510	AT	A	34%	1	LH Vis to RH Vis	-0.06	2.1×10^{-12}	MYO16	MYO16
rs6558407	chr8:144,995,494	T	C	43%	5	LH SalVentAttn to RH SomMot	0.05	4.9×10^{-12}	PLEC	PLEC
rs59154421	chr2:162,813,034	C	CTG	29%	1	LH Limbic to RH Default	-0.06	6.0×10^{-12}	SLC4A10	SLC4A10
rs75650221	chr1:174,421,994	T	C	4%	1	LH Default to caudate	0.13	6.3×10^{-12}	RABGAP1L	--
rs13228652	chr7:83,761,158	C	G	18%	1	LH Vis to RH Vis	0.07	2.8×10^{-11}	SEMA3A	SEMA3A
rs141628902	chr7:43,818,855	T	C	3%	1	LH SalVentAttn to amygdala	0.16	2.9×10^{-11}	BLVRA	--
rs1158069	chr7:19,643,101	A	G	37%	1	RH Vis to caudate	-0.05	4.0×10^{-11}	POLR1F	--
rs975360	chr7:81,411,028	C	T	40%	1	RH Cont to amygdala	-0.05	7.2×10^{-11}	HGF	HGF
rs2796245	chr1:208,030,852	A	G	86%	1	Cross-hemisphere	0.06	7.4×10^{-11}	CD34	--
rs1865249	chr8:142,215,982	C	A	61%	1	LH Cont to RH Cont	0.05	1.1×10^{-10}	SLC45A4	--
rs415978	chr9:118,985,910	A	G	40%	1	RH SomMot to caudate	0.05	1.4×10^{-10}	PAPPA	--
rs1391762	chr4:54,755,157	G	A	82%	1	LH SomMot to RH SomMot	-0.06	1.6×10^{-10}	ENSG00000282278	--
rs6062264	chr20:61,154,871	T	C	28%	1	RH Vis to RH Vis	-0.05	1.8×10^{-10}	GATA5	--
rs5788199	chr10:118,649,704	TA	T	75%	1	LH DorsAttn to amygdala	0.06	1.8×10^{-10}	ENO4, SHTN1	SHTN1
rs551746431	chr2:234,112,457	C	*	2%	1	LH Cont to thalamus	-0.20	1.8×10^{-10}	INPP5D	--
rs115136616	chr5:93,021,116	T	C	5%	1	RH Vis to RH SomMot	0.12	2.0×10^{-10}	FAM172A	--

Freq. = frequency; # measures refers to the number of measures (out of 206) the variant is genome-wide significant for. Effect sizes are oriented with respect to the minor allele (A1). Candidate causal genes were chosen based on a literature review of the nearest genes to each lead variant. rs59154421 (SLC4A10) is referred to by its UK Biobank variant identifier 2:162813034_CTG_C in the Supplementary Data and summary statistics, while rs551746431 (INPP5D) is referred to by its UK Biobank ID 2:234112457_CCAGCTACTCCCAAGTAGTCT_C (the reference allele, CCAGCTACTCCCAAGTAGTCT, is indicated with an asterisk in Table 1 for brevity).

phenotypes⁸⁷. This variant was the eighth-most significant lead variant in the largest schizophrenia GWAS to date⁸⁸, and this variant has also been shown to impair glycosylation in the mouse brain⁸⁹. rs13107325's uniquely strong association with cortico-hippocampal connectivity (see previous section) is consistent with the long literature relating hippocampal size, morphology, function and connectivity to schizophrenia⁹⁰. rs13135092 has 477 reported associations in the GWAS Catalog, including with a wide variety of brain imaging measures.

The fifth-most significant variant, rs12146713 (chr12:106,476,805), is an intronic variant in *NUAK1*. *NUAK1* dose-dependently regulates axon elongation and branching via effects on mitochondrial metabolism and trafficking⁹¹⁻⁹³. *NUAK1* is a candidate gene for autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD) and intellectual disability. In mice, *NUAK1* haploinsufficiency impairs cortical development and has broad-spectrum effects on cognition⁹². The GWAS Catalog reports 73 associations for rs12146713, all brain structure-related.

The sixth-most significant variant, rs35050623 (chr2:37,063,240), is an intergenic variant with *STRN* as the nearest protein-coding gene. *STRN* (striatin) is a calmodulin-binding protein enriched in dendritic spines^{94,95}. In cultured striatal neurons, *STRN* knockdown led to increased dendritic arborization and increased density of stubby spines (a type of dendritic spine with no neck), indicating a role in regulating striatal neurodevelopment⁹⁶. rs35050623 has one GWAS Catalog association, with brain shape.

The eighth-most significant variant, rs4799450 (chr18:35,028,901), is an intronic variant in *CELF4*. *CELF4* is an RNA binding protein and translational regulator of prenatal neocortical subplate layer-based synaptic development⁹⁷. *CELF4* is involved in differentiation and excitation of neurons, corticothalamic development, synaptic transmission, and neuroplasticity and is an important regulator of mRNA stability and translational availability⁹⁸. Deficiency can lead to dysfunctional neuronal excitation and impaired synaptic transmission⁹⁹. *CELF4* dysfunction has been linked to a variety of neuropsychiatric disorders including autism, bipolar disorder,

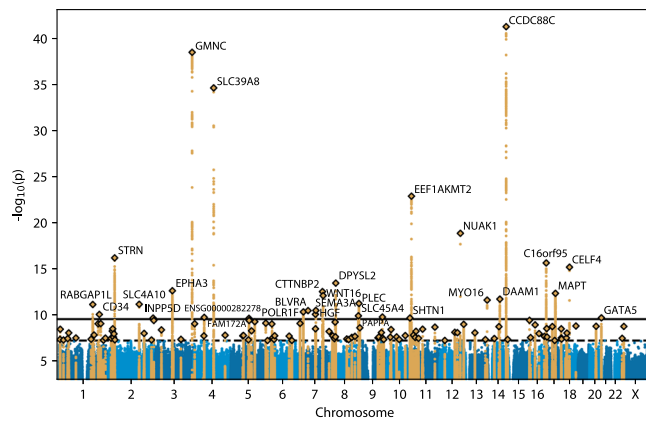


Fig. 3 | Manhattan plot of each variant's minimum p value across the 206 structural connectivity measures. Gold variants indicate linkage disequilibrium (LD) clumps passing nominal genome-wide significance (dashed black line); black diamonds highlight the lead variant in each clump. Nearest genes are labeled for each clump passing Bonferroni-corrected genome-wide significance (solid black line).

schizophrenia, and epilepsy⁹⁸. rs4799450 has one GWAS Catalog association, with white matter microstructure.

The ninth-most significant variant, rs199790004 (chr8:26,427,757), is an intronic variant in *DPYSL2*. *DPYSL2* is a microtubule-stabilizing protein that plays a role in the regulation of axonal outgrowth, dendritic development, synapse elongation and vesicle trafficking during neurodevelopment¹⁰⁰ and is a key regulator of neural stem cell differentiation¹⁰¹. *DPYSL2* is a known schizophrenia (SCZ) risk gene and its expression is altered in schizophrenia patient brains^{102–104}, possibly as a downstream consequence of disrupted mTOR signaling^{103,104}.

The tenth-most significant variant, rs35124509 (chr3:89,521,693), is a missense variant in *EPHA3*, the sole protein-coding gene for over 1 megabase in either direction. Knockdown of *EPHA3* disrupts pathfinding of axons through the corpus callosum¹⁰⁵. *EPHA3* activation reduces neurite outgrowth and, more specifically, induces growth cone collapse¹⁰⁶. *EPHA3* also affects higher-level structural connectivity patterns, regulating excitation-inhibition balance and GABAergic interneuron synaptic density^{107,108}. rs35124509 has 16 GWAS Catalog associations, all with brain imaging measurements.

The eleventh-most significant variant, rs150346963 (chr7:117,625,599), is an intergenic variant located -100 kilobases upstream of *CTTNBP2*. *CTTNBP2* (cortactin binding protein 2) is a cytoskeletal protein predominantly expressed in neurons and controls synapse and dendritic spine formation and maintenance by interacting with cortactin, which binds to and stabilizes actin filaments^{109–111}. *CTTNBP2* is a candidate gene for autism spectrum disorder^{109–113} and mice with autism spectrum disorder-linked mutations in *CTTNBP2* knocked in have impaired synaptic function¹¹⁰. Meanwhile, *CTTNBP2* knockout mice have reduced brain zinc levels – a risk factor for autism – and altered synaptic protein targeting, while zinc supplementation rescues synaptic retention of *CTTNBP2* and improves social behaviors in these mice^{112,113}. rs150346963 has one GWAS Catalog association, with major depressive disorder.

The twelfth-most significant variant, rs118087478 (chr17:44,051,589), is an intronic and 5' UTR variant in *MAPT*. *MAPT* (tau) is best known for its central role in neurodegenerative disorders, but it is also involved in neurodevelopment, where it is highly expressed¹¹⁴ and regulates the assembly and stability of microtubules and their linkage to the plasma membrane¹¹⁵. Microdeletions in the *MAPT* genetic region (at chromosome 17q21.3) have been linked to intellectual disabilities¹¹⁶, which is proposed to be the result of altered tau dosage¹¹⁴, and *Mapt* knockout leads to a reduction in autistic

behaviors in mouse models of autism¹¹⁷. rs118087478 has one GWAS Catalog association, with brain volume.

The 14th-most significant variant, rs79814107 (chr14:59,637,503), is an intronic variant in *DAAMI*. *DAAMI* (disheveled associated activator of morphogenesis 1) is a scaffolding protein that acts downstream of Wnt signaling and controls cell polarity and movement by regulating actin cytoskeleton organization^{118,119}. In mice, deletion of a neural-specific microexon in *Daam1* led to memory defects, reduced long-term potentiation and a decrease in the number of dendritic spines¹²⁰. rs79814107 has been previously associated with brain volume and white matter integrity according to the GWAS Catalog.

The 15th-most significant variant, rs34844662 (chr13:109,687,510) is an intronic variant in *MYO16*. *MYO16* is an unconventional myosin that is involved with cytoskeleton remodeling and is predominantly expressed in the central nervous system¹²¹. Within neurons, *MYO16* interacts with the WAVE regulatory complex which modulates dendritic spine morphology by regulating actin polymerization, particularly in Purkinje cells, and synaptic organization^{121,122}. *MYO16* is a candidate gene for bipolar II disorder¹²³, schizophrenia¹²⁴, and autism¹²⁵.

The 16th-most significant variant, rs6558407 (chr8:144,995,494), is a missense variant in *PLEC*. *PLEC* (plectin) is a 'cytolinker' protein that cross-links cytoskeletal elements with each other, including intermediate filaments, actin filaments, microtubules, and components of the cell and nuclear membranes¹²⁶. Plectin is highly expressed in the central nervous system, especially at the interfaces between glia and pial cells and between glia and endothelial cells, and is thought to be important to blood-brain barrier and pial surface integrity¹²⁷. Plectin has been linked to epilepsy and Alexander disease, a rare demyelinating disease¹²⁸. Mice genetically deficient in the plectin isoform Plc had poorer learning capabilities and long-term memory than wild-type littermates¹²⁹.

The 17th-most significant variant, rs59154421 (chr2:162,813,034), is an intronic variant in *SLC4A10*. *SLC4A10* is a transporter in the plasma membrane that is predominantly expressed in neurons. It helps regulate neuronal pH by taking in bicarbonate, driven by a sodium gradient across the membrane, which helps remove hydrogen ions^{130,131}. Case studies have shown that autosomal recessive loss of function of *SLC4A10* leads to intellectual disability, microcephaly and lateral ventricle abnormalities¹³⁰. *Slc4a10* knockout mice have smaller brain ventricles and behavioral abnormalities, and *SLC4A10* modulates GABA (but not glutamate) release in mouse brain¹³⁰. *SLC4A10* is a candidate gene for autism spectrum disorder¹³² and epilepsy¹³³.

The 19th-most significant variant, rs13228652 (chr7:83,761,158), is an intronic variant in *SEMA3A*. Semaphorins are a family of membrane proteins that orchestrate axon guidance and induce growth cone collapse (leading to their alternate name, collapsins)^{134,135}. *SEMA3A* (semaphorin 3A, also known as collapsin-1 or collapsin) was the first-discovered semaphorin¹³⁴ and acts as an axonal chemorepellent (i.e., it repels growth cones) as well as inducing growth cone collapse. In addition to this primary role, *SEMA3A* also has an important secondary role in inhibiting myelin regeneration in multiple sclerosis, at least in part by inhibiting oligodendrocyte precursor cell differentiation and recruitment to demyelinated lesions^{136–139}.

The 22nd-most significant variant, rs975360 (chr7:81,411,028) is -10 kilobases upstream of the transcription start site of *HGF*. *HGF* (hepatocyte growth factor) is a pleiotropic cytokine that provides trophic (growth, differentiation and survival) support to neurons and is involved in promoting cell movement, axon growth and dendritic morphology^{140,141}. *HGF* is expressed in the cortex and hippocampus during neurodevelopment where it regulates thalamocortical axon growth¹⁴². *HGF*'s cognate receptor, the tyrosine kinase *MET*, is a high-confidence autism risk gene, and *MET* protein levels are lower in the brains of people with autism compared to healthy controls¹⁴³. A mouse

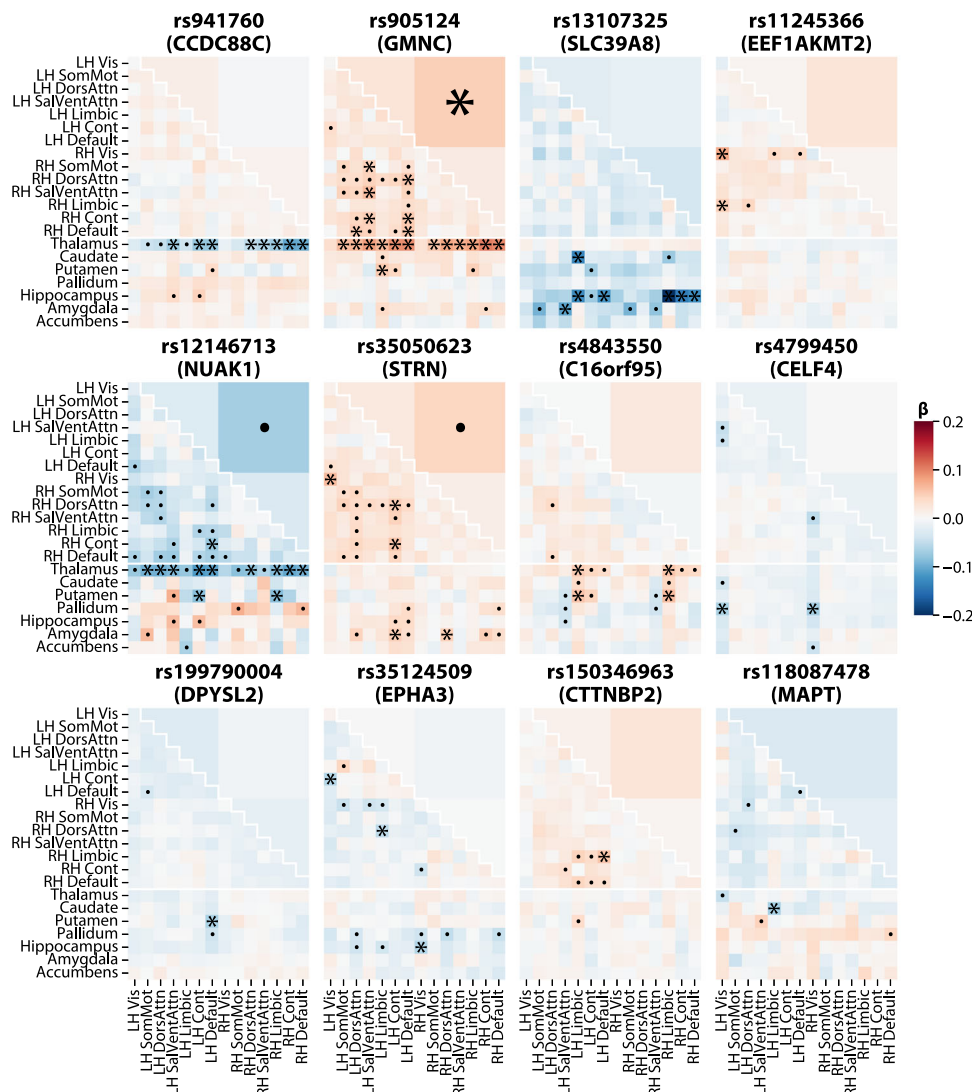


Fig. 4 | Spatial patterns of association with structural connectivity. For brevity, only the 12 most significant lead variants are shown; the remaining 18 are shown in Supplementary Fig. 1. Asterisks (*) indicate associations passing Bonferroni-corrected genome-wide significance ($p < 5 \times 10^{-8} / 206 \approx 2.4 \times 10^{-10}$), while dots (·) indicate associations significant after Bonferroni correction for just the 30 variants times the number of measures ($p < 0.05 / 30 / 206 \approx 8.1 \times 10^{-6}$). Effect sizes (β) are

oriented with respect to the minor allele, so positive effect sizes (red) indicate that the minor allele is associated with increased structural connectivity, while negative effect sizes (blue) indicate that the minor allele is associated with decreased connectivity. LH left-hemisphere, RH right-hemisphere, Vis visual, SomMot somatomotor, DorsAttn dorsal attention, SalVentAttn salience/ventral attention, Limbic limbic, Cont control, Default default mode.

model of decreased HGF and MET expression showed deficiencies in interneuronal migration during neurodevelopment¹⁴⁴.

The 28th-most significant variant, rs5788199 (chr10:118,649,704), is an intronic variant in two overlapping genes, *SHTN1* and *ENO4*. *SHTN1* is necessary for neuronal polarization, the process by which one of the neurites from a developing neuron ultimately becomes the axon and the rest become dendrites. *SHTN1* builds up asymmetrically in the neurite that ultimately becomes the axon, and disrupting this asymmetrical accumulation either by overexpression or knockdown disrupts neuronal polarization¹⁴⁵. *Shtn1* knockout mice exhibited thinning of multiple white matter tracts—the corpus callosum, anterior commissure and hippocampal commissure—as well as absence of the septum¹⁴⁶.

Structural connectivity is highly polygenic and exhibits signatures of negative selection

We next inferred global properties of the genetic architecture of each structural connectivity measure by applying the recently developed SBayesS method¹⁴⁷ (Fig. 1, Fig. 5).

The proportion of heritability (h^2) collectively explained by the 9,423,516 common variants included in the GWAS was relatively modest, ranging from 3.3 to 27.7% (median 13.9%) across the 206 GWAS. As expected, the most heritable measures were also the most replicable between the initial and the follow-up scans (Pearson correlation between heritability and replicability across measures = 0.77), indicating greater stability of these measures over time. Estimates of h^2 from LD score regression¹⁴⁸ were lower, ranging from 1.0 to 18.0% (median 9.1%), but displayed a similar pattern across the 206 measures (Supplementary Fig. 2). The discrepancy in h^2 between the two methods may be attributable to the different assumptions about genetic architecture: in particular, SBayesS assumes a sparse genetic architecture, while LD score regression does not make such an assumption. Heritability was higher for inter-hemisphere connectivity ($h^2 = 27.7\%$) than either left intra-hemisphere ($h^2 = 18.0\%$) or right intra-hemisphere ($h^2 = 16.3\%$) connectivity.

Polygenicity (π) was quite high, with between 0.8% and 9.6% (median 7.5%) of common variants surveyed estimated to have non-zero effects on each of the 206 measures. Strikingly, corticothalamic

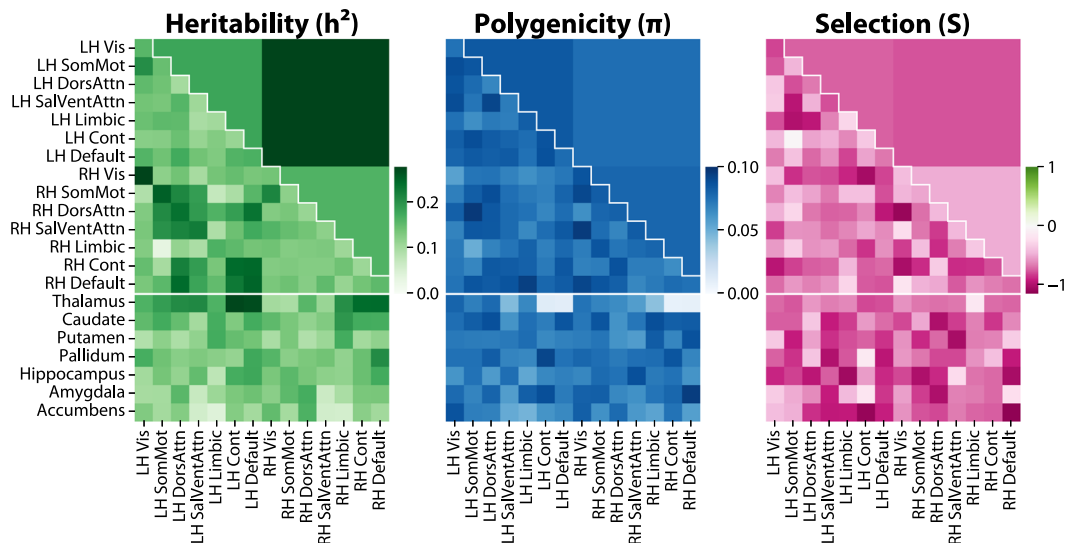


Fig. 5 | Global properties of the genetic architecture of structural connectivity. LH left-hemisphere, RH right-hemisphere, Vis visual, SomMot somatomotor, DorsAttn dorsal attention, SalVentAttn salience/ventral attention, Limbic limbic, Cont control, Default default mode.

connectivity measures involving the default-mode and control networks were among the least polygenic ($\pi = 0.8\text{--}1.2\%$, median 1.0%) despite being among the most heritable ($h^2 = 24.2\text{--}27.5\%$, median 25.7%). The combination of low polygenicity and high heritability suggests that these corticothalamic connections are controlled by more specific gene sets than other types of connections. This observation is consistent with our identification of variants with disproportionate associations with corticothalamic connectivity overall, near *CCDC88C*, *GMNC*, *NUAK1*, and *INPP5D* (Fig. 4).

SBayes also estimates a selection parameter (S), where -1 indicates strong negative selection, 0 indicates no selection, and $+1$ indicates strong positive selection. The degree to which a trait is under selection can be inferred from the relationship between minor allele frequency and GWAS effect size. For instance, if a trait is under strong negative selection, variants with large effects on the trait will tend to be rare, since individuals who carry these variants will have lower evolutionary fitness and be less likely to pass on their genes. 194 of the 206 measures were estimated to be under negative selection (95% confidence interval of S entirely below 0), indicating significantly larger effect sizes among lower-frequency variants. All 206 measures had point estimates of S below 0 , ranging from -1.13 to -0.03 (median -0.70).

Overall, the high polygenicity of these measures indicates that a substantial fraction of the genome is involved in genetically determining structural connectivity, and remains to be discovered by future GWAS. Meanwhile, the fact that these measures appear to be under negative selection suggests that variants affecting structural connectivity tend to have negative consequences on evolutionary fitness.

Structural connectivity is genetically correlated with a range of psychiatric and cognitive traits

We computed genetic correlations (r_g) between our 206 GWAS and 15 well-powered GWAS for neurological, psychiatric, neurodevelopmental, and cognitive traits, using the genetic covariance analyzer (GNOVA) method¹⁴⁹. We chose GNOVA due to its improved correction for sample overlap relative to cross-trait LD score regression¹⁵⁰, another commonly used genetic correlation method. After Bonferroni correction across the 206×15 genetic correlations tested, we observed 18 significant genetic correlations (Supplementary Fig. 3). 6 traits had at least one significant genetic correlation (Fig. 1): bipolar disorder, ADHD, insomnia, Alzheimer's disease, educational attainment, and reaction time. These significant genetic correlations were positive for ADHD ($r_g =$

$0.19\text{--}0.24$), insomnia ($r_g = 0.16$), and Alzheimer's disease ($r_g = 0.23$), indicating that the genetic variants associated with increased connectivity tended to be the same genetic variants associated with increased risk for these conditions, when looking across the whole genome rather than only at genome-wide significant variants. Conversely, the significant genetic correlations were negative for bipolar disorder ($r_g = -0.17$), educational attainment ($r_g = -0.22$ to -0.11), and reaction time ($r_g = -0.18$ to -0.13), indicating that the genetic variants associated with increased connectivity tended to be the same genetic variants associated with reduced risk for bipolar disorder, lower educational attainment, and – somewhat paradoxically given the educational attainment result – better reaction time. On the whole, these genetic correlations support the relevance of genetic factors influencing structural connectivity to a range of neuropsychiatric and cognitive outcomes.

We also computed phenotypic and genetic correlations between the three hemisphere-level measures and 432 diffusion MRI measures based on TBSS and neurite orientation dispersion and density imaging (NODDI) in the UK Biobank (Supplementary Data 3), leveraging prior GWAS of these measures¹⁸. Cross-hemisphere connectivity had Bonferroni-significant phenotypic and genetic correlations with 401 and 197 of the 432 measures, respectively; left-hemisphere connectivity with 381 and 155; and right-hemisphere connectivity with 375 and 100. Phenotypic correlations ranged from -0.60 (between cross-hemisphere connectivity and “Mean OD [orientation dispersion] in cerebral peduncle on FA [fractional anisotropy] skeleton (left)”) to 0.53 (between cross-hemisphere connectivity and “Mean FA in cerebral peduncle on FA skeleton (left)”). Genetic correlations ranged from -0.59 (between cross-hemisphere connectivity and “Mean MO [mean orientation] in body of corpus callosum on FA skeleton”) to 0.53 (between cross-hemisphere connectivity and “Mean ISOVF [isotropic volume fraction] in body of corpus callosum on FA skeleton”). On the whole, this analysis indicates that structural connectivity is phenotypically and genetically related, but nonetheless substantially distinct, from more widely used diffusion MRI measures.

Structural connectivity heritability is enriched in regions with increased chromatin accessibility in oligodendrocytes

We applied partitioned linkage disequilibrium score regression¹⁵¹ to investigate whether heritability for each structural connectivity

measure was enriched in regions with increased chromatin accessibility in each of 6 major brain cell types—astrocytes, excitatory neurons, inhibitory neurons, microglia, oligodendrocytes, and oligodendrocyte precursor cells (OPCs)—measured by single-nucleus chromatin accessibility and messenger RNA expression sequencing (SNARE-seq2) of postmortem adult human motor cortex¹⁵². After Bonferroni correction across the 206×6 cell-type enrichments tested, we found significant enrichments for 7 structural connectivity measures across four cell types (Fig. 1, Supplementary Fig. 4): oligodendrocytes with connectivity within the right-hemisphere default-mode network (67-fold enrichment, 95% confidence interval 32–101, $p = 3 \times 10^{-5}$), microglia with 3 measures, most significantly left-hemisphere dorsal attention to salience/ventral attention connectivity (52-fold [29–76], $p = 4 \times 10^{-7}$), inhibitory neurons with right-hemisphere dorsal attention to right-hemisphere limbic connectivity (49-fold [28–69], $p = 5 \times 10^{-6}$), and astrocytes with 2 measures, most significantly left- to right-hemisphere visual network connectivity (48-fold [27–69], $p = 5 \times 10^{-6}$). There were many sub-significant enrichments as well, particularly for astrocytes and oligodendrocytes. Given that oligodendrocytes are the cell type responsible for myelination and one of the primary functions of oligodendrocytes is to myelinate white matter fibers¹⁵³, the observed oligodendrocyte heritability enrichment suggests that genetic control of structural connectivity may be partially mediated by effects on myelination.

We performed the same cell-type enrichment analysis across 12 cell types from the developing human cortex, using published single-cell assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) data¹⁵⁴. After Bonferroni correction across the 206×12 cell-type enrichments tested, we found significantly enriched heritability in four cell types, all non-neuronal (Supplementary Fig. 5). These cell types were: astrocytes and oligodendrocyte progenitor cells (for right-hemisphere visual network to pallidum connectivity: 70-fold enrichment, 95% confidence interval 34–106, $p = 2 \times 10^{-5}$), endothelial/mural cells (for 6 structural connectivity measures, with enrichments between 98- and 125-fold), microglia (19 measures, 65- to 126-fold), and radial glial cells (18 measures, 35- to 82-fold), stem cells that are progenitors of both neurons and specific glia, namely astrocytes and oligodendrocytes. (We also found significantly disenriched heritability in insular neurons.) These heritability enrichments for multiple fetal cell types suggest that structural connectivity is partially mediated by effects on early brain development, particularly in non-neurons.

Discussion

In this study, we describe the genetic architecture of white-matter structural connectivity via genome-wide association studies of 206 tractography-derived measures across 26,333 UK Biobank participants. We discover 30 loci associated with structural connectivity after rigorous correction for multiple statistical tests performed (with 126 loci at nominal genome-wide significance). These loci pinpoint neurodevelopmental genes with influences on myelination (*SEMA3A*), neurite elongation and guidance (*NUAK1*, *STRN*, *DPYSL2*, *EPHA3*, *SEMA3A*, *HGF*, *SHTN1*), neural cell proliferation and differentiation (*GMNC*, *CELF4*, *HGF*), neuronal migration (*CCDC88C*), cytoskeletal organization (*CTTNBP2*, *MAPT*, *DAAMI*, *MYO16*, *PLEC*), and brain metal transport (*SLC39A8*). In addition to these biological processes, we observe enriched heritability for structural connectivity in regions with increased chromatin accessibility in oligodendrocytes, microglia, inhibitory neurons, astrocytes, and multiple fetal cell types, supporting a role for myelination and early brain development in the genetics of structural connectivity. Our results extend the rich literature of MRI GWAS measuring other aspects of brain structure and function.

We show, for the first time, that structural connectivity is highly polygenic: we estimate that for the participants in our cohort, the

average connectivity measure is influenced by 9.1% of all common (>1% frequency) variants in the genome. This suggests that our 30 loci represent only a small fraction of genetic influences on the structural connectome; much larger sample sizes, possibly in the tens of millions, may be needed to discover the rest¹⁵⁵. This high polygenicity is accompanied by evidence of negative selection: higher-frequency variants tend to have smaller effects on structural connectivity. Together, these results imply a genetic architecture whereby large numbers of disproportionately common variants each exert small effects on structural connectivity, while larger-effect variants tend to be flushed out of the population due to their deleteriousness. Both high polygenicity^{22,155} and strong negative selection²² appear to be characteristic of brain MRI measures in general, rather than being unique properties of structural connectivity in particular. On the other hand, the 18 UK Biobank traits analyzed in the SBayesS paper¹⁴⁷ tend to be less polygenic than our structural connectivity measures, despite having similar evidence of negative selection.

Far from providing a one-to-one map between specific genetic variants and specific regional connectivities, our results indicate widespread pleiotropy of structural connectivity-associated variants. Our 30 lead variants are each associated with a median of 7.5 of the 206 measures, and many are also associated with other aspects of brain structure. The spatial associations of these variants fall into four broad patterns: three variants are disproportionately associated with interhemispheric and corticothalamic connectivity, one with corticothalamic connectivity alone, and eight with inter-hemispheric alone, while the remaining 18 have spatially diffuse associations. Corticothalamic connectivity measures involving the default-mode and control networks are among the least polygenic despite being among the most heritable, providing further evidence in support of corticothalamic connections being controlled by more specific gene sets than other types of connections. Corticothalamic structural connectivity may have outsized importance relative to other types of structural connections, as it underpins cortico-thalamo-cortical loops, modular units of brain organization involved in consciousness and perception¹⁵⁶.

Genetic correlation analyses indicate shared genetic influences between structural connectivity and a range of neuropsychiatric and cognitive traits. The directionality of these associations is somewhat counterintuitive: variants associated with increased structural connectivity tended to be associated with increased risk of ADHD, insomnia, and Alzheimer's disease, decreased risk of bipolar disorder, lower educational attainment, and better reaction time. While cognitive decline has been associated with reduced structural connectivity^{56,157}, and cognitive resilience with increased connectivity¹⁵⁸, in mild cognitive impairment, Alzheimer's disease and/or aging, the relationship between structural connectivity and cognition in healthy individuals after accounting for age has been less well-studied. Note that the Alzheimer's disease GWAS we used included proxy cases for Alzheimer's disease and related dementias, which increases power but can also lead to bias in genetic epidemiological analysis¹⁵⁹.

In parallel to ours, another study also examined the genetic architecture of white-matter structural connectivity in the UK Biobank¹⁶⁰. Unlike our study, the authors used a multivariate GWAS technique, MOSTest¹⁶¹, to boost power by leveraging the fact that GWAS variants associated with structural connectivity tend to be associated with connectivity changes in multiple parts of the brain. While focused less on the mechanisms underlying individual causal gene candidates, the authors nonetheless show a global enrichment of their GWAS associations for many of the same neurodevelopmental processes as we report for our individual causal gene candidates, including neurogenesis, neural differentiation, neural migration, neural projection guidance, and axon development. As in our study, the authors report pervasive shared genetics between structural connectivity and neuropsychiatric disease.

This study has several important limitations. First, the core limitation of our study is the use of tractography to infer structural connectivity. The term “structural connectivity” and the assumption that we can infer connectivity at all from diffusion MRI-based methods is inherently limited; this is true of all cases of using MRI derived metrics to infer neurobiological details. Unlike *ex vivo* methods like histological tract-tracing, tractography is not a direct measure of structural connectivity. The measurements it is based on relate to the diffusion of water molecules in tissue, with tissue structure or organization influencing the diffusivity. This is far removed from true structural connectivity (at the level of axons), and thus we are building at best an estimated model of large-scale trends of connectivity. Tractography has trouble resolving small or intermingled axon bundles, and it is vulnerable to false positive connections^{162,163}. Moreover, tractography is ill-suited to measuring the short-range and unmyelinated connections typical of gray matter. Different tractography algorithms can also have differences based on tractography models and parameters¹⁶⁴, which can vary by tract¹⁶⁵. Our approach attempted to use current best practices to minimize the generation of spurious tracts and assess reliability. As diffusion algorithms and imaging techniques improve, more reliable and interpretable metrics will likely be derived. This could lead to the identification of metrics that relate more specifically to different types of biological variation (e.g. myelination versus packing density) and thereby allow the identification of genetic variation specifically related to different aspects of structural variation. As such, while we were able to identify relationships between these large scale models of structural connectivity and genetics, these results should be interpreted within the limitations of current diffusion imaging approaches. These limitations should be weighed against the unique opportunity and scalability afforded by diffusion MRI-based tractography to uncover genetic influences on the physical wiring of the human brain.

Second, our results were derived from MRI scans of participants aged 40–70 (median 55), and may not generalize to participants of other age groups. In particular, age-related structural connectivity changes are associated with cognitive decline¹⁶⁶, which may be related to our unexpected observation of negative genetic correlations between structural connectivity and cognitive measures.

Third, the 206 structural connectivity measures we defined are inherently somewhat arbitrary, and other measures might uncover complementary biological insights. For instance, we chose to define cortical measures at the level of Yeo 7 networks, which are defined based on functional, rather than structural, connectivity. While we feel this choice has substantial advantages in terms of interpretability, defining cortical networks in a data-driven way, based on the tractography data itself, is an alternative possibility. Other possibilities include considering subcortical structures in a hemisphere-specific manner (e.g. considering left default mode to left amygdala connections separately from left default mode to right amygdala connections), including subcortical-to-subcortical connections, and incorporating measures of overall network topology like network efficiency^{56,59} and rich-club organization^{167,168}. Exemplifying the kinds of insights that can be gained from genetic analyses of these more complex measures, a recent twin study found that heritability of structural connectivity was concentrated in rich-club connections, i.e. those interlinking network hubs¹⁶⁹.

In sum, our results indicate pervasive genetic influences on structural connectivity within the human brain and support their relevance to brain function in health and disease. Future studies using more sophisticated imaging techniques will be necessary to extend these results to gray-matter structural connectivity and develop a truly comprehensive map of genetic influences on the human structural connectome.

Methods

Cohort selection and MRI quality control

In total, 41,489 participants from the UK Biobank were listed as having both T1 (Data-Field 20263, “T1 surface model files and additional structural segmentations”) and diffusion-weighted (Data-Field 20250, “Multiband diffusion brain images - NIFTI”) MRI scans available at baseline. However, of these baseline scans, 1456 lacked diffusion-weighted MRI scans corrected for eddy currents¹⁷⁰, head motion, outlier slices, and gradient distortion (data_ud.nii.gz from Data-Field 20250); a further 8725 lacked eddy-corrected bvec files (data.eddy_r-otated_bvecs from Data-Field 20250); and a further 3 lacked FreeSurfer parcellation images (aseg.mgz from Data-Field 20263). This left 31,309 participants.

We next considered four quality control metrics provided by the UK Biobank: T1 inverse signal-to-noise ratio (“Inverted signal-to-noise ratio in T1”, Data-Field 25734), T1 inverse contrast-to-noise-ratio (“Inverted contrast-to-noise ratio in T1”, Data-Field 25735), number of diffusion MRI outlier slices (“Number of dMRI outlier slices detected and corrected”, Data-Field 25746), and left-to-right head motion as measured by eddy in the diffusion MRI (“Standard deviation of apparent translation in the Y axis as measured by eddy”, Data-Field 25922). All four quality control metrics were defined so that higher values are worse. Noting that the distributions of all four metrics were right-skewed to varying degrees, with a tail of poor-quality scans, we excluded scans where any of the four metrics were more than 3 standard deviations above the mean (Supplementary Fig. 6). 1113 participants had baseline scans excluded, leaving 30,196 participants remaining.

In total, 26,445 of these participants were of European genetic ancestry (according to Pan-UK Biobank, Return 2442) and also had genotyping data deemed suitable for genetic analyses: genotypes available, no mismatch between genetic sex (“Genetic sex”, Data-Field 22001) and self-reported sex (“Sex”, Data-Field 31), no sex chromosome aneuploidy (“Sex chromosome aneuploidy”, Data-Field 22019), and not flagged as “Outliers for heterozygosity or missing rate” (Data-Field 22027). An additional 668 participants were of non-European genetic ancestry (Central/South Asian, African, East Asian, Middle Eastern or Admixed American) and had genotyping data deemed suitable for genetic analyses.

Thus, we ran our tractography pipeline (see next section) on 27,113 baseline scans, of which 26,998 ran successfully: 26,333 of European genetic ancestry and 665 of non-European genetic ancestry. We used the 26,333 European baseline scans for the main GWAS, and the 665 non-European baseline scans for the replication GWAS. To evaluate replicability of the tractography measures themselves, we used a subset of 2400 participants of European genetic ancestry who also had replicate scans that passed all the above-mentioned filters and ran successfully.

Tractography pipeline

Our tractography pipeline, *tractify* (<https://github.com/TIGRLab/tractify>), performs probabilistic tractography using version 3.0.3 of the MRtrix3 diffusion MRI software package⁶⁹. Certain steps also use version 6.0.5 of the Functional Magnetic Resonance Imaging of the Brain Software Library (FSL)¹⁷¹ and version 7.1.1 of the FreeSurfer software package¹⁷².

The pipeline takes the following subject-specific inputs:

1. A brain-extracted, bias-field-corrected T1 scan (brain.mgz from Data-Field 20263), the processing of which is described in detail in the UK Biobank Brain Imaging Documentation.
2. A diffusion-weighted MRI scan corrected for eddy currents, head motion, outlier slices, and gradient distortion (data_ud.nii.gz from Data-Field 20250), the processing of which is also described in the UK Biobank Brain Imaging Documentation.
3. A bval file (bvals from Data-Field 20250).

4. An eddy-corrected bvec file (data.eddy_rotated_bvecs from [Data-Field 20250](#)).
5. A FreeSurfer parcellation of the T1 image (aseg.mgz from [Data-Field 20263](#)), and the following subject-independent inputs:
6. The 2-mm MNI152 standard-space structural template image, MNI152_T1_2mm_brain.nii.gz.
7. The atlas: we concatenated the 200-parcel Schaefer cortical atlas⁷³ mapped to the “Yeo 7” networks⁶⁷, available from https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal/Parcellations/MNI/Schaefer2018_200Parcels_7Networks_order_FSLMNI152_2mm.nii.gz, with 14 subcortical parcels from the Harvard-Oxford atlas (left thalamus, left caudate, left putamen, left pallidum, left hippocampus, left amygdala, left accumbens, right thalamus, right caudate, right putamen, right pallidum, right hippocampus, right amygdala, right accumbens), for a total of 214 parcels. Voxels present in both a Schaefer parcel and a subcortical Harvard-Oxford parcel were assigned to their Schaefer rather than their Harvard-Oxford parcel. This atlas is available as Supplementary Data 4.
8. Compute the MNI152-to-T1 warping transformation by inverting the T1-to-MNI152 warping transformation, using FSL’s `invwarp` command. Provide the T1-to-MNI152 warping transformation (output at step 7) via the `-w` flag, the T1 scan in NIfTI format (output at step 4) via the `-r` flag, and the location to save the MNI152-to-T1 warping transformation via the `-o` flag.
9. Register the atlas from MNI152 space to diffusion space using FSL’s `applywarp` command, by applying the non-linear MNI152-to-T1 warping transformation from step 8, followed by the linear T1-to-diffusion transformation from step 5. Provide the atlas (input #7) via the `-i` flag, the skullstripped mean $b = 0$ image (output at step 3) via the `-r` flag, the MNI152-to-T1 warping transformation (output at step 8) via the `-w` flag, the T1-to-diffusion transformation matrix (output at step 5) via the `--postmat` flag, and the location to save the diffusion-space atlas via the `-o` flag. Specify nearest-neighbors interpolation via the `--interp=nn` flag, and specify the `--ref` flag to treat the warp field as relative.

Part 3: Generate a five-tissue-type image and extract the gray-white matter interface.

The pipeline consists of the following steps for each subject:

Part 1: Extract the mean $b = 0$ image.

1. Perform bias field correction on the diffusion MRI scan (input #2) via the N4 algorithm⁷³, using MRtrix3’s `dwibiascorrect ants` command.
2. Extract the mean $b = 0$ image from the bias-field-corrected scan (output at step 1). First, use MRtrix3’s `dwextract -bzero` command, supplying the bvals and eddy-corrected bvecs (inputs #3 and #4) via the `-fslgrad` flag. Then, average across volumes using MRtrix3’s `mrmath -axis 3 mean` command.
3. Skullstrip the mean $b = 0$ image (output at step 2) using FSL’s `bet` command with the default fractional intensity threshold (`-f`) of 0.5. Enable robust brain center estimation via the `-R` flag. Specify the `-m` flag to generate a binary brain mask in addition to the skullstripped mean $b = 0$ image.

Part 2: Nonlinearly register the atlas to diffusion space (via T1 space, to improve registration quality).

4. Convert the T1 scan (input #1) from MGZ to NIfTI format using MRtrix3’s `mrconvert` command.
5. Linearly register the T1 scan from native space to diffusion space using FSL’s `flirt` command^{74,75}. Provide the T1 scan in NIfTI-format (output at step 4) via the `-in` flag and the skullstripped mean $b = 0$ image (output at step 3) via the `-ref` flag. Use 6 degrees of freedom (`-dof 6`) for the registration. This yields two outputs: the diffusion-space T1 scan (`-out`), and the T1-to-diffusion transformation matrix (`-omat`).
6. Linearly register the T1-weighted image from native space to the MNI152 template using FSL’s `flirt` command. Provide the T1 scan in NIfTI format (output at step 4) via the `-in` flag, the MNI152 template (input #6) via the `-ref` flag, and the location to save the T1-to-MNI152 transformation matrix via the `-omat` flag. Use the default 12 degrees of freedom.
7. Nonlinearly register the T1-weighted image from native space to the MNI152 template by leveraging the linear registration from step 6, using FSL’s `fnirt` command. Provide the T1 scan in NIfTI format (output at step 4) via the `--in` flag, the T1-to-MNI152 transformation matrix (output at step 6) via the `--aff` flag, the MNI152 template (input #6) via the `--ref` flag, and the location to save the warping transformation via the `--cout` flag. Specify `--config=T1_2_MNI152_2mm.cnf` to use FSL’s internal configuration file `T1_2_MNI152_2mm.cnf`, as recommended by the creators of FSL when registering T1-weighted images to the MNI152 template.

10. Convert the FreeSurfer parcellation image (input #5) from MGZ to NIfTI format using MRtrix3’s `mrconvert` command.

11. Register the FreeSurfer parcellation image from T1 space to diffusion space using FSL’s `flirt` command. Provide the parcellation image in NIfTI format (output at step 10) via the `-in` flag, the T1-to-diffusion transformation matrix (output at step 5) via the `-applyxfm -init` flags, and the location to save the diffusion-space parcellation image via the `-out` flag. (We also provide the diffusion-space T1 scan, output at step 5, via the `-ref` flag, but only so that `flirt` knows which voxel and image dimensions to use for the output.) Use nearest-neighbor interpolation (`-interp nearest-neighbour`). Use the default 12 degrees of freedom.

12. Generate a five-tissue-type image from the diffusion-space FreeSurfer parcellation image (output at step 11) using MRtrix3’s `5ttgen freesurfer` command. Specify the `-nocrop` flag to keep the same dimensions as the input image.
13. Extract the gray-white matter interface from the five-tissue-type image (output at step 12) using MRtrix3’s `5tt2gmwmi` command.

Part 4: Perform probabilistic tractography.

14. Generate multi-shell multi-tissue response functions using MRtrix3’s `dwi2response msmt_5tt` command. Provide the diffusion MRI scan (input #2), the five-tissue-type image (output at step 12), the bvals and eddy-corrected bvecs (inputs #3 and #4) via the `-fslgrad` flag, and the binary brain mask (output at step 3) via the `-mask` flag. This yields three outputs: response functions for the white matter, gray matter and cerebrospinal fluid.
15. Estimate a fiber orientation distribution (FOD) via multi-shell multi-tissue constrained spherical deconvolution (MSMT-CSD)⁷⁰ using MRtrix3’s `dwi2fod msmt_csd` command. Provide all of the same inputs as step 14 as well as all of the outputs of step 14. This yields three outputs: orientation distribution function (ODF) images for the white matter, gray matter and cerebrospinal fluid. We will only use the white-matter ODF.
16. Perform probabilistic tractography using MRtrix3’s `tckgen` command. Provide the white-matter ODF (output at step 15). Use anatomically-constrained tractography (ACT)¹⁷⁶ by providing the five-tissue-type image (output at step 12) via the `-act` flag. Seed streamlines from the gray-white matter interface (output at step 13) by providing it via the `-seed_gmwmi` flag. Use the Second-order Integration over Fiber Orientation Distributions (iFOD2) algorithm⁷¹ by specifying `-algorithm iFOD2`. Select 1 million streamlines (`-select 1000000`).
17. Weight each streamline so that the density of reconstructed connections better reflects the density of the underlying white

- matter fibers via the SIFT2 algorithm⁷², as implemented in MRtrix3's `tcksift2` command. Provide the streamlines (output at step 16) and the white-matter ODF (output at step 15).
18. Generate a connectome matrix using MRtrix3's `tck2connectome` command, denoting the weighted number of streamlines connecting each pair of parcels in the atlas (in our case, this matrix has dimension 214×214). Provide the streamlines (output at step 16) and the diffusion-space atlas (output at step 9). Also provide the weight for each streamline (output at step 17) via the `-tck_weights_in` flag. Perform a radial search from each streamline endpoint to locate the nearest parcel by specifying the `-assignment_radial_search` flag¹⁷⁷. Require the matrix to be symmetric and have a diagonal of 0 via the `-symmetric` and `-zero_diagonal` flags. Weighted the entries in the connectome matrix by both the length of the streamline (`-scale_length`), to account for the bias of probabilistic tractography towards shorter connections, and the inverse of the product of the two parcel volumes (`-scale_invnodevol`)¹⁷⁸, to correct for the variability in the size of parcels that would otherwise result in higher connectivity to/from larger parcels solely due to their size.

Phenotype definitions

We defined a total of 3 hemisphere-level connectivity measures (left intra-hemisphere, right intra-hemisphere, inter-hemisphere). Specifically, we averaged the entries of the connectome matrix across all pairs of parcels in the left hemisphere to obtain the overall connectivity within the left hemisphere, and similarly for the right hemisphere. We quantified overall inter-hemispheric connectivity by averaging the entries connecting any parcel in the left hemisphere to any parcel in the right hemisphere.

We defined a total of 105 network-level connectivity measures. Each parcel in the Schaefer atlas has previously been mapped to one of 7 large-scale brain networks, commonly known as the “Yeo 7” networks, which were derived in a data-driven manner based on functional connectivity⁶⁷: visual (“Vis”), somatomotor (“SomMot”), dorsal attention (“DorsAttn”), salience/ventral attention (“SalVentAttn”), limbic, control (“Cont”), and default mode (“Default”). Because each network is present in both hemispheres, there are 14 hemisphere-specific networks, denoted e.g. “LH Vis” for the left-hemisphere visual network and “RH Cont” for the right-hemisphere control network. For each network, we averaged the entries of the connectome matrix across all pairs of parcels in that network to define 14 within-network connectivity measures. Also, for each pair of networks, we averaged the entries of the connectome matrix connecting any parcel in the first network to any parcel in the second network, to define 91 between-network connectivity measures.

Finally, we defined a total of 98 cortical-to-subcortical connectivity measures, between each of the 14 hemisphere-specific Yeo 7 networks and each of the 7 subcortical structures (thalamus, caudate, putamen, pallidum, hippocampus, amygdala, and accumbens).

Genome-wide association studies

We performed genome-wide association studies for the 206 connectivity measures by linearly regressing each measure on each of 9,423,516 single-nucleotide genetic variants (see below) using version 3.2.9 of the regenie genetic analysis toolkit (<https://rgcg.github.io/regenie>)⁷⁴. regenie accounts for sample relatedness and population structure by computing a polygenic risk score (PRS) for the phenotype being associated (step 1), then including this PRS as a covariate when performing the actual association testing (step 2). regenie uses a leave-one-chromosome-out scheme to construct the PRS (e.g. when performing association tests on chromosome 19, a PRS derived from all chromosomes except chromosome 19 is used as a covariate).

For step 1, we used a quality-controlled subset of the UK Biobank's unimputed genotype data, namely the 596,935 variants on the

autosomes and the non-pseudoautosomal regions of the X chromosome with minor allele frequency $>1\%$, $<10\%$ missingness and Hardy-Weinberg equilibrium $p > 1 \times 10^{-15}$ (with mid-P correction) across all 425,630 UK Biobank participants of European genetic ancestry (according to Pan-UK Biobank, Return 2442) deemed suitable for genetic analyses (according to the criteria stated in the “Cohort selection and MRI quality control” section, above). Our missingness and Hardy-Weinberg equilibrium cutoffs are the same as those used by the flagship UK Biobank exome sequencing study⁷⁹. We performed this quality control using version 2.0.0 of the plink GWAS software package¹⁸⁰. For computational efficiency, regenie constructs the PRS using stacked ridge regression, by partitioning the genome into non-overlapping blocks of 200 markers (where 200 is specified using the “-bsize” option), running ridge regression within each block, and then running a second level of ridge regression to aggregate across blocks.

We ran step 2 on a larger set of 9,423,516 genotyped or imputed autosomal and non-pseudoautosomal X-chromosome variants with INFO score >0.8 , using the same quality control filters as in step 1: minor allele frequency $>1\%$, $<10\%$ missingness and Hardy-Weinberg equilibrium $p > 1 \times 10^{-15}$ (with mid-P correction). As described in the flagship UK Biobank study⁹, variants were imputed to two reference panels, the Haplotype Reference Consortium⁷⁵ and a combined UK10K⁷⁶ and 1000 Genomes Phase 3⁷⁷ panel; the imputed genotypes for the two panels were then combined, using the HRC imputation for variants present in both panels.

We covaried for age (“Age when attended assessment center”, Data-Field 21003, standardized to mean 0 and variance 1), sex, age \times sex, age², age² \times sex, genotyping array (Axiom versus BiLEVE), scanner site ($n = 22$, “UK Biobank assessment center”, Data-Field 54), total intracranial volume (“Volume of EstimatedTotalIntraCranial (whole brain)”, Data-Field 26521; median-imputed for 9 people for whom it was missing), and the top 10 genotype principal components (“Genetic principal components”, Data-Field 22009) to control for population structure. Prior to running regenie, we applied a rank-based inverse normal transformation to the residuals using the Blom transformation ($c = 3/8$)¹⁸¹. This ensures that the GWAS phenotype is normally distributed even though the connectivity measures are highly non-normally distributed, nor are they particularly well approximated by a log-normal distribution (Supplementary Fig. 7).

Regenie assumes full dosage compensation for variants on the X chromosome (except for variants in the pseudoautosomal region, which were analyzed the same way as autosomal variants), coding alleles as 0/1/2 for females and 0/2 for males¹⁸²; we excluded the Y chromosome.

Identification of independent genome-wide significant variants

To identify independent genome-wide significant variants, we clumped together variants within 5 megabases and with linkage disequilibrium (LD) $R^2 > 0.01$, using the `--clump` command from plink version 2.0.0 (<https://www.cog-genomics.org/plink/2.0/postproc#clump>), with options `--clump-p1 5e-8 --clump-p2 0.001 --clump-r2 0.01 --clump-kb 5000`. For Fig. 3, we performed this clumping on the minimum p value for each variant across the 206 measures (i.e. the p -values shown in the Manhattan plots). For Supplementary Data 1, we instead performed this clumping separately for each of the 206 GWAS.

Ideogram of associated genomic loci

We plotted the ideogram of associated genomic loci in Fig. 1 with PhenoGram (<http://visualization.ritchielab.org/phenograms/plot>)¹⁸³.

p value inflation and linkage disequilibrium score regression intercepts and heritabilities

We observed acceptably mild p value inflation, with genomic inflation factors (λ_{GC}) of 1.014 to 1.093 (median 1.044) across the 206 GWAS. Most of this inflation was due to polygenicity rather than confounding,

with linkage disequilibrium (LD) score regression intercepts¹⁴⁸ of only 0.996 to 1.016 (median 1.006).

We used the `--h2` command from the `ldsc` software package (<https://github.com/bulik/ldsc>) to compute these LD score regression intercepts, as well as the heritabilities in Supplementary Fig. 2. Before running `--h2`, we used `ldsc's munge_sumstats.py` script to reformat each GWAS's summary statistics into the format recognized by `ldsc`. Rather than using predefined LD scores, we elected to compute our own LD scores for the 9,423,516 variants in our GWAS from the UK Biobank's imputed genotypes themselves. We did so by (1) randomly subsampling 1000 of the 26,333 participants (for computational tractability), (2) converting imputed variants to hardcalls (since `ldsc` does not support fractional genotypes), and (3) running `ldsc's --l2` command with the UK Biobank's hardcall genotypes provided via the `--bfile` flag. When running `--l2`, we used a 1-megabase window (`--ld-window-kb 1000`) instead of the usual 1-centimorgan window (`--ld-window-cm 1`), since the UK Biobank's imputed genotype files do not report centimorgan distances.

Heritability, polygenicity and selection analysis with SBayesS

We used the `SBayesS` method¹⁴⁷ to estimate heritability (h^2), polygenicity (π) and selection (S) parameters for each of our 206 GWAS. We ran the `--sbayes S` command from the Genome-wide Complex Trait Bayesian analysis (GCTB) software package (<https://cnsgenomics.com/software/gctb>) on each GWAS's summary statistics, with the "Sparse matrix (including MHC regions)" from <https://cnsgenomics.com/software/gctb/#LDmatrices> (`ukbEURu_imp_v3_HM3_n50k.chisq10.ldm.sparse`) as the LD matrix and a fixed random seed of 0. `SBayesS` uses Markov Chain Monte Carlo (MCMC) sampling to infer h^2 , π and S ; we used the default settings of running 10,000 iterations of the Markov chain (`--chain-length 10000`) and sampling the parameters from every 10th iteration (`--thin 10`), starting at iteration 2000 (`--burn-in 2000`), thus generating 800 estimates of each parameter (i.e. from iterations 2000, 2010, 2020, ... 9990). We obtained a single estimate per parameter by taking the mean across the 800 estimates, and confidence intervals by taking the 2.5th and 97.5th percentiles of these 800 estimates.

Genetic correlation analysis

We used the genetic covariance analyzer (GNOVA) method¹⁴⁹ (github.com/qlu-lab/GNOVA-2.0) to compute genetic correlations between the 206 structural connectivity measures and 15 brain-related traits:

1. Major depression¹⁸⁴, with summary statistics from the PGC ("mdd2019edinburgh" at <https://www.med.unc.edu/pgc/download-results>, under "Genome-wide summary statistics from a meta-analysis of PGC and UK Biobank", filename `PGC_UKB_depression_genome-wide.txt`)
2. Bipolar disorder¹⁸⁵, with summary statistics from the PGC ("bip2021" at <https://www.med.unc.edu/pgc/download-results>, filename `pgc-bip2021-all.vcf.tsv.gz`)
3. Schizophrenia⁸⁸, with summary statistics from the PGC ("scz2022" at <https://www.med.unc.edu/pgc/download-results>, filenames `PGC3_SCZ_wave3.primary.autosome.public.v3.vcf.tsv.gz` and `PGC3_SCZ_wave3.primary.chrX.public.v3.vcf.tsv.gz`)
4. Autism spectrum disorder (ASD)¹⁸⁶, with summary statistics from the PGC ("asd2019" at <https://www.med.unc.edu/pgc/download-results>, filename `iPSYCHPGC_ASD_Nov2017.gz`)
5. Attention-deficit hyperactivity disorder (ADHD)¹⁸⁷, with summary statistics from the PGC and iPSYCH ("Summary statistics ADHD meta-analysis, Jan2022 Release" at <https://ipsych.dk/en/research/downloads>, filename `ADHD_meta_Jan2022_iPSYCH1_iPSYCH2_deCODE_PGC.meta.2.zip`)
6. Problematic alcohol use¹⁸⁸, with summary statistics from the Million Veteran Program (available through the Database of Genotypes and Phenotypes (dbGaP) by application at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001672.v6.p1, filename `PAU_MVPI_MVP2_PGC_UKB_Dec11.txt.gz`)

7. Insomnia¹⁸⁹, with summary statistics from https://ctg.cncr.nl/software/summary_statistics (filename `insomnia_ukb2b_EUR_sumstats_20190311_with_chrX_mac_100.txt.gz`)
8. Alzheimer's disease¹⁹⁰, with summary statistics from the GWAS Catalog (accession `GCST90027158`, filename `GCST90027158_buildGRCh38.tsv.gz`)
9. Parkinson's disease¹⁹¹, with summary statistics from <https://drive.google.com/file/d/1FZ9UL99LAqyWnyNBxxlx6qOUlIfAnubin> (filename `nallsEtAl2019_excluding23andMe_allVariants.tab.gz`)
10. Amyotrophic lateral sclerosis (ALS)¹⁹², with summary statistics from the GWAS Catalog (accession `GCST90027164`, filename `GCST90027164_buildGRCh37.tsv.gz`)
11. Epilepsy¹⁹³, with summary statistics from http://www.epigad.org/download/final_sumstats.zip, filename `ILAE3_Caucasian_GGE_final.tbl`. We used summary statistics for genetic generalized epilepsy (GGE), which had substantially greater GWAS signal than non-subtype-specific epilepsy.
12. Stroke¹⁹⁴, with summary statistics from the GWAS Catalog (accession `GCST90104539`, filename `GCST90104539_buildGRCh37.tsv.gz`)
13. Educational attainment¹⁹⁵, with summary statistics from <https://www.thessgac.com> under "Summary Statistics for Okbay et al. (2022)" (login required, filename `EA4_additive_excl_23andMe.txt.gz`)
14. Intelligence¹⁹⁶, with summary statistics from https://ctg.cncr.nl/software/summary_statistics (filename `SavageJansen_2018_intelligence_metaanalysis.txt.gz`)
15. Reaction time¹⁹⁷, with summary statistics from the GWAS Catalog (accession `GCST006268`, filename `Davies2018_UKB_RT_summary_results_29052018.txt`). The sign of the GWAS beta coefficients are flipped in this GWAS, so that a positive beta indicates that a variant is associated with lower (i.e. better) reaction time, and a negative beta indicates that a variant is associated with higher (i.e. worse) reaction time.

To account for case-control imbalance in analyses of case-control GWAS, the effective sample size $N_{eff} = 4 / (1 / N_{cas} + 1 / N_{con})$ can be used as a measure of sample size, where N_{cas} and N_{con} are either total number of cases and controls in the GWAS, or (more accurately) the number of cases and controls with non-missing genotypes for each variant. For meta-analyses of multiple GWAS cohorts, it has been argued that N_{eff} should be computed per cohort and then summed across cohorts¹⁹⁸, since each cohort has a different case-control ratio. Unfortunately, most publicly available case-control summary statistics either do not provide N_{eff} or compute it using the total N_{cas} and N_{con} instead of the per-cohort N_{cas} and N_{con} , which can lead to substantial bias in estimates of heritability (though not necessarily genetic correlation)¹⁹⁸. In these cases, we inferred N_{eff} for each variant using the formula $N_{eff} = ((4 / (2 \times MAF \times (1 - MAF) \times INFO)) - BETA^2) / SE^2$, where MAF is the variant's minor allele frequency in the GWAS sample, $INFO$ its imputation information score, $BETA$ its effect size, and SE the standard error of this effect size¹⁹⁹. If $INFO$ was not provided, we excluded the $INFO$ term. If SE was not provided, we back-calculated it using the variant's effect size and p value. If MAF was not provided (or was only provided for a different sample, such as 1000 Genomes), we fell back to using each variant's total N_{cas} and N_{con} across all cohorts (or, if per-variant N_{cas} and N_{con} were not available, the total N_{cas} and N_{con} across all cohorts) and calculated N_{eff} as $4 / (1 / N_{cas} + 1 / N_{con})$. For our own structural connectivity GWASs and other quantitative-trait GWASs (e.g. reaction time), we did not compute an effective sample size and simply used the total sample size.

We harmonized the alleles of the 15 summary statistics, as well as our own GWAS summary statistics, to the European subset of 1000

Genomes Phase 3⁷⁷ (storage.googleapis.com/broad-alkesgroup-public/LDSCORE/1000G_Phase3_plinkfiles.tgz) using `ldsc's munge_sumstats.py` script. We then removed variants with missing data so that the munged summary statistics would be compatible with GNOVA. Finally, we ran GNOVA to compute genetic correlations, specifying the European subset of 1000 Genomes Phase 3 as a reference panel for computing LD scores.

We also used GNOVA to perform genetic correlations between our three hemisphere-level measures and 432 GWAS of diffusion MRI measures based on tract-based spatial statistics (TBSS) and neurite orientation dispersion and density imaging (NODDI)¹⁸. We obtained summary statistics for these GWAS from the table at open.win.ox.ac.uk/ukbiobank/big40/BIG40-IDPs_v4/IDPs.html, subsetting to the 432 diffusion MRI measures in the UK Biobank's "dMRI skeleton" phenotype category (Category 134; biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=134). We followed the same procedure as for the 15 summary statistics above, using the N(all) column in each summary statistics file as the sample size.

Cell-type enrichment analysis

We used partitioned LD score regression¹⁵¹ from the `ldsc` software package to perform cell-type enrichment analyses across human cortical cell types. We used two chromatin accessibility datasets, one adult and one developmental.

First, we included regions for 6 cell types from postmortem adult human motor cortex, derived from single-nucleus chromatin accessibility and messenger RNA expression sequencing (SNARE-seq2). We obtained these regions from Table S14c of¹⁵²: "SNARE-Seq2 Differentially Accessible Regions (DARs, q value < 0.001 and $\log(\text{fold change}) > 1$) identified by subclass for human M1". Since these regions were only available for neuronal subclasses, but a subclass-specific analysis would likely be underpowered, we merged regions across subclasses. We obtained excitatory neuron-specific regions by merging regions specific to layer 2–3 intratelencephalic (L2-3 IT), layer 5 extratelencephalic (L5 ET), layer 5 intratelencephalic (L5 IT), layer 5-6 non-projecting pyramidal (L5-6 NP), layer 6 corticothalamic (L6 CT), layer 6 intratelencephalic (L6 IT), *CAR3*-expressing layer 6 intratelencephalic (L6 IT Car3), and layer 6b (L6b) excitatory neurons. We obtained inhibitory neuron-specific regions by merging regions specific to LAMP5, PVALB, SNCG, SST, SST CHODL, and VIP inhibitory neurons. We note that this choice does not capture excitatory neuron-specific regions that are present in multiple excitatory neuron subclasses, or inhibitory neuron-specific regions that are present in multiple inhibitory neuron subclasses. Besides excitatory and inhibitory, we included the following non-neuronal cell types in our analysis: astrocytes; microglia and perivascular macrophages (Micro-PVM), which we refer to as 'microglia' for simplicity; oligodendrocytes (Oligo); and oligodendrocyte progenitor cells (OPC). We did not include regions specific to vascular leptomeningeal cells (VLMCs) and endothelial cells due to their rarity. Thus, we included a total of 6 cell types in the analysis.

Second, we included regions for 12 cell types from the developing human cortex, derived from single-cell assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq). We obtained these regions from Table S2 of¹⁵⁴.

We created a binary annotation file for each of these 18 cell types, with a 1 or 0 for each variant depending on whether or not it was situated in a region with increased chromatin accessibility in that cell type. For instance, for the SNARE-Seq2 dataset, 29,398 of the 9,423,516 variants in our GWAS (0.35%) were situated in astrocyte-specific open chromatin regions, 112,321 (1.33%) in excitatory-specific regions, 87,192 (1.03%) in inhibitory-specific regions, 35,754 (0.42%) in microglia-specific regions, 30,394 (0.36%) in oligodendrocyte-specific regions, and 36,246 (0.43%) in OPC-specific regions. We also created an "intercept" annotation file in which all of the 9,423,516 variants were assigned a 1.

We computed LD scores for each of these binary annotation files by applying `ldsc's --l2` command with a 1-megabase window (`--ld-wind-kb 1000`), providing the UK Biobank's hardcall genotypes with the `--bfile` flag, and providing the annotation file with the `--annot` and `--thin-annot` flags. Aside from providing an annotation file, this is the same way we computed LD scores for all variants (see the "p value inflation and linkage disequilibrium score regression intercepts" subsection, above). Note that the LD scores for the intercept annotation are the same as the LD scores for all variants.

Finally, for each of the 206 structural connectivity GWAS and each of the 18 cell types, we ran stratified LD score regression using `ldsc's --h2` command. We provided the GWAS's summary statistics under the `--h2` flag, the LD scores for both the intercept and the cell type under the `--ref-ld` flag, the LD scores for the intercept under the `--w-ld` flag, and the allele frequencies for each variant across our 26,333 participants (computed with `plink 1.9's --freq` command) under the `--frqfile` flag. We also specified the `--overlap-annot` flag to indicate that the two annotations (for the intercept and the cell type) overlap with each other, and the `--print-coefficients` flag to print the regression coefficients for each annotation.

This analysis yields an enrichment (and a standard error and p value for this enrichment) for each GWAS and cell type. This enrichment represents how many times greater heritability is explained by the average variant overlapping a region with increased chromatin accessibility in that cell type, relative to the average variant not overlapping such a region. The use of LD score regression makes this analysis robust to the confounding effects of LD: without it, an annotation could erroneously be considered enriched for heritability even if the entirety of the GWAS signal underlying this enrichment were coming from variants in LD not overlapping the annotation.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Researchers can apply for access to the UK Biobank at ukbiobank.ac.uk/enable-your-research/apply-for-access. Genome-wide summary statistics have been deposited to the European Bioinformatics Institute GWAS Catalog (<https://www.ebi.ac.uk/gwas>) under accession numbers GCST90302648 through GCST90302853. The 206 structural connectivity measures for each participant will be made available through the UK Biobank Returns Catalog (biobank.ndph.ox.ac.uk/ukb/docs.cgi?id=1) to all researchers with UK Biobank access.

Code availability

The tractography pipeline used in this study, `tractify`, is available at <https://github.com/TIGRLab/tractify>.

References

- Zhang, K. & Sejnowski, T. J. A universal scaling law between gray matter and white matter of cerebral cortex. *Proc. Natl. Acad. Sci. Usa.* **97**, 5621–5626 (2000).
- Purves, D. et al. Increased Conduction Velocity as a Result of Myelination. in *Neuroscience. 2nd edition* (Sinauer Associates, 2001).
- Sporns, O., Tononi, G. & Kötter, R. The human connectome: A structural description of the human brain. *PLoS Comput. Biol.* **1**, e42 (2005).
- Bassett, D. S. & Sporns, O. Network neuroscience. *Nat. Neurosci.* **20**, 353–364 (2017).
- Moseley, M. E. et al. Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. *Radiology* **176**, 439–445 (1990).

6. Doran, M. et al. Normal and abnormal white matter tracts shown by MR imaging using directional diffusion weighted sequences. *J. Comput. Assist. Tomogr.* **14**, 865–873 (1990).
7. Aung, W. Y., Mar, S. & Benzinger, T. L. S. Diffusion tensor MRI as a biomarker in axonal and myelin damage. *Imaging Med.* **5**, 427 (2013).
8. Thompson, P. M. et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
9. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
10. Hibar, D. P. et al. Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–229 (2015).
11. Adams, H. H. et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci.* **19**, 1569–1582 (2016).
12. Hibar, D. P. et al. Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 1–12 (2017).
13. Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
14. Zhao, B. 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 (2019).
15. Zhao, B. 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 (2019).
16. Satizabal, C. L. et al. Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet.* **51**, 1624–1636 (2019).
17. Grasby, K. L. et al. The genetic architecture of the human cerebral cortex. *Science* **367**, eaay6690 (2020).
18. Smith, S. M. et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat. Neurosci.* **24**, 737 (2021).
19. Zhao, B. et al. Common genetic variation influencing human white matter microstructure. *Science* **372**, eabf3736 (2021).
20. Sha, Z. et al. The genetic architecture of structural left–right asymmetry of the human brain. *Nat. Hum. Behav.* **5**, 1226–1239 (2021).
21. Chambers, T. et al. Genetic common variants associated with cerebellar volume and their overlap with mental disorders: a study on 33,265 individuals from the UK-Biobank. *Mol. Psychiatry* **27**, 2282–2290 (2022).
22. Makowski, C. et al. Discovery of genomic loci of the human cerebral cortex using genetically informed brain atlases. *Science* **375**, 522–528 (2022).
23. Brouwer, R. M. et al. Genetic variants associated with longitudinal changes in brain structure across the lifespan. *Nat. Neurosci.* **25**, 421–432 (2022).
24. Zhao, B. et al. Common variants contribute to intrinsic human brain functional networks. *Nat. Genet.* **54**, 508–517 (2022).
25. Wang, C. et al. Phenotypic and genetic associations of quantitative magnetic susceptibility in UK Biobank brain imaging. *Nat. Neurosci.* **25**, 818–831 (2022).
26. Warriar, V. et al. Genetic insights into human cortical organization and development through genome-wide analyses of 2,347 neuroimaging phenotypes. *Nat. Genet.* **55**, 1483–1493 (2023).
27. Kochunov, P. et al. Relationship between white matter fractional anisotropy and other indices of cerebral health in normal aging: tract-based spatial statistics study of aging. *Neuroimage* **35**, 478–487 (2007).
28. Conturo, T. E. et al. Tracking neuronal fiber pathways in the living human brain. *Proc. Natl. Acad. Sci. Usa.* **96**, 10422 (1999).
29. Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J. & Aldroubi, A. In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* **44**, 625–632 (2000).
30. Jeurissen, B., Descoteaux, M., Mori, S., Leemans, A. & Diffusion, M. R. I. fiber tractography of the brain. *NMR Biomed.* **32**, e3785 (2019).
31. Medaglia, J. D., Lynall, M. E. & Bassett, D. S. Cognitive network neuroscience. *J. Cogn. Neurosci.* **27**, 1471–1491 (2015).
32. Basile, G. A. et al. White matter substrates of functional connectivity dynamics in the human brain. *Neuroimage* **258**, 119391 (2022).
33. Zhang, F. et al. Quantitative mapping of the brain’s structural connectivity using diffusion MRI tractography: A review. *Neuroimage* **249**, 118870 (2022).
34. van den Heuvel, M. P., Mandl, R. C. W., Stam, C. J., Kahn, R. S. & Hulshoff Pol, H. E. Aberrant Frontal and Temporal Complex Network Structure in Schizophrenia: A Graph Theoretical Analysis. *J. Neurosci.* **30**, 15915–15926 (2010).
35. Zalesky, A. et al. Disrupted axonal fiber connectivity in schizophrenia. *Biol. Psychiatry* **69**, 80–89 (2011).
36. Wang, Q. et al. Anatomical insights into disrupted small-world networks in schizophrenia. *Neuroimage* **59**, 1085–1093 (2012).
37. van den Heuvel, M. P. et al. Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry* **70**, 783–792 (2013).
38. Zhang, R. et al. Disrupted brain anatomical connectivity in medication-naïve patients with first-episode schizophrenia. *Brain Struct. Funct.* **220**, 1145–1159 (2015).
39. Griffa, A. et al. Characterizing the connectome in schizophrenia with diffusion spectrum imaging. *Hum. Brain Mapp.* **36**, 354 (2015).
40. Zhao, X. et al. Abnormal rich-club organization associated with compromised cognitive function in patients with Schizophrenia and their unaffected parents. *Neurosci. Bull.* **33**, 445–454 (2017).
41. Roberts, G. et al. Structural dysconnectivity of key cognitive and emotional hubs in young people at high genetic risk for bipolar disorder. *Mol. Psychiatry* **23**, 413–421 (2018).
42. Deng, Y. et al. Tractography-based classification in distinguishing patients with first-episode schizophrenia from healthy individuals. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **88**, 66–73 (2019).
43. Cui, L.-B. et al. Connectome-based patterns of first-episode medication-naïve patients with Schizophrenia. *Schizophr. Bull.* **45**, 1291–1299 (2019).
44. Ji, E. et al. Increased and decreased superficial white matter structural connectivity in Schizophrenia and bipolar disorder. *Schizophr. Bull.* **45**, 1367–1378 (2019).
45. van den Heuvel, M. P. et al. Evolutionary modifications in human brain connectivity associated with schizophrenia. *Brain* **142**, 3991–4002 (2019).
46. Levitt, J. J. et al. Miswiring of frontostriatal projections in Schizophrenia. *Schizophr. Bull.* **46**, 990–998 (2020).
47. Houenou, J. et al. Increased white matter connectivity in euthymic bipolar patients: diffusion tensor tractography between the subgenual cingulate and the amygdalo-hippocampal complex. *Mol. Psychiatry* **12**, 1001–1010 (2007).
48. Sweet, J. A. et al. Cingulum bundle connectivity in treatment-refractory compared to treatment-responsive patients with bipolar disorder and healthy controls: a tractography and surgical targeting analysis. *J. Neurosurg.* 1–13 (2022).
49. Korgaonkar, M. S., Fornito, A., Williams, L. M. & Grieve, S. M. Abnormal structural networks characterize major depressive disorder: A connectome analysis. *Biol. Psychiatry* **76**, 567–574 (2014).
50. Hong, S. B. et al. Connectomic disturbances in attention-deficit/hyperactivity disorder: a whole-brain tractography analysis. *Biol. Psychiatry* **76**, 656–663 (2014).

51. Cha, J. et al. Neural correlates of aggression in medication-naïve children with ADHD: Multivariate analysis of morphometry and tractography. *Neuropsychopharmacology* **40**, 1717–1725 (2015).
52. Griffiths, K. R. et al. Structural brain network topology underpinning ADHD and response to methylphenidate treatment. *Transl. Psychiatry* **11**, 1–9 (2021).
53. Thomas, C., Humphreys, K., Jung, K.-J., Minshew, N. & Behrmann, M. The anatomy of the callosal and visual-association pathways in high-functioning autism: a DTI tractography study. *Cortex* **47**, 863–873 (2011).
54. Hong, S. et al. Detecting abnormalities of corpus callosum connectivity in autism using magnetic resonance imaging and diffusion tensor tractography. *Psychiatry Res.* **194**, 333–339 (2011).
55. Nickel, K. et al. Altered transcallosal fiber count and volume in high-functioning adults with autism spectrum disorder. *Psychiatry Res Neuroimaging* **322**, 111464 (2022).
56. Lo, C.-Y. et al. Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease. *J. Neurosci.* **30**, 16876–16885 (2010).
57. Prasad, G., Nir, T. M., Toga, A. W. & Thompson, P. M. Tractography density and network measures in Alzheimer's disease. *2013 IEEE 10th Int. Symp. Biomed. Imaging* **2013**, 692–695 (2013).
58. Zhan, L. et al. Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease. *Front. Aging Neurosci.* **7**, 48 (2015).
59. Shu, N. et al. Diffusion tensor tractography reveals disrupted topological efficiency in white matter structural networks in multiple sclerosis. *Cereb. Cortex* **21**, 2565–2577 (2011).
60. Li, Y. et al. Diffusion tensor imaging based network analysis detects alterations of neuroconnectivity in patients with clinically early relapsing-remitting multiple sclerosis. *Hum. Brain Mapp.* **34**, 3376–3391 (2013).
61. Shu, N. et al. Disrupted topological organization of structural and functional brain connectomes in clinically isolated syndrome and multiple sclerosis. *Sci. Rep.* **6**, 29383 (2016).
62. Fleischer, V. et al. Increased structural white and grey matter network connectivity compensates for functional decline in early multiple sclerosis. *Mult. Scler.* **23**, 432–441 (2017).
63. Llufriu, S. et al. Structural networks involved in attention and executive functions in multiple sclerosis. *Neuroimage Clin.* **13**, 288–296 (2017).
64. Shu, N. et al. Progressive brain rich-club network disruption from clinically isolated syndrome towards multiple sclerosis. *Neuroimage Clin.* **19**, 232–239 (2018).
65. Schiavi, S. et al. Sensory-motor network topology in multiple sclerosis: Structural connectivity analysis accounting for intrinsic density discrepancy. *Hum. Brain Mapp.* **41**, 2951–2963 (2020).
66. Schiavi, S. et al. Classification of multiple sclerosis patients based on structural disconnection: A robust feature selection approach. *J. Neuroimaging* <https://doi.org/10.1111/jon.12991>. (2022).
67. Thomas Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125 (2011).
68. Jahanshad, N. et al. Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. *Proc. Natl Acad. Sci. USA.* **110**, 4768–4773 (2013).
69. Tournier, J.-D. et al. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* **202**, 116137 (2019).
70. Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A. & Sijbers, J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *Neuroimage* **103**, 411–426 (2014).
71. Tournier, J.-D., Calamante, F. & Connelly, A. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. *Proc. Int. Soc. Magn. Reson. Med.* **18**, 1670 (2010).
72. Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *Neuroimage* **119**, 338–351 (2015).
73. Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
74. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
75. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
76. UK10K Consortium. et al. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
77. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
78. Steinhart, Z. & Angers, S. Wnt signaling in development and tissue homeostasis. *Development* **145**, dev146589 (2018).
79. Ekici, A. B. et al. Disturbed Wnt Signalling due to a Mutation in CCDC88C Causes an Autosomal Recessive Non-Syndromic Hydrocephalus with Medial Diverticulum. *MSY* **1**, 99–112 (2010).
80. Ruggeri, G. et al. Bi-allelic mutations of CCDC88C are a rare cause of severe congenital hydrocephalus. *Am. J. Med. Genet. A* **176**, 676–681 (2018).
81. Marguet, F. et al. Neuropathological hallmarks of fetal hydrocephalus linked to CCDC88C pathogenic variants. *Acta Neuropathol. Commun.* **9**, 104 (2021).
82. Zhou, F. et al. Gmnc is a master regulator of the multiciliated cell differentiation program. *Curr. Biol.* **25**, 3267–3273 (2015).
83. Kyrrousi, C. et al. Mcidas and GemC1 are key regulators for the generation of multiciliated ependymal cells in the adult neurogenic niche. *Development* **142**, 3661–3674 (2015).
84. Lalioti, M.-E. et al. GemC1 is a critical switch for neural stem cell generation in the postnatal brain. *Glia* **67**, 2360–2373 (2019).
85. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
86. Wang, C.-Y. et al. ZIP8 is an iron and zinc transporter whose cell-surface expression is up-regulated by cellular iron loading. *J. Biol. Chem.* **287**, 34032 (2012).
87. Park, J. H. et al. SLC39A8 Deficiency: A Disorder of Manganese Transport and Glycosylation. *Am. J. Hum. Genet.* **97**, 894 (2015).
88. Trubetskov, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
89. Mealer, R. G. et al. The schizophrenia-associated variant in SLC39A8 alters protein glycosylation in the mouse brain. *Mol. Psychiatry* **27**, 1405–1415 (2022).
90. Lieberman, J. A. et al. Hippocampal dysfunction in the pathophysiology of schizophrenia: a selective review and hypothesis for early detection and intervention. *Mol. Psychiatry* **23**, 1764–1772 (2018).
91. Courchet, J. et al. Terminal axon branching is regulated by the LKB1-NUAK1 kinase pathway via presynaptic mitochondrial capture. *Cell* **153**, 1510–1525 (2013).
92. Courchet, V. et al. Haploinsufficiency of autism spectrum disorder candidate gene NUAK1 impairs cortical development and behavior in mice. *Nat. Commun.* **9**, 1–14 (2018).
93. Lanfranchi, M. et al. The AMPK-related kinase NUAK1 controls cortical axons branching through a local modulation of mitochondrial metabolic functions. *bioRxiv* 2020.05.18.102582 <https://doi.org/10.1101/2020.05.18.102582>. (2020).
94. Castets, F. et al. A novel calmodulin-binding protein, belonging to the WD-repeat family, is localized in dendrites of a subset of CNS neurons. *J. Cell Biol.* **134**, 1051–1062 (1996).

95. Bartoli, M., Monneron, A. & Ladant, D. Interaction of calmodulin with striatin, a WD-repeat protein present in neuronal dendritic spines. *J. Biol. Chem.* **273**, 22248–22253 (1998).
96. Li, D., Musante, V., Zhou, W., Picciotto, M. R. & Nairn, A. C. Striatin-1 is a B subunit of protein phosphatase PP2A that regulates dendritic arborization and spine development in striatal neurons. *J. Biol. Chem.* **293**, 11179 (2018).
97. Salamon, I. et al. Celf4 controls mRNA translation underlying synaptic development in the prenatal mammalian neocortex. *Nat. Commun.* **14**, 6025 (2023).
98. Bryant, C. D. & Yazdani, N. RNA-binding proteins, neural development and the addictions. *Genes Brain Behav.* **15**, 169–186 (2016).
99. Wagnon, J. L. et al. CELF4 Regulates Translation and Local Abundance of a Vast Set of mRNAs, Including Genes Associated with Regulation of Synaptic Function. *PLoS Genet.* **8**, e1003067 (2012).
100. Desprez, F., Ung, D. C., Vourc'h, P., Jeanne, M. & Laumonnier, F. Contribution of the dihydropyrimidinase-like proteins family in synaptic physiology and in neurodevelopmental disorders. *Front. Neurosci.* **17**, 1154446 (2023).
101. Xiong, L.-L. et al. DPYSL2 is a novel regulator for neural stem cell differentiation in rats: revealed by Panax notoginseng saponin administration. *Stem Cell Res. Ther.* **11**, 1–16 (2020).
102. Nomoto, M. et al. Clinical evidence that a dysregulated master neural network modulator may aid in diagnosing schizophrenia. *Proc. Natl Acad. Sci. USA.* **118**, e2100032118 (2021).
103. Izumi, R. et al. Dysregulation of DPYSL2 expression by mTOR signaling in schizophrenia: Multi-level study of postmortem brain. *Neurosci. Res.* **175**, 73–81 (2022).
104. Feuer, K. L., Peng, X., Yovo, C. K. & Avramopoulos, D. DPYSL2/CRMP2 isoform B knockout in human iPSC-derived glutamatergic neurons confirms its role in mTOR signaling and neurodevelopmental disorders. *Mol. Psychiatry* <https://doi.org/10.1038/s41380-023-02186-w>. (2023).
105. Nishikimi, M., Oishi, K., Tabata, H., Torii, K. & Nakajima, K. Segregation and pathfinding of callosal axons through EphA3 signaling. *J. Neurosci.* **31**, 16251 (2011).
106. Shi, G., Yue, G. & Zhou, R. EphA3 functions are regulated by collaborating phosphotyrosine residues. *Cell Res.* **20**, 1263–1275 (2010).
107. Brennaman, L. H. et al. Polysialylated NCAM and EphrinA/EphA regulate synaptic development of GABAergic interneurons in prefrontal cortex. *Cereb. Cortex* **23**, 162 (2013).
108. Sullivan, C. S., Kümper, M., Temple, B. S. & Maness, P. F. The neural cell adhesion molecule (NCAM) promotes clustering and activation of EphA3 receptors in GABAergic interneurons to induce ras homolog gene family, member A (RhoA)/Rho-associated protein kinase (ROCK)-mediated growth cone collapse. *J. Biol. Chem.* **291**, 26262 (2016).
109. Xie, Y. et al. Dendritic spine in autism genetics: Whole-exome sequencing identifying de novo variant of CTTNBP2 in a quad family affected by autism spectrum disorder. *Children* **10**, 80 (2023).
110. Shih, P.-Y. et al. Autism-linked mutations of CTTNBP2 reduce social interaction and impair dendritic spine formation via diverse mechanisms. *Acta Neuropathologica Commun.* **8**, 1–19 (2020).
111. Lin, Y.-C., Frei, J. A., Kilander, M. B. C., Shen, W. & Blatt, G. J. A subset of autism-associated genes regulate the structural stability of neurons. *Front. Cell. Neurosci.* **10**, 220700 (2016).
112. Shih, P.-Y. et al. CTTNBP2 controls synaptic expression of zinc-related autism-associated proteins and regulates synapse formation and autism-like behaviors. *Cell Rep.* **31**, 107700 (2020).
113. Shih, P.-Y. et al. Phase separation and zinc-induced transition modulate synaptic distribution and association of autism-linked CTTNBP2 and SHANK3. *Nat. Commun.* **13**, 1–20 (2022).
114. Sapir, T., Frotscher, M., Levy, T., Mandelkew, E. M. & Reiner, O. Tau's role in the developing brain: implications for intellectual disability. *Hum. Mol. Genet.* **21**, 1681–1692 (2012).
115. Washbourne, P. Synapse assembly and neurodevelopmental disorders. *Neuropsychopharmacology* **40**, 4–15 (2014).
116. Shaw-Smith, C. et al. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat. Genet.* **38**, 1032–7 (2006).
117. Tai, C. et al. Tau reduction prevents key features of autism in mouse models. *Neuron* **106**, 421 (2020).
118. Habas, R., Kato, Y. & He, X. Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel Formin homology protein Daam1. *Cell* **107**, 843–854 (2001).
119. Liu, W. et al. Mechanism of activation of the Formin protein Daam1. *Proc. Natl. Acad. Sci. USA.* **105**, 210–215 (2008).
120. Poliński, P. et al. A novel regulatory mechanism of actin cytoskeleton dynamics through a neural microexon in DAAM1 is important for proper memory formation. *bioRxiv* 2023.01.12.523772 <https://doi.org/10.1101/2023.01.12.523772>. (2023).
121. Telek, E., Kengyel, A. & Bugyi, B. Myosin XVI in the nervous system. *Cells* **9**, 1903 (2020).
122. Roesler, M. K. et al. Myosin XVI regulates actin cytoskeleton dynamics in dendritic spines of purkinje cells and affects pre-synaptic organization. *Front. Cell. Neurosci.* **13**, 453778 (2019).
123. Kao, C. F. et al. Identification of susceptible loci and enriched pathways for bipolar II disorder using genome-wide association studies. *Int. J. Neuropsychopharmacol.* **19**, pyw064 (2016).
124. Rodriguez-Murillo, L. et al. Fine mapping on chromosome 13q32–34 and brain expression analysis implicates MYO16 in Schizophrenia. *Neuropsychopharmacology* **39**, 934–943 (2013).
125. Liu, Y. F. et al. Autism and intellectual disability-associated KIRREL3 interacts with neuronal proteins MAPIB and MYO16 with potential roles in neurodevelopment. *PLoS One* **10**, e0123106 (2015).
126. Castañón, M. J., Walko, G., Winter, L. & Wiche, G. Plectin-intermediate filament partnership in skin, skeletal muscle, and peripheral nerve. *Histochem. Cell Biol.* **140**, 33–53 (2013).
127. Lie, A. A. et al. Plectin in the human central nervous system: pre-dominant expression at pia/glia and endothelia/glia interfaces. *Acta Neuropathol.* **96**, 215–221 (1998).
128. Potokar, M. & Jorgačevski, J. Plectin in the central nervous system and a putative role in brain astrocytes. *Cells* **10**, 2353 (2021).
129. Valencia, R. G. et al. Plectin dysfunction in neurons leads to tau accumulation on microtubules affecting neuriteogenesis, organelle trafficking, pain sensitivity and memory. *Neuropathol. Appl. Neurobiol.* **47**, 73 (2021).
130. Fasham, J. et al. SLC4A10 mutation causes a neurological disorder associated with impaired GABAergic transmission. *Brain* **146**, 4547–4561 (2023).
131. Kipper, F. C. et al. Embryonic periventricular endothelial cells demonstrate a unique pro-neurodevelopment and anti-inflammatory gene signature. *Sci. Rep.* **10**, 1–16 (2020).
132. Berg, J. M. & Geschwind, D. H. Autism genetics: searching for specificity and convergence. *Genome Biol.* **13**, 1–16 (2012).
133. Krepischi, A. C. et al. Two distinct regions in 2q24.2-q24.3 associated with idiopathic epilepsy. *Epilepsia* **51**, 2457–2460 (2010).

134. Luo, Y., Raible, D. & Raper, J. A. Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones. *Cell* **75**, 217–227 (1993).
135. Kolodkin, A. L., Matthes, D. J. & Goodman, C. S. The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules. *Cell* **75**, 1389–1399 (1993).
136. Williams, A. et al. Semaphorin 3A and 3F: key players in myelin repair in multiple sclerosis? *Brain* **130**, 2554–2565 (2007).
137. Syed, Y. A. et al. Inhibition of CNS remyelination by the presence of semaphorin 3A. *J. Neurosci.* **31**, 3719–3728 (2011).
138. Piaton, G. et al. Class 3 semaphorins influence oligodendrocyte precursor recruitment and remyelination in adult central nervous system. *Brain* **134**, 1156–1167 (2011).
139. Boyd, A., Zhang, H. & Williams, A. Insufficient OPC migration into demyelinated lesions is a cause of poor remyelination in MS and mouse models. *Acta Neuropathol.* **125**, 841–859 (2013).
140. Gutierrez, H., Dolcet, X., Tolcos, M. & Davies, A. HGF regulates the development of cortical pyramidal dendrites. *Development* **131**, 3717–3726 (2004).
141. Desole, C. et al. HGF and MET: From brain development to neurological disorders. *Front. Cell Dev. Biol.* **9**, 683609 (2021).
142. Levitt, P., Eagleson, K. L. & Powell, E. M. Regulation of neocortical interneuron development and the implications for neurodevelopmental disorders. *Trends Neurosci.* **27**, 400–406 (2004).
143. Campbell, D. B. et al. Disruption of cerebral cortex MET signaling in autism spectrum disorder. *Ann. Neurol.* **62**, 243–250 (2007).
144. Powell, E. M., Mars, W. M. & Levitt, P. Hepatocyte growth factor/scatter factor is a motogen for interneurons migrating from the ventral to dorsal telencephalon. *Neuron* **30**, 79–89 (2001).
145. Toriyama, M. et al. Shootin1: a protein involved in the organization of an asymmetric signal for neuronal polarization. *J. Cell Biol.* **175**, 147 (2006).
146. Baba, K. et al. Gradient-reading and mechano-effector machinery for netrin-1-induced axon guidance. *Elife* **7**, e34593 (2018).
147. Zeng, J. et al. Widespread signatures of natural selection across human complex traits and functional genomic categories. *Nat. Commun.* **12**, 1–12 (2021).
148. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
149. Lu, Q. et al. A powerful approach to estimating annotation-stratified genetic covariance via GWAS summary statistics. *Am. J. Hum. Genet.* **101**, 939–964 (2017).
150. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
151. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
152. Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
153. Kuhn, S., Gritti, L., Crooks, D. & Dombrowski, Y. Oligodendrocytes in development, myelin generation and beyond. *Cells* **8**, 1424 (2019).
154. Ziffra, R. S. et al. Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).
155. Matoba, N., Love, M. I. & Stein, J. L. Evaluating brain structure traits as endophenotypes using polygenicity and discoverability. *Hum. Brain Mapp.* **43**, 329 (2022).
156. Shepherd, G. M. G. & Yamawaki, N. Untangling the cortico-thalamo-cortical loop: Cellular pieces of a knotty circuit puzzle. *Nat. Rev. Neurosci.* **22**, 389–406 (2021).
157. Voineskos, A. N. et al. Age-related decline in white matter tract integrity and cognitive performance: a DTI tractography and structural equation modeling study. *Neurobiol. Aging* **33**, 21–34 (2012).
158. Farrar, D. C. et al. Retained executive abilities in mild cognitive impairment are associated with increased white matter network connectivity. *Eur. Radiol.* **28**, 340–347 (2018).
159. Wu, Y. et al. Pervasive biases in proxy GWAS based on parental history of Alzheimer’s disease. *bioRxiv* 2023.10.13.562272 <https://doi.org/10.1101/2023.10.13.562272>. (2023).
160. Sha, Z., Schijven, D., Fisher, S. E. & Francks, C. Genetic architecture of the white matter connectome of the human brain. *bioRxiv* 2022.05.10.491289 <https://doi.org/10.1101/2022.05.10.491289>. (2023)
161. van der Meer, D. et al. Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* **11**, 1–9 (2020).
162. Maier-Hein, K. H. et al. The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.* **8**, 1–13 (2017).
163. Bullock, D. N. et al. A Taxonomy of the Brain’s White Matter: Twenty-One Major Tracts for the 21st Century. *Cereb. Cortex* <https://doi.org/10.1093/cercor/bhab500>. (2022).
164. Seider, N. A. et al. Accuracy and reliability of diffusion imaging models. *Neuroimage* **254**, 119138 (2022).
165. Reid, L. B., Cespedes, M. I. & Pannek, K. How many streamlines are required for reliable probabilistic tractography? Solutions for microstructural measurements and neurosurgical planning. *Neuroimage* **211**, 116646 (2020).
166. Madole, J. W. et al. Aging-sensitive networks within the human structural connectome are implicated in late-life cognitive declines. *Biol. Psychiatry* **89**, 795–806 (2021).
167. van den Heuvel, M. P. & Sporns, O. Rich-club organization of the human connectome. *J. Neurosci.* **31**, 15775–15786 (2011).
168. Ball, G. et al. Rich-club organization of the newborn human brain. *Proc. Natl. Acad. Sci. USA* **111**, 7456–7461 (2014).
169. Arnatkeviciute, A. et al. Genetic influences on hub connectivity of the human connectome. *Nat. Commun.* **12**, 1–14 (2021).
170. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125**, 1063–1078 (2016).
171. Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
172. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774 (2012).
173. Tustison, N. J. et al. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
174. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5**, 143–156 (2001).
175. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
176. Smith, R. E., Tournier, J. D., Calamante, F. & Connelly, A. Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *Neuroimage* **62**, 1924–1938 (2012).
177. Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. The effects of SIFT on the reproducibility and biological accuracy of the structural connectome. *Neuroimage* **104**, 253–265 (2015).
178. Hagmann, P. et al. Mapping the Structural Core of Human Cerebral Cortex. *PLoS Biology* **6**, e159 (2008).
179. Backman, J. D. et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).

180. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
181. Mark Beasley, T., Erickson, S. & Allison, D. B. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav. Genet.* **39**, 580 (2009).
182. Clayton, D. Testing for association on the X chromosome. *Biostatistics* **9**, 593–600 (2008).
183. Wolfe, D., Dudek, S., Ritchie, M. D. & Pendergrass, S. A. Visualizing genomic information across chromosomes with PhenoGram. *BioData Min.* **6**, 18 (2013).
184. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
185. Mullins, N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
186. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
187. Demontis, D. et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat. Genet.* **55**, 198–208 (2023).
188. Zhou, H. et al. Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat. Neurosci.* **23**, 809 (2020).
189. Watanabe, K. et al. Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nat. Genet.* **54**, 1125–1132 (2022).
190. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer’s disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
191. Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson’s disease: a meta-genome wide association study. *Lancet Neurol.* **18**, 1091 (2019).
192. van Rheenen, W. 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).
193. International League Against Epilepsy Consortium on Complex EpilepsiesBerkovic, S. F., Cavalleri, G. L. & Koeleman, B. P. C. Genome-wide meta-analysis of over 29,000 people with epilepsy reveals 26 loci and subtype-specific genetic architecture. *medRxiv* 2022.06.08.22276120 (2022).
194. Mishra, A. et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature* <https://doi.org/10.1038/s41586-022-05165-3>. (2022).
195. Okbay, A. et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**, 437–449 (2022).
196. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912 (2018).
197. Davies, G. et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 1–16 (2018).
198. Grotzinger, A. D., de la Fuente, J., Nivard, M. G. & Tucker-Drob, E. M. Pervasive downward bias in estimates of liability scale heritability in GWAS meta-analysis: A simple solution. *medRxiv* 2021.09.22.21263909 (2021).
199. Privé, F., Arbel, J., Aschard, H. & Vilhjálmsson, B. J. Identifying and correcting for misspecifications in GWAS summary statistics and polygenic scores. *bioRxiv* 2021.03.29.437510 <https://doi.org/10.1101/2021.03.29.437510>. (2022).
200. Vallat, R. *Pinguin: statistics in python*. *J. Open Source Softw.* **3**, 1026 (2018).

Acknowledgements

We gratefully acknowledge Dan Felsky, Erin Dickie, Peter Zhukovsky, John Griffiths and Kelly Shen for helpful discussion, John Griffiths for creating the tractography image in Fig. 1, and Milos Milic for assistance with MRI data curation. MW is supported by a Banting Fellowship from the Canadian Institutes of Health Research. CH acknowledges the support of the CAMH Foundation, Canadian Institutes of Health Research, and National Institute of Mental Health. SJT acknowledges the support of the CAMH Discovery Fund, Krembil Foundation, Kavli Foundation, McLaughlin Foundation, Natural Sciences and Engineering Research Council of Canada (RGPIN-2020-05834 and DGEER-2020-00048), and Canadian Institutes of Health Research (NGN-171423 and PJT-175254) and the Simons Foundation Autism Research Initiative. This research was conducted under the auspices of UK Biobank application 61530, “Multimodal subtyping of mental illness across the adult lifespan through integration of multi-scale whole-person phenotypes”.

Author contributions

MW performed the GWAS and post-GWAS analyses. IK and MW researched candidate causal genes at each GWAS locus. SM developed the tractify tractography pipeline, and adapted it to the UK Biobank with assistance from NJF, CH and MW. SJT and CH supervised the study. SEM provided methodological input. MW wrote the manuscript with assistance from all co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-46023-2>.

Correspondence and requests for materials should be addressed to Michael Wainberg, Colin Hawco or Shreejoy J. Tripathy.

Peer review information : *Nature Communications* thanks Li Shen and Hongtu Zhu for their contribution to the peer review of this work.


Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Michael Wainberg ^{1,2,3,4} , **Natalie J. Forde**⁵, **Salim Mansour**⁶, **Isabel Kerrebijn**^{2,4}, **Sarah E. Medland** ^{7,8,9},
Colin Hawco^{3,4,6,11}  & **Shreejoy J. Tripathy** ^{1,3,4,10,11} 

¹Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada. ²Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, ON, Canada. ³Department of Psychiatry, University of Toronto, Toronto, ON, Canada. ⁴Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada. ⁵Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. ⁶Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada. ⁷QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia. ⁸School of Psychology, University of Queensland, Brisbane, QLD, Australia. ⁹Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia. ¹⁰Department of Physiology, University of Toronto, Toronto, ON, Canada. ¹¹These authors contributed equally: Colin Hawco, Shreejoy J. Tripathy.  e-mail: m.wainberg@utoronto.ca; colin.hawco@camh.ca; shreejoy.tripathy@camh.ca