



OPEN

Novel genetic variants associated with brain functional networks in 18,445 adults from the UK Biobank

Heidi Foo¹ , Anbupalam Thalamuthu¹, Jiyang Jiang¹, Forrest C. Koch¹, Karen A. Mather^{1,2}, Wei Wen¹ & Perminder S. Sachdev^{1,3}

Here, we investigated the genetics of weighted functional brain network graph theory measures from 18,445 participants of the UK Biobank (44–80 years). The eighteen measures studied showed low heritability (mean $h^2_{\text{SNP}} = 0.12$) and were highly genetically correlated. One genome-wide significant locus was associated with strength of somatomotor and limbic networks. These intergenic variants were located near the *PAX8* gene on chromosome 2. Gene-based analyses identified five significantly associated genes for five of the network measures, which have been implicated in sleep duration, neuronal differentiation/development, cancer, and susceptibility to neurodegenerative diseases. Further analysis found that somatomotor network strength was phenotypically associated with sleep duration and insomnia. Single nucleotide polymorphism (SNP) and gene level associations with functional network measures were identified, which may help uncover novel biological pathways relevant to human brain functional network integrity and related disorders that affect it.

During aging, the human brain undergoes functional changes which affect the integration of information within and between functional brain networks, and these have been shown to be associated with behavioral changes¹. By modelling large-scale brain networks using graph theory, defined by a collection of nodes (brain regions) and edges (magnitude of temporal correlation in functional magnetic resonance imaging (fMRI) activity between two brain regions)^{2,3}, it is possible to investigate aging-related topological changes. Previous functional graph theory studies have mainly considered edge presence or absence represented as a binary variable^{4–6}, which precludes information about variations in connectivity weights between networks. Given that connection weights exhibit high heterogeneity over several orders of magnitude⁷, studying weighted brain networks may provide greater insights into their underlying hierarchy and organizational principles.

Genetics may play an important role in influencing changes in the functional topology of the aging brain. Graph theoretical brain functional measures are reported to be heritable^{8,9}. While a previous study¹⁰ has looked at genetics of functional connectivity from resting-state fMRI images, to date, there has not been any population-based study investigating the genetic contribution to functional connectivity using graph theory measures. Studying the genetic architecture of graph theory brain functional measures has important implications¹¹—firstly, it can identify genes associated with network topology; and secondly, it may provide insights into the underlying biological mechanisms of macroscopic network topology in aging and how it alters during disease states.

Here, we address the question of how genetics is associated with the integrity of functional connectivity as measured by graph theory measures in resting-state fMRI (rs-fMRI) data. We assessed graph theory measures, including global efficiency, characteristic path length, Louvain modularity, transitivity, local efficiency and strength of default, dorsal attention, frontoparietal, limbic, salience, somatomotor, and visual networks, which are typically examined and found to change in aging⁶ as well as multiple neuropathological processes^{12–14}. A UK Biobank sample comprising 18,445 participants of British ancestry was used in this study. We first estimated single nucleotide polymorphism (SNP) heritability (h^2). Subsequently, genome-wide association studies (GWAS) were performed to identify genetic variants associated with each graph theory measure. Gene-based association

¹Centre for Healthy Brain Aging, CHeBA, School of Psychiatry, University of New South Wales Medicine, Kensington, Sydney, NSW 2052, Australia. ²Neuroscience Research Australia, Randwick, Sydney, NSW 2031, Australia. ³Neuropsychiatric Institute, Euroa Centre, Prince of Wales Hospital, Randwick, Sydney, NSW 2031, Australia. ✉email: heidi.foo@student.unsw.edu.au

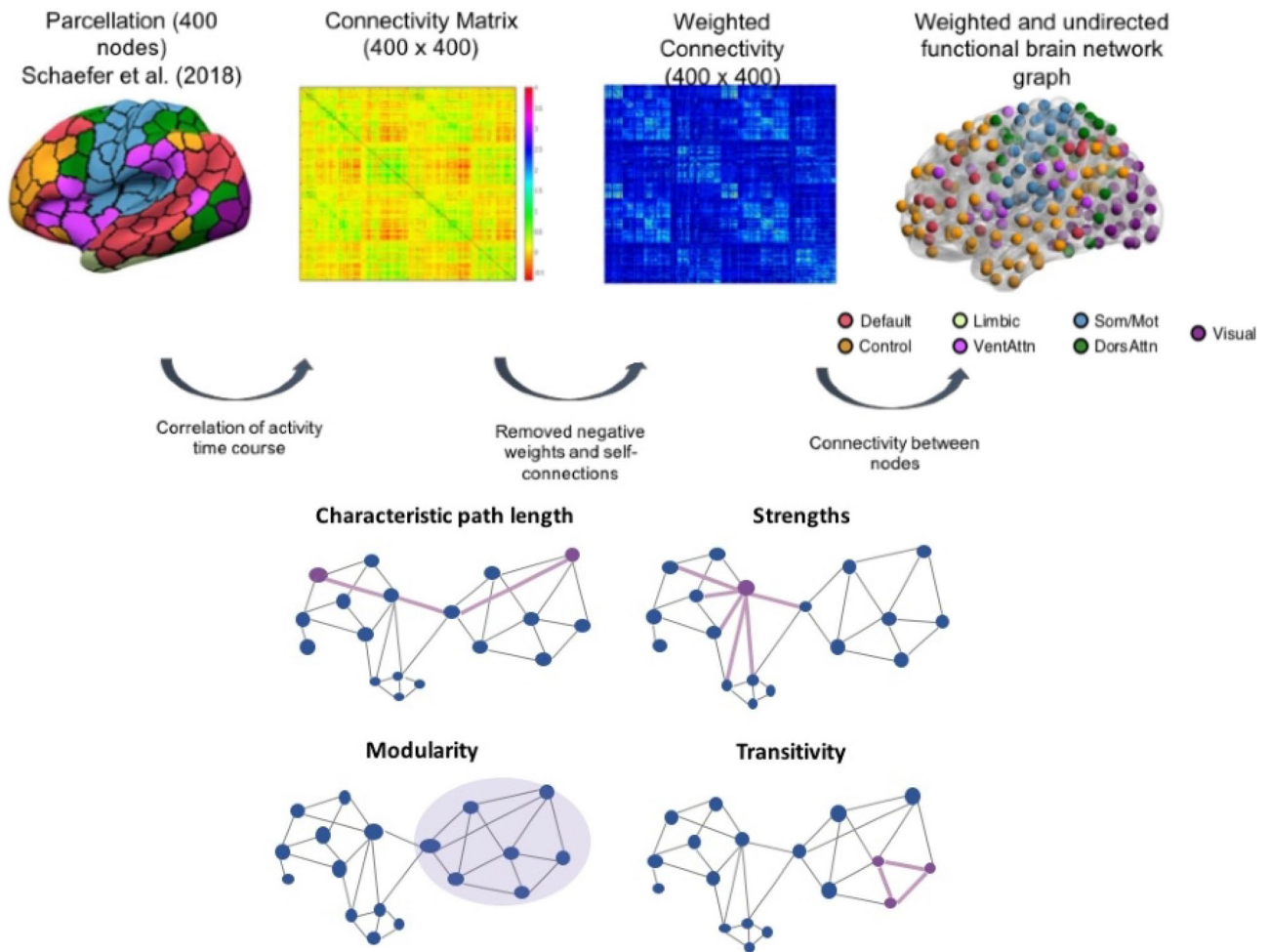


Figure 1. Schematic representation of brain network construction using graph theory analysis. After pre-processing, the brain was divided into different parcels using the Schaefer et al. (2018) parcellation scheme. Subsequently, activity time course was extracted from each region to create the correlation matrix. We used the correlation matrix and removed all the self-connected and negative weights to derive a corresponding weighted undirected brain network matrix and functional brain network graph. Lastly, we used the network matrix to calculate the sets of topological graph theory measures.

analysis was carried out to uncover gene-level associations, and the functional consequences of the significant genetic variants were explored.

Results

Demographics and graph theory measures. After imaging and genetic pre-processing and quality control, 2153 UK Biobank participants were excluded resulting in a final sample of 18,445 participants of British ancestry. There were 9773 females and 8672 males with a mean age (sd) of 62.47 (7.47). Eighteen weighted graph theory measures were derived by parcellating the rs-fMRI data using the Schaefer atlas into 400 regions¹⁵, which is a fine-grained parcellation scheme based on Yeo-7 network. These 18 measures include: global efficiency and characteristic path length (network integration); modularity, transitivity, and local efficiency of 7 networks (network segregation); and strength of 7 networks. Supplementary Table 1 defines each of the measures and provides evidence of their association with aging and neuropathological diseases. Supplementary Table 2 shows the mean and standard deviation of the demographics and graph theory measures. Figure 1 shows the brain network reconstruction using rs-fMRI data.

Phenotypic correlations between the measures were examined (Supplementary Fig. 1). High correlations were observed between the following measures: characteristic path length and modularity were negatively correlated with all other graph theory measures ($r = -0.661$ to -0.989); global efficiency and transitivity were positively correlated with local efficiency and strength of all the networks ($r = 0.753$ to 0.952); local efficiencies of networks were positively correlated with each other ($r = 0.784$ to 0.902); and strengths of networks were positively correlated with each other ($r = 0.772$ to 0.927).

SNP heritability estimates and genetic correlations. SNP heritability, h^2_{SNP} , was estimated using the proportion of variance in each graph theory measure that is explained by GWAS SNPs. All graph theory meas-

ures, except for strength of the visual network, were significantly heritable ($p < 0.05$), ranging from $h^2_{\text{SNP}} = 0.07$ for local efficiency of visual network to $h^2_{\text{SNP}} = 0.17$ for the strength of limbic network, with a mean h^2_{SNP} of 0.12 (Supplementary Table 3; Fig. 2A). Notably, higher h^2_{SNP} estimates (0.11–0.17) were observed for global efficiency, characteristic path length, transitivity, and strength of limbic, somatomotor, default, salience, and frontoparietal networks compared to the other measures. In addition, genetic and environmental correlations were examined. Strong genetic correlations between the network measures were observed. All measures were positively associated with each other with the exception of Louvain modularity and characteristic path length being negatively correlated with all other measures (Supplementary Table 4; Fig. 2B).

Genome-wide association study. GWAS of the rs-fMRI data for each individual graph theory measure ($n = 18$) were carried out using an additive genetic model adjusted for age, age², sex, age × sex, age² × sex, head motion from resting-state fMRI, head position, volumetric scaling factor needed to normalize for head size, genotyping array and 10 genetic principal components.

At the genome-wide significance level of $p < 5 \times 10^{-8}$ (unadjusted for the number of measures assessed), there were 31 SNPs significantly associated with nine of the 18 graph theory measures namely, global efficiency, characteristic path length, Louvain modularity, local efficiency of default and somatomotor networks, strength of default, limbic, salience, and somatomotor networks (Supplementary Table 5). Supplementary Fig. 2 shows the Manhattan and quantile–quantile plots of all the 18 graph theory measures. However, after adjusting for the number of independent tests using this method by Nyholt and colleagues¹⁶ ($n = 6$ tests for this study; p -threshold = $5 \times 10^{-8}/6 = 8.33 \times 10^{-9}$), only fourteen variants from a single locus remained significant with strength of the somatomotor network (lead SNP: rs12616641), and one of which also was significant for strength of the limbic network (rs62158161) (Table 1; Fig. 3A,B). All SNPs were located in an intergenic region near *PAX8* (Paired box gene 8) on chromosome 2 (BP 114065390–114092549) (Fig. 3C).

The conditional and joint association (COJO) analysis using the GWAS summary data of the network measures did not identify any additional SNPs apart from the top associated within each locus (Supplementary Table 5). There is high linkage disequilibrium in the genomic region (114065390–114110568) on chromosome 2 (Supplementary Fig. 3).

Multivariate association test. Since all network measures were highly correlated, to increase power a multivariate analysis using GWAS summary statistics from multiple network measures was performed. Based on their network properties, pooled summary statistics as implemented in metaUSAT¹⁷ were obtained for (i) global efficiency and characteristic path length; (ii) modularity and transitivity; (iii) local efficiency of all networks; and (iv) strength across all networks. Multivariate analysis of the combined strength network measures yielded 31 significant associations at $p < 1.25 \times 10^{-8}$ (adjusted for four tests). Out of this, 23 were found in GWAS of somatomotor strength measure reported in supplementary table 5 and the remaining 8 SNPs were found for combined strength measure (Table 2). No additional hits were found in the multivariate analysis of the other three grouped measures.

Gene-based association analysis. Gene-based association analysis was performed using the software MAGMA¹⁸ (Table 3). Our results showed that five genes – *SLC25A33* (Solute Carrier Family 25 Member 33), *TMEM201* (Transmembrane Protein 201), *ZEB1* (Zinc Finger E-Box Binding Homeobox 1), *SH2B3* (SH2B Adaptor Protein 3), and *ATXN2* (Ataxin 2) – were associated with global efficiency, characteristic path length, and strength of default, dorsal attention, and somatomotor networks after adjusting for the number of independent tests ($n = 6$) and number of genes ($n = 18,319$) i.e. p -threshold $< 0.05/(6 \times 18,319) = 4.56 \times 10^{-7}$. Supplementary Table 6 describes the genes associated with the graph theory measures.

Functional annotations. We assessed the potential functions of the 39 significant SNPs from the GWAS ($n = 31$) and multivariate association tests ($n = 8$) in SNPnexus¹⁹. The results showed that many of the network associated variants were associated with other phenotypes including sleep patterns, psychiatric disorders, coronary artery disease, cholesterol, and blood pressure (Supplementary Tables 7 and 8). None of the variants were found to be eQTLs.

Gene expression enrichment analysis of the list of significant genes in our results was subsequently performed using the Functional Mapping and Annotation (FUMA) platform²⁰. They were expressed in the brain but also across other tissue types (Supplementary Fig. 4). Moreover, there was no enrichment in brain tissue types (Supplementary Table 9).

Genetic correlations with other traits. Given that the most robust GWAS result was observed for the strength of somatomotor network, we examined its genetic correlations with other traits via linkage disequilibrium (LD) score regression (LDSC)²¹ (Supplementary Table 10). Results showed that strength of somatomotor network was genetically correlated with other traits, such as nervous feelings, sleep traits, neuroticism, depressive symptoms, blood pressure, education (unadjusted $p \leq 0.05$). However, none of them passed multiple comparisons correction.

Correlations with other associated traits. As the SNPs in our study have been associated with sleep and insomnia in previous GWAS studies (Supplementary Table 7), we explored the phenotypic correlations between the graph theory and sleep-related measures in the UK Biobank data (Supplementary Tables 11 and 12). Self-reported sleep duration and insomnia were significantly associated with the graph theory measures.

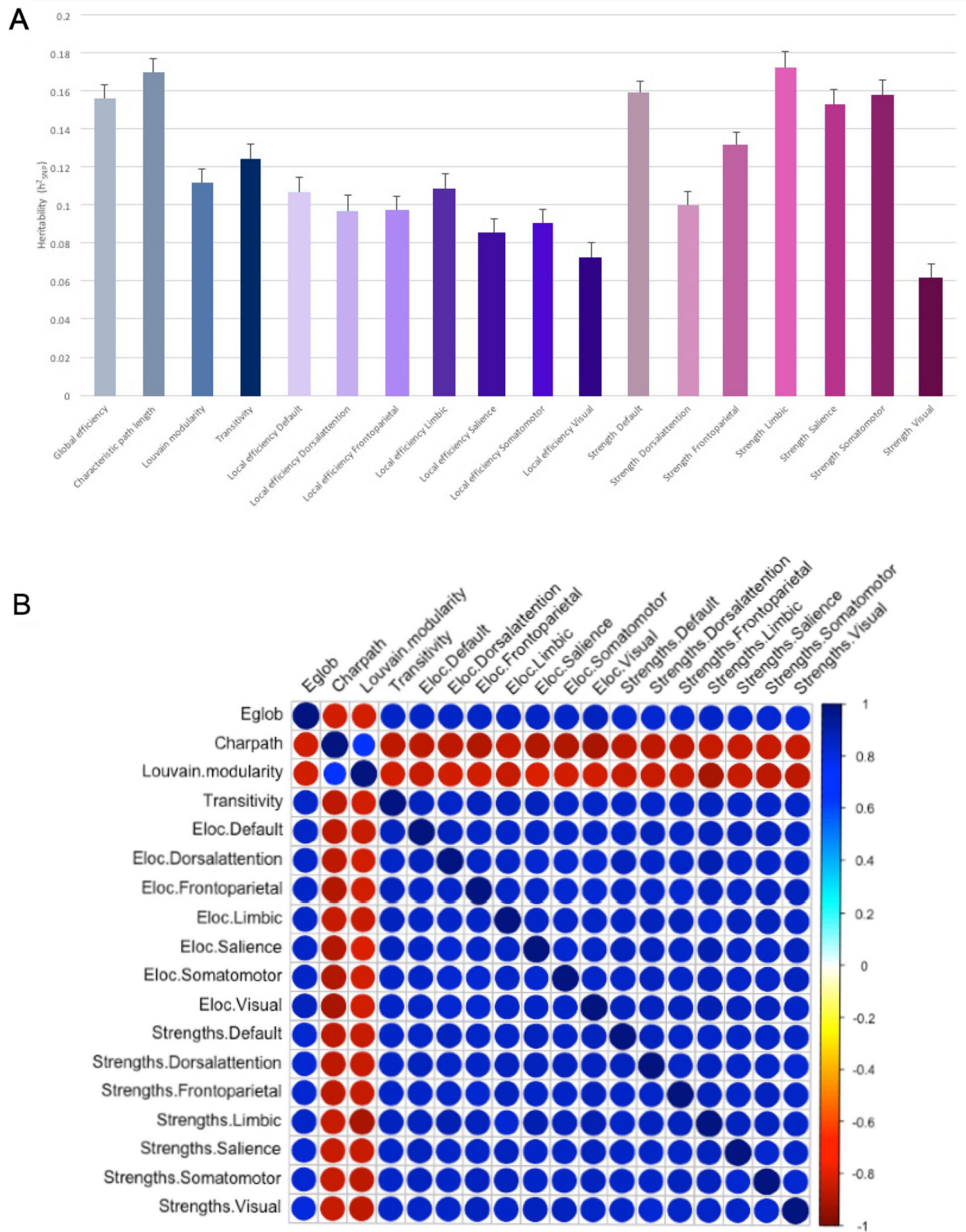


Figure 2. Genetic and environmental correlations between the 18 weighted graph theory measures. (A) represents the heritability estimates for each of the graph theory measures. Gray represents the network integration as characterized by global efficiency and characteristic path length; blue represents network segregation as characterized by Louvain modularity and transitivity; purple represents networks of local efficiency; and pink represents strength of the networks. (B) Genetic correlations were estimated using LDSC (<https://github.com/bulik/ldsc>). Strong genetic and environmental correlations between the network measures were observed (see Supplementary Table 4 for more details). Abbreviations: Eglob, global efficiency; Charpath, characteristic path length; Eloc, local efficiency.

SNP and measures	Chr	Position	Function	Nearest gene	A1/A2	EAF	β	SE	p
Strength Limbic									
rs62158161	2	114065572	Intergenic	PAX8	C/T	0.749	0.069	0.012	4.70E-09
Strength Somatomotor									
rs62158160	2	114065390	Intergenic	PAX8	C/T	0.742	0.066	0.011	6.90E-09
rs62158161	2	114065572	Intergenic	PAX8	C/T	0.749	0.069	0.012	2.40E-09
rs62158166	2	114077218	Intergenic	PAX8	G/C	0.773	0.073	0.012	1.20E-09
rs62158168	2	114078381	Intergenic	PAX8	C/G	0.773	0.073	0.012	1.20E-09
rs12616641*	2	114079248	Intergenic	PAX8	C/A	0.774	0.073	0.012	9.90E-10
rs62158169	2	114081827	Intergenic	PAX8	C/T	0.784	0.072	0.012	3.50E-09
rs62158170	2	114082175	Intergenic	PAX8	A/G	0.784	0.073	0.012	2.60E-09
rs199993536	2	114082628	Intergenic	PAX8	T/A	0.783	0.072	0.012	3.10E-09
rs6737318	2	114083120	Intergenic	PAX8	A/G	0.779	0.070	0.012	6.20E-09
rs62158206	2	114084596	Intergenic	PAX8	T/C	0.779	0.071	0.012	4.10E-09
rs7556815	2	114085785	Intergenic	PAX8	G/A	0.780	0.071	0.012	4.70E-09
rs2863957	2	114089551	Intergenic	PAX8	C/A	0.779	0.071	0.012	4.10E-09
rs1823125	2	114090412	Intergenic	PAX8	A/G	0.779	0.071	0.012	3.30E-09
rs60873293	2	114092549	Intergenic	PAX8	G/T	0.779	0.070	0.012	5.70E-09

Table 1. GWAS genome-wide significant results for brain functional network measures. Linear regression models were adjusted for age, age², sex, age \times sex, age² \times sex, head motion from resting-state fMRI, head position, volumetric scaling factor needed to normalize for head size, and 10 genetic principal components. P values are two-tailed. Only results that survived Bonferroni correction ($p < 5 \times 10^{-8}/6$) were reported here. *rs12616641 was the lead SNP. Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; A1, coded allele; A2, non-coded allele; EAF, effect allele frequency; β , beta; SE, standard error.

Strength of somatomotor network showed the most significant association with sleep duration ($p = 8.33 \times 10^{-11}$) and insomnia ($p = 0.0019$).

Discussion

Functional graph theory measures reflect the underlying functional topography of the brain. To date, this is the first study investigating the genetics of weighted functional graph theory measures. We present h^2_{SNP} estimates and results from GWAS of graph theory measures using resting-state fMRI data from 18,445 UK Biobank participants. We identified significant SNPs and gene associations that survived multiple correction testing with six of the 18 graph theory measures including global efficiency, characteristic path length, and strength of default, dorsal attention, limbic, and somatomotor networks. The novel contributions of this paper are the identification of new genetic associations at the variant, locus, and gene levels, providing insights into the genetic architecture of graph theory metrics using resting-state fMRI data.

Similar to the study by Elliot and colleagues¹⁰ who showed that resting-state fMRI connectivity edges had the lowest levels of h^2_{SNP} we found low h^2_{SNP} across all the graph theory measures. This is in contrast to previous classical twin study design studies that have shown moderate to high heritability of 0.52 to 0.64 for global efficiency, 0.47 to 0.61 for mean clustering coefficient, and 0.38 to 0.59 for modularity⁸. One of the plausible explanations is that h^2_{SNP} typically provides smaller heritability estimates due to uncaptured rare genetic variants compared to those provided by the classic twin study design¹⁰. High genetic correlations observed between graph theory measures may be due to the high phenotypic correlations between the measures.

The GWAS results found that strength of limbic and somatomotor networks were associated with SNPs located in an intergenic region near *PAX8* (Paired box gene 8) on chromosome 2, which is the closest gene to the top SNP. The *PAX8* gene family encodes transcription factors which are essential during development and tissue homeostasis²². Specifically, *PAX8* protein is considered as a master regulator for key cellular processes in DNA repair, replication, and metabolism²³. It has also been shown to regulate several genes involved in the production of thyroid hormone²⁴, essential for brain development and function such as neuronal differentiation, synaptogenesis, and dendritic proliferation^{25,26}. A previous study has also linked reductions in intrinsic functional connectivity in the somatomotor network to participants with subclinical hypothyroidism compared to controls²⁷. Interestingly, studies have also found subclinical thyroid dysfunction to be associated with sleep quality^{28,29}. Other genes located in this intergenic region near *PAX8* include *CBWD2* and *CHCHD5*, which have also been associated with sleep duration in prior studies^{30,31}. Considering that we also observed phenotypic correlations between sleep duration, insomnia and graph theory measures, it is possible that variants in or near *PAX8* and other genes in this region may play a role in the regulation of genes associated with functional brain network properties and sleep.

Results from multivariate SNP-based analyses from combined network strength measures found associations with SNPs on chromosome 2, which were associated with inflammation and oncogenesis. The majority of these SNPs were located in intergenic regions close to non-coding RNA genes. *LINC01826* (long intergenic non-protein coding RNA 1826) has been associated with inflammatory responses of the vascular endothelial cells, which

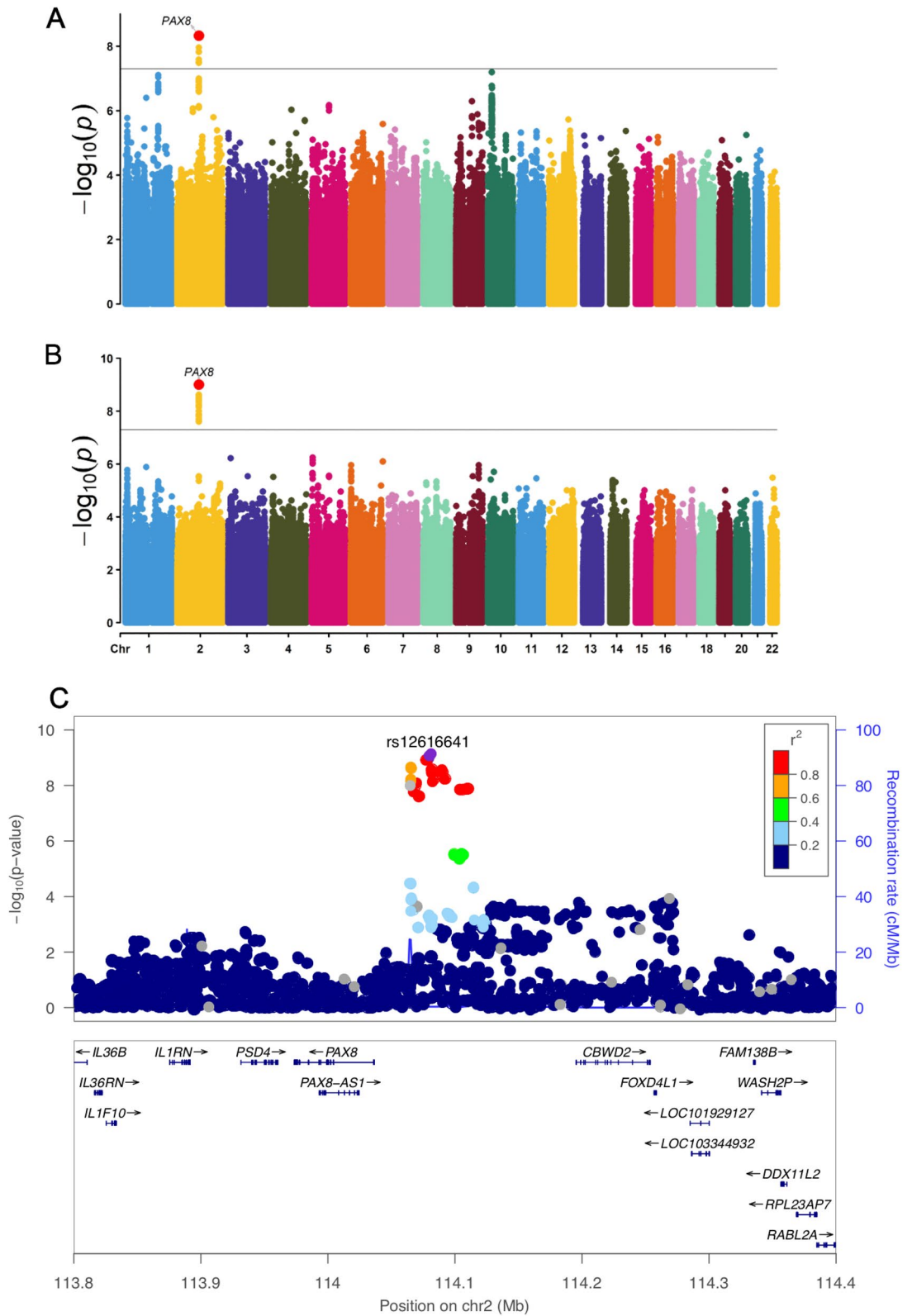


Figure 3. GWAS Manhattan Plots for strength of limbic and somatomotor networks and the locus zoom plot for the identified chromosome 2 region. (A) represents the Manhattan plot for strength of limbic network; (B) represents the Manhattan plot for strength of somatomotor network. For each of the Manhattan plots, each point represents a single genetic variant plotted according to its genomic position (x-axis) and its association with the relevant graph theory measure is shown by the corresponding $-\log_{10}(P)$ values on the y-axis. Linear regression models were adjusted for age, age², sex, age × sex, age² × sex, head motion from resting-state fMRI, head position, volumetric scaling factor needed to normalize for head size, genotyping array and 10 genetic principal components. The black solid line represents the classical GWAS significance threshold of $p < 5 \times 10^{-8}$. (C) The Locus zoom plot showing the chromosome 2 locus significantly associated with both the strength of somatomotor and limbic networks. Rs12616641 is the lead SNP.

Measure	SNP	Chr	Position	Function	Nearest gene	<i>p</i>
Combined network strength	rs145868127	2	144721440	ncRNA_exonic	<i>LOC101928386</i>	3.92E-18
	rs2680724	2	134822146	Intergenic	<i>MIR3679</i>	4.04E-18
	rs62165320	2	131541110	Intergenic	<i>AMER3</i>	5.08E-18
	rs2661030	2	123837766	Intergenic	<i>LINC01826</i>	5.33E-11
	rs2661035	2	123839945	Intergenic	<i>LINC01826</i>	5.45E-11
	rs1880544	2	121222694	ncRNA_exonic	<i>LINC01101</i>	1.35E-10
	rs2089478	2	123837764	Intergenic	<i>LINC01826</i>	1.56E-10
	rs12474078	2	211330092	ncRNA_intronic	<i>LANCL1-AS1</i>	2.50E-09

Table 2. Multivariate SNP-based analyses for combined network strength measure. Only results that survived Bonferroni correction ($p < 5 \times 10^{-8} / 4 = 1.25 \times 10^{-8}$) are reported. Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; ncRNA, non-coding RNA; *LOC101928386*, Uncharacterized LOC101928386; *MIR3679*, MicroRNA 3679; *AMER3*, APC Membrane Recruitment Protein 3; *LINC01826*, Long Intergenic Non-Protein Coding RNA 1826; *LINC01101*, Long Intergenic Non-Protein Coding RNA 1101; *LANCL1-AS1*, LANCL1 Antisense RNA 1.

Gene	Graph theory measures	Chr	Position	NSNPs	<i>P</i>
<i>SLC25A33</i>	Global efficiency	1	9599528	69	5.64E-08
	Characteristic path length	1	9599528	69	5.23E-08
<i>TMEM201</i>	Global efficiency	1	9648932	43	1.83E-08
	Characteristic path length	1	9648932	43	1.48E-08
	Strength dorsal attention	1	9648932	43	1.36E-07
	Strength somatomotor	1	9648932	43	2.09E-07
<i>ZEB1</i>	Global efficiency	10	31607825	323	3.27E-07
	Characteristic path length	10	31607825	323	2.29E-07
<i>SH2B3</i>	Strength default	12	111843720	50	2.95E-07
<i>ATXN2</i>	Strength default	12	111890018	149	2.48E-07

Table 3. Gene-based association analysis identified five significant gene-level associations for brain functional network measures. Gene-based association analysis was performed via MAGMA¹⁸, which uses 1000G reference panel for calculation of LD between the SNPs and gene coordinates based on NCBI build 37. Only genes that passed the Bonferroni correction ($p < 4.56 \times 10^{-7}$) are reported here. Abbreviations: Chr, chromosome; NSNP, number of single nucleotide polymorphisms. *SLC25A33*, Solute Carrier Family 25 Member 33; *TMEM201*, Transmembrane Protein 201; *ZEB1*, Zinc Finger E-Box Binding Homeobox 1; *SH2B3*, SH2B Adaptor Protein 3; *ATXN2*, Ataxin 2.

are important in the development of cardio-cerebrovascular diseases³². *MIR3679* (microRNA 3679) is a short non-coding RNA involved in post-transcriptional regulation of gene expression, which has been postulated to function as a tumor suppressor due to lower levels observed in patients with diffuse glioma than controls³³. *LINC01101* (long intergenic non-protein coding RNA 1101), on the other hand, has been down-regulated in cervical cancer³⁴. The identified genes may contribute to understanding the relationship between strength of brain networks and disease.

Gene-based association analysis showed associations with genes involved in neuronal differentiation/development, cancer, and susceptibility to neurodegenerative diseases. *ZEB1* has been implicated in neuronal glioblastoma³⁵ whereas *SLC25A33* has been associated with insulin/insulin-like growth factor 1 (IGF-1) necessary for metabolism, cell growth, and survival³⁶ and showed higher expression in transformed fibroblasts and cancer cell lines compared to non-transformed cells^{36,37}. *TMEM201* mice, which undergo accelerated senescence, exhibited an early onset age-related decline in antibody response and have a higher rate of mortality³⁸. *ATXN2* belongs to a class of genes associated with microsatellite-expansion diseases, where an interrupted CAG repeat expansion has been associated with brain-related diseases including spinocerebellar ataxia type 2 (SCA2), frontotemporal lobar degeneration (FTLD), and amyotrophic lateral sclerosis (ALS)^{39,40}. A neighbouring gene *SH2B3* to *ATXN2* has also been implicated in increased ALS risk⁴⁰. Consistent with previous studies that identified functional graph theory measures that were associated with the most common FTLD (behavioral variant of frontotemporal dementia)^{41,42}, we observed similar graph theory measures to be associated with the *ATXN2* gene. This implies that *ATXN2* may be involved in the relationship between the disruption of brain networks and neurological/neuromuscular disorders.

In addition, we also observed that brain functional networks are associated with self-reported sleep traits. Consistent with a previous study showing that individuals with chronic insomnia also showed disrupted global and local properties of the brain involving networks such as default mode, dorsal attention, and sensory-motor⁴³, we found that insomnia was most significantly associated with decreased strength of somatomotor network.

Therefore, it seems that changes in sleep quality may have a profound impact on the brain functional networks. Given that we observed genes associated with these networks, it is possible that altered networks may be driving the sleep abnormality. The reverse is also possible, with sleep disturbance influencing the activity in the networks. The directionality of the relationship between brain networks and sleep traits may be determined using longitudinal data in future studies.

The strengths of this study include the well characterized sample and its large sample size, and uniform MRI methods. The results, however, should be interpreted with caution. While using weighted undirected matrix circumvents issues surrounding filtering/thresholding the connectivity matrix to maintain significant edge weights represented in a binary matrix, there are inherent difficulties associated with the interpretation of the results. As brain signals recorded from resting-state fMRI are typically noisy, it is possible that edge weights may be affected by non-neural contributions⁴⁴. Despite this, with careful denoising of the resting-state fMRI data^{45,46} and covarying for motion, it is possible to minimize the noise in the data. In addition, previous studies have suggested that stronger edge weights make greater contributions in the computation of graph metrics than lower weight connections^{47,48}. This implies that when evaluating weighted graphs, false positive connections based on lower correlations may have a less disruptive impact on the network topology⁴⁹. Given that the brain is a complex system with hierarchical network structure, studying weighted networks, as was done in this study, may provide a more holistic representation of the brain functional network. Future studies may benefit from investigating the genetic effects between binarized and weighted graph theory metrics. Moreover, given the high correlations between the graph theory measures suggest that these measures may account for the same phenotype, or at least almost account for the same variability in the population, performing GWAS on each of these measures may not be necessary. However, these measures have shown to affect different disease states in older age. Moreover, this is an exploratory study and the findings presented are purely correlative. Despite so, the findings from the paper provide a new preliminary support for combining resting-state fMRI, graph theory methods, and GWAS to identify genetic variants associated with the various graph theory measures. It may also shed light to direct future studies in determining what graph theory imaging phenotypes should be included. In order to further understand the mechanisms about brain function at a more basic level, it may be useful for future studies to use developmental and adult human brain gene expression data to associate the expression of a single gene or genes with specific graph theory measures and phenotypes. Furthermore, given that the UK Biobank has recently released additional data, replication may be possible in the future.

In summary, this is the first study to investigate the genetics of weighted functional graph theory measures in a large and well characterized cohort. We observed multiple SNPs and genes associated with weighted graph theory measures, which have been observed to be implicated in sleep duration, neuronal differentiation/development, cancer, and susceptibility to neurodegenerative diseases. Our findings may help in the identification of novel biological pathways relevant to human brain functional network integrity and disease.

Methods

Study population. Our study was approved by the UK Biobank in December 2018 (Application number: 45262). rs-fMRI data for 20,598 participants with British ancestry was downloaded in March 2019⁵⁰. The imaging assessment took place at three different assessment centres with the majority assessed in Manchester, more recently scans were undertaken at Newcastle and Reading, UK. The UK Biobank study was conducted under approval from the NHS National Research Ethics Service (approval letter dated 17th June 2011, ref. 11/NW/0382), project 10279. All data and materials are available via UK Biobank (<http://www.ukbiobank.ac.uk>). Only individuals with both genetics and rs-fMRI data were included in this study.

Image processing. Briefly, the UK Biobank structural T1-weighted MRI scans were acquired on three 3 T Siemens Skyra MRI scanners (software platform VD13) at three sites (Reading, Newcastle, and Manchester) using a 32-channel receiving head coil and a 3D MPRAGE protocol (1.0 × 1.0 × 1.0 mm resolution, matrix 208 × 256 × 256, inversion time (TI)/repetition time (TR) = 880/2,000 ms, in-plane acceleration 2). An extensive overview of the data acquisition protocols and image processing carried out on behalf of the UK Biobank can be found elsewhere⁵¹. Rs-fMRI data previously pre-processed by the UK Biobank⁵¹ were used. Briefly, motion correction, intensity normalization, highpass temporal filtering, echo-planar imaging (EPI) unwarping, and gradient distortion correction were performed. Subsequently, structured artefacts were removed by ICA + FIX processing (Independent component analysis followed by FMRIB's ICA-based X-noiseifier)^{52–54}. We removed participants with motion of > 2 mm/degrees of translation/rotation. After image quality control and removal of participants with head motion outliers, we excluded 1,626 participants and 18,972 participants remained.

Graph theory analyses. The Schaefer atlas¹⁵ was used as it met the following requirements: (1) it integrated both local gradient and global similarity approaches (i.e. Markov Random Field model) for parcellation, which showed greater homogeneity than four other previously published parcellations implying that it will not overestimate local connectivity of the regions; (2) it revealed neurobiologically meaningful features of the brain organization; and (3) it was based of the Yeo 7-parcels atlas⁵⁵ and provided a more fine-grained parcellation, which allowed us to look at an average of the parcels across the 7 networks for local efficiency and nodal strength. 3dNetCorr command from Analysis of Functional Neuroimages (AFNI)⁵⁶ was used to produce network adjacency matrix for each participant. The mean time-series for each region was correlated with the mean time-series for all other regions and extracted for each participant. Partial correlation, r , between all pairs of signals was computed to form a 400-by-400 (Schaefer atlas) connectivity matrix, which was then Fisher z -transformed. Self-connections and negative correlations were set to zero. The main network analysis was performed on positive weighted networks. Given that connection weights in brain networks can vary across magnitude, undirected

weighted connectivity matrices were used instead of binary connectivity matrices. The higher the weight, the stronger the functional connectivity is between the brain regions⁴⁴.

All graph theory measures were quantified by using the brain connectivity toolbox (BCT)². Global-level measure included global efficiency, characteristic path length, transitivity, and Louvain modularity. A single value is derived for each of these four measures and quantitatively represent the whole brain network. Network level measures, such as local efficiency and strength, were estimated for each node and averaged across the all parcels within each network. Subsequently, we averaged the left–right hemisphere to derive a value for each node and averaged within each network to derive a value for each of the 7 networks.

To assess integration of information, we calculated global efficiency and characteristic path length^{2,57}. Global efficiency represents how effectively the information is transmitted at a global level and is the average inverse shortest path length while the latter measures the integrity of the network and how fast and easily information can flow within the network. To assess network segregation, which characterizes the specialized processing of the brain at a local level, we calculated the Louvain modularity, transitivity, and local efficiency indices^{2,57}. Louvain modularity is a community detection method, which iteratively transforms the network into a set of communities, each consisting of a group of nodes. Higher modularity values indicate denser within-modular connections but sparser connections between nodes that are in different modules. Transitivity refers to the sum of all the clustering coefficients around each node in the network and is normalized collectively. Local efficiency is a node-specific measure and is defined relative to the sub-graph comprising of the immediate neighbours of a node. Finally, strength (weighted degree) is described as the sum of all neighbouring edge weights². High connectivity strength indicates stronger connectivity between the regions, which provides an estimation of functional importance of each network.

Genotype data in UK Biobank. Genetic data for approximately 500,000 were available and full details on the genetics data used were described previously⁵⁸. The samples were collected from stored blood samples in the UK Biobank and genotyped either using the UK Biobank or the UK Biobank axiom array. Genotyping was performed on 33 batches of ~4700 samples by Affymetrix (High Wycombe, UK). Further details on the UK Biobank sample pre-processing are available here <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155583>.

An imputed data set was made available where the UK Biobank interim release was imputed to a reference set that consisted of >92 million autosomal variants imputed from the Haplotype Reference Consortium (HRC)⁵⁹ and UK10K + 1000 Genomes resources reference panels. After SNP QC filters (MAF > 0.1% and imputation information score > 0.3), 9926107 SNPs were used in the GWAS analysis.

Further quality control measures were performed. Due to the confounds associated with population structure⁶⁰, only samples reported to have recent British ancestry were used in the GWAS analysis. Outliers, including those with high missingness, relatedness, quality control failure, and sex mismatch, were removed. The final UK Biobank sample, after genotyping quality control and including those with rs-fMRI data, was n = 18,445 participants.

Sleep behavioral data in UK Biobank. Self-reported sleep data, namely sleep duration and frequency of insomnia, were used. Sleep duration was recorded as the number of reported hours of sleep in every 24 h and frequency of insomnia was recorded as never/rarely, sometimes, or usually. More details can be found on <http://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=1160> and <http://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=1200>.

Potential confounds. In accordance with previous GWAS of brain imaging phenotypes in the UK Biobank¹⁰, we controlled for similar confounding variables in this study. In addition to age at scanning and sex (i.e. age, age², age × sex, age² × sex), covariates relating to imaging parameters and genetic ancestry were also included. These included: head motion from resting-state fMRI, head position, volumetric scaling factor needed to normalize for head size, and the 10 genetic principal components.

SNP heritability. Heritability analysis, which is defined as the proportion of observed phenotypic variance explained by additive genetic factors of all common autosomal variants⁶¹, estimates the relative contribution of genes and the environment on a phenotype. Using BOLT-REML⁶² implemented in BOLT-LMM v2.3⁶³, we estimated the heritability accounted by autosomal SNPs among the graph theory measures. BOLT-REML uses multiple component modelling to partition SNP heritability and applies Monte Carlo algorithm for variance component analysis.

Genome-wide association analysis. BOLT-LMM v2.3⁶³ was used to conduct GWAS for the graph theory measures in the UK Biobank sample and adjusted for potential confounds. To correct for multiple hypothesis testing, we estimated the number of independent tests used based on Nyholt et al. method¹⁶ and derived n = 6 independent tests for this study. The genome-wide significant threshold was set at $p < (5 \times 10^{-8} / 6) = 8.33 \times 10^{-9}$. Findings with unadjusted threshold of $p < 5 \times 10^{-8}$ were also reported in Supplementary Table 5. Quantile and Manhattan plots were also presented for each of the graph theory measure in Supplementary Fig. 2. Manhattan and QQ plots were made using the R-package⁶⁴. Locus Zoom⁶⁵ was used for the visualization of the nearest genes within a ± 500-kilobase genomic region for the strength of somatomotor and limbic networks based on the hg19 UCSC Genome Browser assembly (Fig. 3).

Linkage disequilibrium and independent SNPs. The linkage disequilibrium plot of the associated genomic region was made using the LD plot function in the R package “gaston”⁶⁶. To identify independent SNPs associated with each of the network measures a stepwise model selection method as implemented in the COJO⁶⁷(cojo-slct) procedure of GCTA⁶⁸ package with default parameters was used.

Multivariate association analysis. Multivariate association analysis was conducted using metaUSAT¹⁷. The method uses summary statistics from individual studies and is suitable for correlated traits. MetaUSAT derives strength from two methods of multivariate association tests (score test and multivariate analysis variance test) as it uses convex linear combination of two test statistics. Data-driven minimum p-value corresponding to the best-linear combination is obtained. The significant threshold was set as $p < (5 \times 10^{-8} / 4) = 1.25 \times 10^{-8}$.

Gene-based association analysis and functional mapping. Gene-based association analysis was performed via MAGMA¹⁸ (v1.07, <https://ctg.cncr.nl/software/magma/>). MAGMA uses 1000G reference panel for calculation of LD between the SNPs and gene coordinates based on NCBI build 37. SNPs were mapped to a gene if they were within ± 5 kb of the gene co-ordinates. Gene-based association test statistic was derived using the default option, which is the sum of $-\log(\text{SNP } p\text{-value})$. The Bonferroni correction was used for significance of the gene-based association tests ($p\text{-value of the gene-based tests} / (\text{number of independent tests} \times \text{number of genes tested})$). In other words, significant threshold was set as $p < 0.05 / (6 \times 18,319) = 4.56 \times 10^{-7}$.

Functional annotation. We performed lookups for variants that passed the suggestive GWAS threshold of $p < 5 \times 10^{-8}$ to investigate the previously reported associations with the other traits. NHGRI-EBI GWAS Catalogue⁶⁹ included previous GWAS publications (Supplementary Table 7) and SNPnexus¹⁹ included all other publications (Supplementary Table 8).

FUMA²⁰ gene2func online platform (version 1.3.4, <http://fuma.ctglab.nl/>) was used to explore the functional consequences of significant genes. All the genes in the GWAS analysis and the gene-based analysis after nominal significance were used as input (a total of 18 genes). Using the FUMA platform, the GTEx v7 30 general tissue types data set was used for tissue specificity analyses. Differentially expressed gene (DEG) sets were pre-calculated by performing two-tailed t-test for any one of the tissue type against all others. Expression values were normalized (zero-mean) following a \log_2 transformation of expression values. Using the genes as background gene set, 2×2 enrichment sets were performed. Genes with $p\text{-value} \leq 0.05$ were after Bonferroni correction were defined as differentially expressed and colored in red in Supplementary Fig. 3.

Genetic correlation estimation with LDSC. LD Hub (v1.9.1, <http://ldsc.broadinstitute.org/ldhub/>) was used to estimate the genetic correlation between graph theory measures and corresponding traits. Summary statistics were uploaded to LD hub where it calculates the genetic correlations using the LDSC software (v1.0.0, <https://github.com/bulik/ldsc>).

Test of associations with graph theory measures. The graph theory measures were normalized using ranked transformed using the `rntransform()` function in R from GeneABEL package⁷⁰ and age were z-transformed for regression analysis. Regression model similar to GWAS analysis was used to test the association of self-reported sleep duration and frequency of insomnia variables with the graph theory measures. Only results with a two-tailed $p < 0.05/18$ (number of graph theory measures) were considered significant. Data management, derivation of summary statistics and other statistical analyses, and correlation plots were performed using R (V 4.0.0) software⁶⁴.

Received: 2 March 2021; Accepted: 7 July 2021

Published online: 16 July 2021

References

- Chan, M. Y., Park, D. C., Savalia, N. K., Petersen, S. E. & Wig, G. S. Decreased segregation of brain systems across the healthy adult lifespan. *Proc. Natl. Acad. Sci.* **111**, E4997. <https://doi.org/10.1073/pnas.1415122111> (2014).
- Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **52**, 1059–1069. <https://doi.org/10.1016/j.neuroimage.2009.10.003> (2010).
- Bertolero, M. A., Yeo, B. T. T., Bassett, D. S. & D’Esposito, M. A mechanistic model of connector hubs, modularity and cognition. *Nat. Hum. Behav.* **2**, 765–777. <https://doi.org/10.1038/s41562-018-0420-6> (2018).
- Geerligs, L., Renken, R. J., Saliassi, E., Maurits, N. M. & Lorist, M. M. A Brain-wide study of age-related changes in functional connectivity. *Cerebral Cortex (New York, N.Y.: 1991)* **25**, 1987–1999. <https://doi.org/10.1093/cercor/bhu012> (2015).
- Cohen, J. R. & D’Esposito, M. The segregation and integration of distinct brain networks and their relationship to cognition. *J. Neurosci.: Off. J. Soc. Neurosci.* **36**, 12083–12094. <https://doi.org/10.1523/jneurosci.2965-15.2016> (2016).
- Song, J. *et al.* Age-related reorganizational changes in modularity and functional connectivity of human brain networks. *Brain Connect.* **4**, 662–676. <https://doi.org/10.1089/brain.2014.0286> (2014).
- Markov, N. T. *et al.* The role of long-range connections on the specificity of the macaque interareal cortical network. *Proc. Natl. Acad. Sci.* **110**, 5187. <https://doi.org/10.1073/pnas.1218972110> (2013).
- Sinclair, B. *et al.* Heritability of the network architecture of intrinsic brain functional connectivity. *Neuroimage* **121**, 243–252. <https://doi.org/10.1016/j.neuroimage.2015.07.048> (2015).
- Fornito, A. *et al.* Genetic influences on cost-efficient organization of human cortical functional networks. *J. Neurosci.: Off. J. Soc. Neurosci.* **31**, 3261–3270. <https://doi.org/10.1523/jneurosci.4858-10.2011> (2011).

10. Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216. <https://doi.org/10.1038/s41586-018-0571-7> (2018).
11. Thompson, P. M., Ge, T., Glahn, D. C., Jahanshad, N. & Nichols, T. E. Genetics of the connectome. *Neuroimage* **80**, 475–488. <https://doi.org/10.1016/j.neuroimage.2013.05.013> (2013).
12. Munilla, J. *et al.* Construction and analysis of weighted brain networks from SICE for the Study of Alzheimer's Disease. *Front. Neuroinform.* <https://doi.org/10.3389/fninf.2017.00019> (2017).
13. Khazaei, A., Ebrahimi-Zadeh, A. & Babajani-Feremi, A. Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. *Clin. Neurophysiol.: Off. J. Int. Federat. Clin. Neurophysiol.* **126**, 2132–2141. <https://doi.org/10.1016/j.clinph.2015.02.060> (2015).
14. Lebedev, A. *et al.* Large-scale resting state network correlates of cognitive impairment in Parkinson's disease and related dopaminergic deficits. *Front. Syst. Neurosci.* <https://doi.org/10.3389/fnsys.2014.00045> (2014).
15. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex (New York, N.Y. : 1991)* **28**, 3095–3114. <https://doi.org/10.1093/cercor/bhx179> (2018).
16. Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74**, 765–769. <https://doi.org/10.1086/383251> (2004).
17. Ray, D. & Boehnke, M. Methods for meta-analysis of multiple traits using GWAS summary statistics. *Genet. Epidemiol.* **42**, 134–145. <https://doi.org/10.1002/gepi.22105> (2018).
18. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219. <https://doi.org/10.1371/journal.pcbi.1004219> (2015).
19. Dayem Ullah, A. Z. *et al.* SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res.* **46**, W109–W113. <https://doi.org/10.1093/nar/gky399> (2018).
20. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826. <https://doi.org/10.1038/s41467-017-01261-5> (2017).
21. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241. <https://doi.org/10.1038/ng.3406> (2015).
22. Blake, J. A. & Ziman, M. R. Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development* **141**, 737. <https://doi.org/10.1242/dev.091785> (2014).
23. Ruiz-Llorente, S. *et al.* Genome-wide analysis of Pax8 binding provides new insights into thyroid functions. *BMC Genomics* **13**, 147–147. <https://doi.org/10.1186/1471-2164-13-147> (2012).
24. Pasca di Magliano, M., Di Lauro, R. & Zannini, M. Pax8 has a key role in thyroid cell differentiation. *Proc. Natl. Acad. Sci. USA* **97**, 13144–13149. <https://doi.org/10.1073/pnas.240336397> (2000).
25. Williams, G. R. Neurodevelopmental and neurophysiological actions of thyroid hormone. *J. Neuroendocrinol.* **20**, 784–794. <https://doi.org/10.1111/j.1365-2826.2008.01733.x> (2008).
26. Bernal, J. Thyroid hormone receptors in brain development and function. *Nat. Clin. Pract. Endocrinol. Metab.* **3**, 249–259. <https://doi.org/10.1038/ncpendmet0424> (2007).
27. Kumar, M. *et al.* Alteration in intrinsic and extrinsic functional connectivity of resting state networks associated with subclinical hypothyroid. *J. Neuroendocrinol.* <https://doi.org/10.1111/jne.12587> (2018).
28. Kim, W. *et al.* Association between sleep duration and subclinical thyroid dysfunction based on nationally representative data. *J. Clin. Med.* <https://doi.org/10.3390/jcm8112010> (2019).
29. Song, L. *et al.* The Association Between Subclinical Hypothyroidism and Sleep Quality: A population-based study. *Risk Manag. Healthc. Policy* **12**, 369–374. <https://doi.org/10.2147/RMHP.S234552> (2019).
30. Veatch, O. J., Keenan, B. T., Gehrman, P. R., Malow, B. A. & Pack, A. I. Pleiotropic genetic effects influencing sleep and neurological disorders. *Lancet Neurol.* **16**, 158–170. [https://doi.org/10.1016/S1474-4422\(16\)30339-8](https://doi.org/10.1016/S1474-4422(16)30339-8) (2017).
31. Doherty, A. *et al.* GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat. Commun.* **9**, 5257. <https://doi.org/10.1038/s41467-018-07743-4> (2018).
32. Lin, Z. *et al.* Let-7e modulates the inflammatory response in vascular endothelial cells through ceRNA crosstalk. *Sci. Rep.* **7**, 42498. <https://doi.org/10.1038/srep42498> (2017).
33. Ohno, M. *et al.* Assessment of the diagnostic utility of serum MicroRNA classification in patients with diffuse glioma. *JAMA Netw. Open* **2**, e1916953–e1916953. <https://doi.org/10.1001/jamanetworkopen.2019.16953> (2019).
34. Iancu, I. V. *et al.* LINC01101 and LINC00277 expression levels as novel factors in HPV-induced cervical neoplasia. *J. Cell Mol. Med.* **21**, 3787–3794. <https://doi.org/10.1111/jcmm.13288> (2017).
35. Yu, W., Liang, S. & Zhang, C. Aberrant miRNAs regulate the biological hallmarks of glioblastoma. *NeuroMol. Med.* **20**, 452–474. <https://doi.org/10.1007/s12017-018-8507-9> (2018).
36. Favre, C., Zhdanov, A., Leahy, M., Papkovsky, D. & O'Connor, R. Mitochondrial pyrimidine nucleotide carrier (PNC1) regulates mitochondrial biogenesis and the invasive phenotype of cancer cells. *Oncogene* **29**, 3964–3976. <https://doi.org/10.1038/nc.2010.146> (2010).
37. Floyd, S. *et al.* The insulin-like growth factor-I-mTOR signaling pathway induces the mitochondrial pyrimidine nucleotide carrier to promote cell growth. *Mol. Biol. Cell* **18**, 3545–3555. <https://doi.org/10.1091/mbc.e06-12-1109> (2007).
38. Shimada, A. & Hasegawa-Ishii, S. Senescence-accelerated Mice (SAMs) as a model for brain aging and immunosenescence. *Aging Dis.* **2**, 414–435 (2011).
39. Fournier, C. *et al.* Interrupted CAG expansions in ATXN2 gene expand the genetic spectrum of frontotemporal dementias. *Acta Neuropathol. Commun.* **6**, 41. <https://doi.org/10.1186/s40478-018-0547-8> (2018).
40. Lahut, S. *et al.* ATXN2 and its neighbouring gene SH2B3 are associated with increased ALS risk in the Turkish population. *PLoS ONE* **7**, e42956–e42956. <https://doi.org/10.1371/journal.pone.0042956> (2012).
41. Reyes, P. *et al.* Functional connectivity changes in behavioral, semantic, and nonfluent variants of frontotemporal dementia. *Behav. Neurol.* **9684129–9684129**, 2018. <https://doi.org/10.1155/2018/9684129> (2018).
42. Agosta, F. *et al.* Brain network connectivity assessed using graph theory in frontotemporal dementia. *Neurology* **81**, 134–143. <https://doi.org/10.1212/WNL.0b013e31829a33f8> (2013).
43. Li, Z. *et al.* Disrupted brain network topology in chronic insomnia disorder: A resting-state fMRI study. *NeuroImage: Clin.* **18**, 178–185. <https://doi.org/10.1016/j.nicl.2018.01.012> (2018).
44. Fallani, F. D. V., Richiardi, J., Chavez, M. & Achard, S. Graph analysis of functional brain networks: practical issues in translational neuroscience. *Philos. Trans. R. Soc. B: Biol. Sci.* **369**, 20130521. <https://doi.org/10.1098/rstb.2013.0521> (2014).
45. Parkes, L., Fulcher, B., Yucel, M. & Fornito, A. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *Neuroimage* **171**, 415–436. <https://doi.org/10.1016/j.neuroimage.2017.12.073> (2018).
46. Power, J. D., Plitt, M., Laumann, T. O. & Martin, A. Sources and implications of whole-brain fMRI signals in humans. *Neuroimage* **146**, 609–625. <https://doi.org/10.1016/j.neuroimage.2016.09.038> (2017).
47. Drakesmith, M. *et al.* Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data. *Neuroimage* **118**, 313–333. <https://doi.org/10.1016/j.neuroimage.2015.05.011> (2015).
48. Ginestet, C. E., Fournel, A. P. & Simmons, A. Statistical network analysis for functional MRI: summary networks and group comparisons. *Front. Comput. Neurosci.* **8**, 51. <https://doi.org/10.3389/fncom.2014.00051> (2014).

49. van den Heuvel, M. P. *et al.* Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *Neuroimage* **152**, 437–449. <https://doi.org/10.1016/j.neuroimage.2017.02.005> (2017).
50. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779. <https://doi.org/10.1371/journal.pmed.1001779> (2015).
51. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400–424. <https://doi.org/10.1016/j.neuroimage.2017.10.034> (2018).
52. Beckmann, C. F. & Smith, S. M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* **23**, 137–152. <https://doi.org/10.1109/tmi.2003.822821> (2004).
53. Griffanti, L. *et al.* ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* **95**, 232–247. <https://doi.org/10.1016/j.neuroimage.2014.03.034> (2014).
54. Salimi-Khorshidi, G. *et al.* Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, 449–468. <https://doi.org/10.1016/j.neuroimage.2013.11.046> (2014).
55. Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165. <https://doi.org/10.1152/jn.00338.2011> (2011).
56. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res. Int. J.* **29**, 162–173 (1996).
57. Deco, G., Tononi, G., Boly, M. & Kringelbach, M. L. Rethinking segregation and integration: contributions of whole-brain modeling. *Nat. Rev. Neurosci.* **16**, 430–439. <https://doi.org/10.1038/nrn3963> (2015).
58. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209. <https://doi.org/10.1038/s41586-018-0579-z> (2018).
59. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283. <https://doi.org/10.1038/ng.3643> (2016).
60. Marchini, J., Cardon, L. R., Phillips, M. S. & Donnelly, P. The effects of human population structure on large genetic association studies. *Nat. Genet.* **36**, 512–517. <https://doi.org/10.1038/ng1337> (2004).
61. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304. <https://doi.org/10.1038/ng.3941> (2017).
62. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392. <https://doi.org/10.1038/ng.3431> (2015).
63. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908. <https://doi.org/10.1038/s41588-018-0144-6> (2018).
64. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2020).
65. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337. <https://doi.org/10.1093/bioinformatics/btq419> (2010).
66. Hervé Perdry, C. D.-R., Deepak Bandyopadhyay, Lutz Kettner. Gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models. (2020).
67. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375. <https://doi.org/10.1038/ng.2213> (2012).
68. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011> (2011).
69. 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. <https://doi.org/10.1093/nar/gky1120> (2019).
70. Karssen, L. C., van Duijn, C. M. & Aulchenko, Y. S. The GenABEL Project for statistical genomics. *F1000Res* **5**, 914–914. <https://doi.org/10.12688/f1000research.8733.1> (2016).

Author contributions

H.F. devised the study; H.F., and A.T. jointly designed the analyses, analyzed the data, interpreted and wrote the paper with technical support and guidance with data interpretation from K.M., W.W., and P.S.; H.F., A.T., J.Y.J., and F.C.K. contributed to the statistical concepts; and P.S. supervised the work. All authors commented on the manuscript.

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/s41598-021-94182-9>.

Correspondence and requests for materials should be addressed to H.F.

Reprints and permissions information is available at 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) 2021