

ARTICLE

Open Access

Expression quantitative trait loci-derived scores and white matter microstructure in UK Biobank: a novel approach to integrating genetics and neuroimaging

Miruna C. Barbu¹, Athina Spiliopoulou^{2,3}, Marco Colombo^{1,2}, Paul McKeigue², Toni-Kim Clarke¹, David M. Howard^{1,4}, Mark J. Adams¹, Xueyi Shen¹, Stephen M. Lawrie¹, Andrew M. McIntosh^{1,5} and Heather C. Whalley¹

Abstract

Expression quantitative trait loci (eQTL) are genetic variants associated with gene expression. Using genome-wide genotype data, it is now possible to impute gene expression using eQTL mapping efforts. This approach can be used to analyse previously unexplored relationships between gene expression and heritable in vivo measures of human brain structural connectivity. Using large-scale eQTL mapping studies, we computed 6457 gene expression scores (eQTL scores) using genome-wide genotype data in UK Biobank, where each score represents a genetic proxy measure of gene expression. These scores were then tested for associations with two diffusion tensor imaging measures, fractional anisotropy ($N_{FA} = 14,518$) and mean diffusivity ($N_{MD} = 14,485$), representing white matter structural integrity. We found FDR-corrected significant associations between 8 eQTL scores and structural connectivity phenotypes, including global and regional measures ($\beta_{\text{absolute}} FA = 0.0339-0.0453$; $MD = 0.0308-0.0381$) and individual tracts ($\beta_{\text{absolute}} FA = 0.0320-0.0561$; $MD = 0.0295-0.0480$). The loci within these eQTL scores have been reported to regulate expression of genes involved in various brain-related processes and disorders, such as neurite outgrowth and Parkinson's disease (*DCAKD*, *SLC35A4*, *SEC14L4*, *SRA1*, *NMT1*, *CPNE1*, *PLEKHM1*, *UBE3C*). Our findings indicate that eQTL scores are associated with measures of in vivo brain connectivity and provide novel information not previously found by conventional genome-wide association studies. Although the role of expression of these genes regarding white matter microstructural integrity is not yet clear, these results suggest it may be possible, in future, to map potential trait- and disease-associated eQTL to in vivo brain connectivity and better understand the mechanisms of psychiatric disorders and brain traits, and their associated imaging findings.

Introduction

Expression quantitative trait loci (eQTL) are genetic variants which are proximally (cis) or distally (trans) associated with variation in the expression of genes¹. Previous animal and human studies have found that

changes in gene expression lead to phenotypic variation, including adaptive phenotypic changes and evolutionary developments. In humans, for instance, cis-regulatory mutations lead to differences in lactase (*LCT*) gene expression, resulting in lactase persistence in adulthood². With respect to psychiatric disorders, major depressive disorder (MDD) and bipolar disorder have been associated with decreased expression of prodynorphin messenger RNA (mRNA), which is involved in regulation of mood and expressed in limbic-related areas within the brain (e.g., amygdala, hippocampus)³⁻⁵. These findings

Correspondence: Miruna C. Barbu (s1571976@sms.ed.ac.uk)

¹Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

²Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

Full list of author information is available at the end of the article.

© The Author(s) 2020



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

indicate the importance of cis-regulatory mutations to trait variance.

Variation in gene regulation leads to differences in individual phenotypes, indicating that eQTL may play a role in susceptibility to disease^{6,7}. To test this hypothesis, methods which combine gene expression data with genome-wide association studies (GWAS) summary statistics have been developed. These approaches may provide further insight into the potential causal pathways and genes involved in specific disorders, or predict the regulatory roles of single nucleotide polymorphisms (SNPs) in linkage disequilibrium (LD) with previously associated variants⁸. Previous studies have found that genetic variation may explain some of the variance in levels of gene expression in human tissues, including post-mortem brain tissue^{9–12}. In one such study, Zou et al.¹³ conducted an expression genome-wide association study (eGWAS) on post-mortem brains of individuals with Alzheimer's disease (AD) and other brain pathologies (non-AD; including progressive supranuclear palsy). They found 2980 *cis*SNPs associated with both AD and non-AD conditions. By investigating brain eQTL in post-mortem tissue therefore, researchers have been able to discover associations between gene expression and disease states in the brain.

Using brain tissue in order to investigate gene expression levels is however problematic, due to limitations such as small sample sizes and possible expression level differences in post-mortem versus ante-mortem brains¹⁴. As such, alternative approaches have therefore been investigated. One such approach is using eQTL measured from whole blood gene expression as a proxy for brain gene expression; an approach supported by important benefits such as greater sample size and easier accessibility¹⁵. Although it is recommended that wherever possible gene expression levels should be measured in a tissue-specific manner, considerable overlap has been demonstrated between blood and brain eQTL, supporting the validity of the approach¹⁴.

Neuroimaging measures provide a novel opportunity to investigate whether eQTL are significantly associated with in vivo brain phenotypes, and thereby increasing our knowledge of the role of eQTL in the wider context of psychiatric disorders. White matter microstructure, as measured by diffusion tensor imaging (DTI), is consistently heritable across tracts^{16–18} and is compromised in several psychiatric disorders. Generally, decreased microstructural integrity of white matter is characterised by lower directionality of water molecule diffusion (reduced fractional anisotropy, FA) and less constrained water molecule diffusion (increased mean diffusivity, MD). Consistent findings across studies have indicated higher MD and lower FA in individuals suffering from MDD, for example^{19,20}. Investigating the regulatory loci associated with white matter microstructure in health and

disease may aid in the detection of molecular mechanisms influencing disease through aberrant structural brain connectivity.

Within the current study, we derived eQTL scores based on two well-powered whole-blood eQTL studies^{21,22}. We then used GENOSCORES, a database of filtered summary statistics of publicly-available GWAS covering multiple phenotypes, including gene expression, to calculate eQTL scores (<https://pm2.phs.ed.ac.uk/genoscores/>).

The resultant eQTL-based genetic scores can be considered proxies for the expression of particular genes, which can then be tested for association with traits of interest. Here, we analysed their association with white matter microstructure as measured by FA and MD in UK Biobank. We used participants from the October 2018 UK Biobank neuroimaging release ($N_{FA} = 14,518$; $N_{MD} = 14,485$). The purpose of the study was to utilise a novel approach to investigate associations between regulatory SNPs and white matter microstructure. This approach could lead to further specialised investigation into psychiatric and neurological disorders, as well as other brain-related traits, such as cognition and behaviour.

Methods and materials

UK Biobank (UKB)

UK Biobank is a health resource aiming to prevent, diagnose and treat numerous disorders. It is comprised of 502,617 individuals whose genetic and environmental data (e.g., lifestyle, medications) were collected between 2006 and 2010 in the United Kingdom (<http://www.ukbiobank.ac.uk/>). UKB received ethical approval from the Research Ethics Committee (reference: 11/NW/0382). This study has been approved by the UKB Access Committee (Project #4844). Written informed consent was obtained from all participants.

Study population

In the current study, individuals were excluded if they participated in studies such as the Psychiatric Genomics Consortium (⁵¹; PGC) MDD GWAS or Generation Scotland (⁵²; Scottish Family Health Study), or if the individuals were biologically related to another participant, to remove overlap of genetic samples. For the brain imaging sample, a quality check performed by UK Biobank ensured that no abnormal scans were included in subsequent analyses²³. We additionally excluded individuals whose global measures for FA and MD lay more than three standard deviations from the sample mean^{19,24}. This resulted in 14,518 individuals with FA values ($N_{female} = 7561$ (52%); $N_{male} = 6957$ (48%); mean age: 63.14 ± 7.4 ; age range: 45.92–80.67) and 14,485 individuals with MD values ($N_{female} = 7552$ (52%); $N_{male} = 6933$ (48%); mean age: 63.12 ± 7.39 ; age range: 45.92–80.67).

Genotyping and eQTL score calculation

A total of 488,363 UKB blood samples ($N_{\text{female}} = 264,857$; $N_{\text{male}} = 223,506$; <http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22001>) were genotyped using the UK BiLEVE array ($N = 49,949$; <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=149600>) and the UK Biobank Axiom array ($N = 438,417$; <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=149601>). Details of genotyping and quality control are described in more detail by Hagenaaers et al.²⁵ and Bycroft et al.²⁶.

From GENOSCORES, we used eQTL analysis summary statistics from two studies of whole-blood eQTL^{21,22}. Briefly, Gusev et al.²¹ developed a novel approach aimed at identifying associations between gene expression and complex traits in cases where gene expression level is not directly measured. These authors reported eQTL based on a sample of 1414 individuals with whole-blood expression measured using the Illumina HumanHT-12 version 4 Expression BeadChip. Westra et al.²² performed a large eQTL meta-analysis in 5311 samples across 7 studies from peripheral blood, with gene expression measured using Illumina whole-genome Expression BeadChips (HT12v3, HT12v4 or H8v2 arrays). Their aim was to investigate the magnitude of the effect of cis and trans SNPs on gene expression, as well as to observe whether mapping eQTL in peripheral blood could uncover biological pathways associated with complex traits and disease. Further details of data acquisition and protocols are described in more detail in the two studies^{21,22}.

We computed a total of 10,884 eQTL scores (N Gusev study = 3801; N Westra study = 7083) for individuals included in the imaging sample ($N_{\text{FA}} = 14,518$; $N_{\text{MD}} = 14,485$) from the SNPs found in GENOSCORES, using a p -value threshold of 1×10^{-5} (0.00001). We then excluded overlapping eQTL scores between the two studies (i.e., scores for which SNPs affect expression of the same gene in both studies) by only including the score where a SNP had the lowest p -value, i.e., most significant association. The final eQTL score list was 6457 (N Gusev study = 3286; N Westra study = 3171). These scores were used as input variables in subsequent statistical analyses. Figure S1 in Supplementary Materials provides a summary of the score derivation process.

Briefly, eQTL scores were computed as a sum of the genotypes for an individual (g , scored as 0, 1, 2 copies of the reference allele) weighted by the effect size estimate (β_t) for the trait of interest t . In order to adjust for LD, vector β_t was pre-multiplied by the generalised inverse of the SNP-SNP correlation matrix R estimated from the 1000 Genomes reference panel, limited to the individuals with European ancestry.

The formula to compute the eQTL score for trait t for an individual (i) is therefore:

$$\text{score}(i, t) = g_i R^{-1} \beta_t$$

Magnetic resonance imaging (MRI) acquisition

In the current study, imaging-derived phenotypes (IDPs) produced by UKB were used. MRI acquisition and pre-processing procedures for white matter tracts were performed by UKB using standardised protocols (https://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf). Briefly, images were acquired in Manchester ($N_{\text{FA}} = 12,248$; $N_{\text{MD}} = 12,221$) and Newcastle ($N_{\text{FA}} = 2270$; $N_{\text{MD}} = 2264$) on a standard Siemens Skyra 3 T scanner with a 32-channel radio-frequency (RF) receive head coil and later pre-processed using the FMRIB Software Library (FSL), and parcellation of white matter tracts was conducted using AutoPtx²³. Individual white matter tracts belonging to each tract category can be observed in Table S1 in Supplementary Materials.

Owing to the fact that head position and RF coil in the scanner may affect data quality and subsequent pre-processing, three scanner brain position variables were also generated by UKB, with the aim of being used as confounding variables in subsequent analyses. These are lateral brain position—X (<http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25756>), transverse brain position—Y (<http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25757>) and longitudinal brain position—Z (<http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25758>). The three variables were included as covariates in the statistical analysis described below.

Statistical methods

All analyses were conducted using R (version 3.2.3) in a Linux environment. The R code for the current analyses is available in Supplementary Materials, section 3. For each white matter tract, we used generalized linear mixed models (function “lme” in package “nlme”) for bilateral brain regions, which were included as dependent variables. The eQTL scores were included as independent variables separately in each model, with additional covariates: age, age², sex, fifteen genetic principal components to control for population stratification, three MRI head position coordinates, MRI site and genotype array, while hemisphere was included as a within-subject variable. For unilateral tracts, as well as global measures and white matter tract categories of FA and MD, also included in the models as dependent variables, we used a general linear model (function “lm”), using the same covariates as above, without hemisphere included as a separate term, and again including the eQTL scores as independent variables separately in each model.

For global measures and white matter tract categories of FA and MD, we applied principal component analysis (PCA) on the white matter tracts of interest (all 27 for

Table 1 Information regarding eQTL scores with significant associations for both FA and MD-measured tracts.

Score name & eQTL type	N SNPs in score	Regulated gene	Study from which score is calculated	Gene function
DCAKD_eQTL_cis	8	Depospho-CoA Kinase Domain Containing (DCAKD)	Gusev et al.	Expressed in glioma; ubiquitous expression in brain; implicated in a number of psychiatric and neurological disorders ^{31–34}
SLC35A4_eQTL_cis	12	Solute Carrier Family 35 Member A4 (SLC35A4)	Gusev et al.	Expressed in brain ³⁶
SEC14L4_eQTL_cis	1	SEC14 Like Lipid Binding 4 (SEC14L4)	Westra et al.	Specific function not yet determined; may be implicated in neurodegeneration ⁴¹
SRA1_eQTL_cis	15	Steroid Receptor RNA Activator 1 (SRA1)	Westra et al.	Involved in regulation of many NR (nuclear receptor) and non-NR activities (e.g. chromatin organisation); may be associated with idiopathic hypogonadotropic hypogonadism ^{37,38}
NMT1_eQTL_cis	7	N-Myristoyltransferase 1 (NMT1)	Westra et al.	Ubiquitous expression in brain; may be implicated in brain tumours ^{47–49}
CPNE1_eQTL_cis	1	Copine 1 (CPNE1)	Westra et al.	May regulate molecular events at the interface of the cell membrane and cytoplasm; expressed during brain development and implicated in neurite outgrowth in rats ^{44–46}
PLEKHM1_eQTL_cis	5	Plectstrin Homology and RUN Domain Containing M1 (PLEKHM1)	Gusev et al.	Protein encoded by this gene is important for bone resorption; may play critical role in vesicular transport in the osteoclast ^{42,43}
UBE3C_eQTL_cis	4	Ubiquitin Protein Ligase E3C (UBE3C)	Westra et al.	Expressed in brain; may be implicated in Parkinson's disease ^{39,40}

global measures; 12 for association fibres; 6 for thalamic radiations; 9 for projection fibres) in order to extract a latent measure. Scores of the first unrotated component were extracted and set as dependent variables in general linear models. False discovery rate (FDR) correction using the “p.adjust” function in R ($q < 0.05$) was applied across the eQTL scores and the individual white matter tracts ($N_{\text{tests}} = 98,855$), and separately across eQTL scores and global and tract categories ($N_{\text{tests}} = 25,828$).

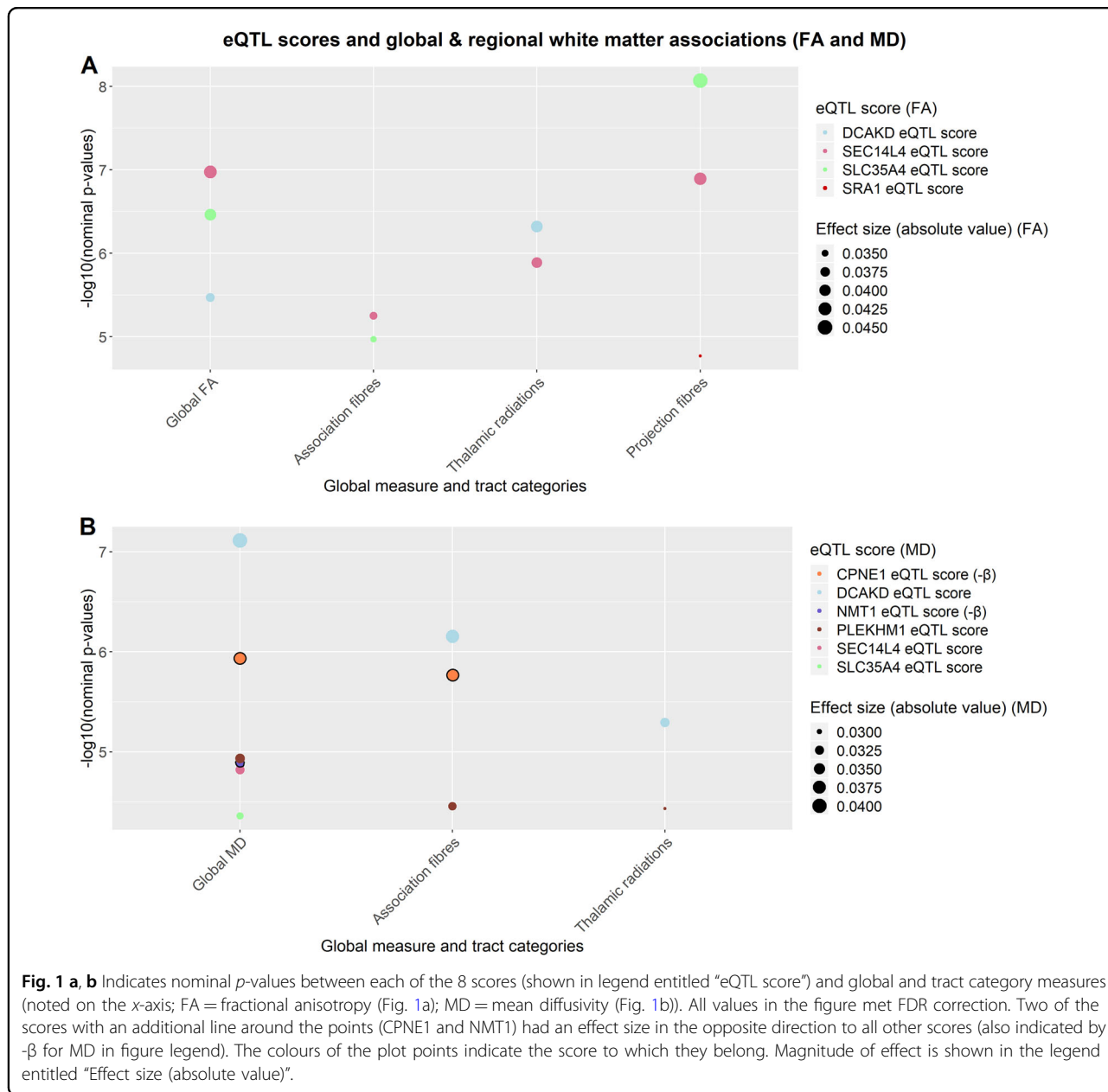
In addition, we included quadratic terms for each eQTL score in linear models to test non-linear relationships between the eQTL scores and global and regional measures of white matter microstructure. Analyses included all covariates mentioned above, with the only addition being the quadratic term of each eQTL score included alongside the original score. Results depicting the proportion of variance explained by the quadratic terms are included in Tables S2–S3 and Figs. S2–S3 in the Supplementary Materials.

Finally, for the top significant findings in the manuscript, we ran linear models where there were overlapping genes between the two discovery datasets^{21,22} (i.e. eQTL scores that regulate the expression of the same gene in both Gusev et al. and Westra et al.), and report results from both datasets for comparison (Tables S4–S12 in Supplementary Materials).

Results

There were several eQTL scores that showed significant associations with a number of global measures, tract categories, and white matter tracts post FDR correction ($q < 0.05$); Table 1; Figs. 1 (a, b) and 2 (a, b); Tables S13–S20 in Supplementary Materials). In total, 25 scores were significantly associated with FA values ($\beta_{\text{absolute}} = 0.032–0.056$) and 24 scores with MD values ($\beta_{\text{absolute}} = 0.029–0.048$) in several tracts (see Tables S21–S22 and Figs. S4–S5 in Supplementary Materials). Among these scores, 8 were associated with white matter tracts measured by both FA and MD. The primary findings reported in this manuscript focus on these 8 overlapping scores, as these were considered to provide the most consistent information with regards to gene expression within white matter tracts as measured by two different DTI scalars (see Tables 2 and 3), further findings are presented in Tables S23–S24 in Supplementary Materials.

The Allen Brain Atlas²⁷ was used to investigate gene expression patterns across brain regions in 6 donors. Figs. S6–S8 and Table S25 in Supplementary Materials provide a summary of each donor's individual gene expression level as well as a mean gene expression level across participants in a number of brain structures.



The effect of the 8 scores on FA measures of white matter microstructure

Table 2.

The effect of the 8 scores on MD measures of white matter microstructure

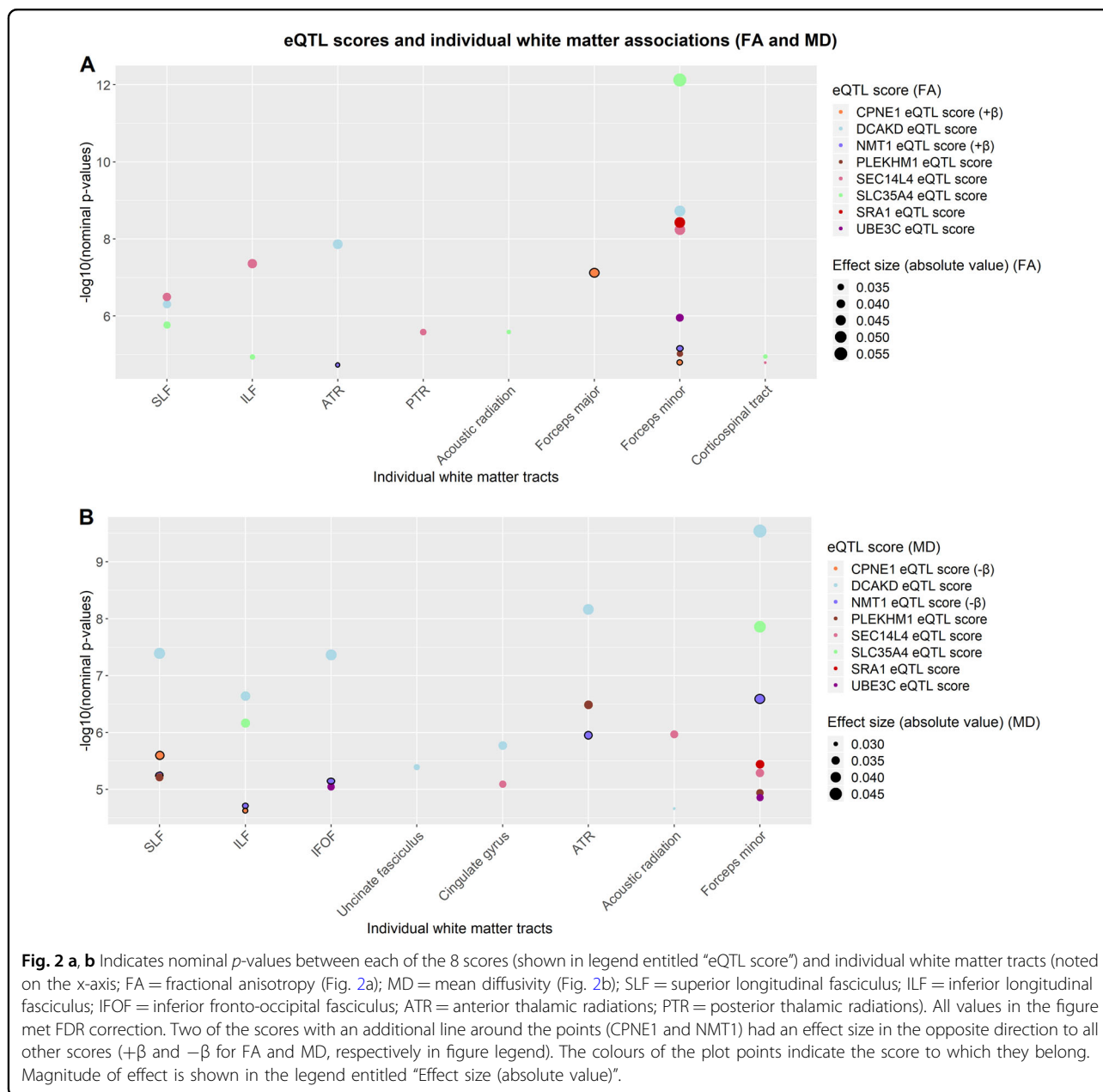
Table 3.

Genome-wide associations between score SNPs and white matter tracts

Using a previously published GWAS of imaging traits²⁸, we next sought to observe the association between the SNPs comprising each of the 8 scores ($N_{total} = 53$; SNP list

can be found in Table S26 in Supplementary Materials) with those found previously for the white matter tracts of interest (i.e., the tracts which showed post-FDR significant associations). This SNP look-up was performed in order to observe whether our analysis of eQTL scores, comprising SNPs which together regulate the expression of a single gene, yielded any novel associations with white matter tracts which were not previously found in conventional GWAS.

We used the Brain Imaging Genetics (BIG) database (<http://big.stats.ox.ac.uk/>) to extract the effect size and p -value of each SNP of interest as associated with the white matter tracts, as provided in Elliott et al.²⁸. As GWAS for



global and tract category measures were not performed in the original study²⁸, we performed these GWAS as part of the current project (i.e., GWAS for global measures, association fibres, thalamic radiations and projection fibres). Our GWAS parameters and quality check procedures are described in more detail in Supplementary Materials, section 11. p -Values and effect size of each SNP for each individual white matter tract of interest (left and right hemispheres separately from Elliott et al.²⁸), as well as for global and tract categories (run locally), are also contained in Figure S9 in Supplementary Materials. Briefly, only one SNP across three eQTL scores was previously found to reach genome-wide significance with

forceps minor (FA)²⁸, projection fibres (FA) and global FA (GWAS run locally): rs2237077.

Discussion

In this study, we utilised a novel approach to investigate whether eQTL scores corresponding to the expression of specific genes in whole blood were significantly and specifically associated with white matter tracts in $N > 14,000$ individuals. We found significant associations in white matter microstructure as measured by both FA and MD for a number of scores (FA_{N scores} = 25; MD_{N scores} = 24). Of these, 8 scores were found to be significantly associated with various white matter tracts as measured by

Table 2 Significant associations between eQTL scores and global measures, category, and individual white matter tracts (FA); FDR = false discovery rate; for each score, tracts are arranged from global to individual tracts.

Score, white matter tracts	Effect size	SD	t-value	p-value	p-value, FDR corrected	R ² (%)
DCAKD eQTL score						
Global FA	-0.037	0.008	-4.647	1.37×10^{-8}	0.009	0.132
Thalamic radiations	-0.040	0.008	-5.038	1.88×10^{-9}	0.003	0.159
Superior longitudinal fasciculus	-0.039	0.008	-5.033	4.89×10^{-7}	0.004	0
Anterior thalamic radiations	-0.043	0.008	-5.680	3.45×10^{-7}	<0.001	0
Forceps minor	-0.047	0.008	-6.012	1.07×10^{-5}	<0.001	0.221
SLC35A4 eQTL score						
Global FA	-0.040	0.008	-5.100	1.12×10^{-5}	0.002	0.156
Association fibres	-0.035	0.008	-4.404	2.57×10^{-6}	0.020	0.114
Projection fibres	-0.045	0.008	5.761	1.15×10^{-5}	<0.001	0.201
Corticospinal tract	-0.033	0.007	-4.395	1.70×10^{-6}	0.034	0
Acoustic radiation	-0.033	0.007	-4.704	7.56×10^{-13}	0.013	0
Inferior longitudinal fasciculus	-0.034	0.008	-4.389	1.15×10^{-5}	0.034	0
Superior longitudinal fasciculus	-0.037	0.008	-4.789	1.05×10^{-7}	0.010	0
Forceps minor	-0.056	0.008	-7.175	5.60×10^{-6}	<0.001	0.312
SEC14L4 eQTL score						
Global FA	-0.042	0.008	-5.320	1.28×10^{-7}	0.001	0.167
Association fibres	-0.036	0.008	-4.543	1.63×10^{-5}	0.012	0.118
Thalamic radiations	-0.039	0.008	-4.843	2.61×10^{-6}	0.005	0.141
Projection fibres	-0.042	0.008	5.285	3.19×10^{-7}	0.001	0.167
Corticospinal tract	-0.032	0.007	-4.312	4.39×10^{-8}	0.042	0.842
Posterior thalamic radiation	-0.035	0.008	-4.701	5.76×10^{-9}	0.013	0.919
Superior longitudinal fasciculus	-0.039	0.008	-5.114	3.19×10^{-7}	0.003	0.672
Inferior longitudinal fasciculus	-0.042	0.008	-5.477	1.69×10^{-5}	0.001	0.842
Forceps minor	-0.046	0.008	-5.827	3.76×10^{-9}	<0.001	0.200
SRA1 eQTL score						
Projection fibres	-0.034	0.008	4.303	1.83×10^{-5}	0.027	0.109
Forceps minor	-0.046	0.008	-5.898	6.99×10^{-6}	<0.001	0.207
NMT1 eQTL score						
Anterior thalamic radiations	0.032	0.008	4.286	1.58×10^{-5}	0.043	0.162
Forceps minor	0.035	0.008	4.496	7.49×10^{-8}	0.027	0.123
CPNE1 eQTL score						
Forceps minor	0.034	0.008	4.319	9.40×10^{-6}	0.042	0.108
Forceps major	0.044	0.008	5.382	7.49×10^{-8}	0.001	0.181
PLEKHM1 eQTL score						
Forceps minor	-0.035	0.008	-4.432	1.37×10^{-8}	0.034	0.119
UBE3C eQTL score						
Forceps minor	-0.038	0.008	-4.872	1.12×10^{-6}	0.008	0.141

R² represents variance explained (%) by each score for each tract.

Table 3 Significant associations between eQTL scores and individual white matter tracts (MD); FDR = false discovery rate; for each score, tracts are arranged from global to individual tracts.

Score, white matter tracts	Effect size	SD	t-value	p-value	p-value, FDR corrected	R ² (%)
DCAKD eQTL score						
Global MD	0.040	0.008	5.376	7.72 × 10 ⁻⁸	0.002	0.166
Thalamic radiations	0.033	0.007	4.563	5.10 × 10 ⁻⁶	0.013	0.109
Association fibres	0.038	0.008	4.964	6.97 × 10 ⁻⁷	0.005	0.145
Acoustic radiation	0.030	0.007	4.247	2.18 × 10 ⁻⁵	0.047	0
Uncinate fasciculus	0.031	0.007	4.609	4.09 × 10 ⁻⁶	0.016	0
Cingulate gyrus	0.035	0.007	4.789	1.70 × 10 ⁻⁶	0.009	0
Inferior longitudinal fasciculus	0.038	0.007	5.177	2.29 × 10 ⁻⁷	0.002	0
Anterior thalamic radiations	0.040	0.007	5.796	6.92 × 10 ⁻⁹	<0.001	0
Inferior fronto-occipital fasciculus	0.041	0.008	5.481	4.31 × 10 ⁻⁸	0.001	0
Superior longitudinal fasciculus	0.042	0.008	5.490	4.08 × 10 ⁻⁸	0.001	0
Forceps minor	0.048	0.008	6.309	2.90 × 10 ⁻¹⁰	<0.001	0.232
SLC35A4 eQTL score						
Global MD	0.031	0.008	4.089	4.35 × 10 ⁻⁵	0.042	0.088
Inferior longitudinal fasciculus	0.036	0.007	4.968	6.86 × 10 ⁻⁷	0.004	0
Forceps minor	0.043	0.008	5.677	1.39 × 10 ⁻⁸	<0.001	0.181
SEC14L4 eQTL score						
Global MD	0.033	0.008	4.330	1.50 × 10 ⁻⁵	0.028	0.094
Cingulate gyrus	0.033	0.007	4.465	8.07 × 10 ⁻⁶	0.025	0.593
Acoustic radiation	0.034	0.007	4.878	1.08 × 10 ⁻⁶	0.006	2.044
Forceps minor	0.035	0.008	4.560	5.15 × 10 ⁻⁶	0.019	0.110
SRA1 eQTL score						
Forceps minor	0.035	0.008	4.635	3.60 × 10 ⁻⁶	0.016	0.116
NMT1 eQTL score						
Global MD	-0.033	0.008	-4.363	1.29 × 10 ⁻⁵	0.026	0.112
Inferior longitudinal fasciculus	-0.031	0.007	-4.270	1.97 × 10 ⁻⁵	0.045	0.137
Inferior fronto-occipital fasciculus	-0.034	0.008	-4.485	7.37 × 10 ⁻⁶	0.023	0.153
Anterior thalamic radiations	-0.034	0.007	-4.870	1.13 × 10 ⁻⁶	0.006	0.110
Superior longitudinal fasciculus	-0.034	0.008	-4.536	5.79 × 10 ⁻⁶	0.020	0.088
Forceps minor	-0.039	0.008	-5.154	2.59 × 10 ⁻⁷	0.003	0.157
CPNE1 eQTL score						
Global MD	-0.037	0.008	-4.865	1.16 × 10 ⁻⁶	0.005	0.125
Association fibres	-0.037	0.008	-4.787	1.71 × 10 ⁻⁶	0.006	0.126
Inferior longitudinal fasciculus	-0.031	0.007	-4.230	2.35 × 10 ⁻⁵	0.050	0.249
Superior longitudinal fasciculus	-0.036	0.008	-4.706	2.56 × 10 ⁻⁶	0.012	0.180
PLEKHM1 eQTL score						
Global MD	0.033	0.008	4.386	1.16 × 10 ⁻⁵	0.025	0.112
Thalamic radiations	0.030	0.007	4.128	3.68 × 10 ⁻⁵	0.042	0.091
Association fibres	0.032	0.008	4.140	3.50 × 10 ⁻⁵	0.042	0.102

Table 3 continued

Score, white matter tracts	Effect size	SD	t-value	p-value	p-value, FDR corrected	R ² (%)
Forceps minor	0.033	0.008	4.388	1.15 × 10 ⁻⁵	0.030	0.112
Superior longitudinal fasciculus	0.034	0.008	4.522	6.17 × 10 ⁻⁶	0.021	0
Anterior thalamic radiations	0.036	0.007	5.110	3.26 × 10 ⁻⁷	0.003	0
UBE3C eQTL score						
Forceps minor	0.033	0.008	4.347	1.39 × 10 ⁻⁵	0.035	0.104
Inferior fronto-occipital fasciculus	0.033	0.008	4.441	9.01 × 10 ⁻⁶	0.027	0

R² represents variance explained (%) by each score for each tract.

both FA and MD. In particular, the largest effect was seen for the association between forceps minor (FA) and the eQTL score for *SLC35A4*, and across several tracts measured by MD for the eQTL score for *DCAKD*. Although these eQTL were derived from whole blood, there is evidence of expression in the brain for some of the genes, outlined in further detail below. These findings also provided novel information not previously found by conventional genome-wide association studies.

All 8 scores were associated with white matter microstructural integrity of the forceps minor as measured by FA (7 of which were also associated with MD values). The forceps minor forms the anterior part of the corpus callosum, connecting homologous regions of the prefrontal cortex between hemispheres. It is postulated to be involved in numerous cognitive and behavioural skills, such as decision making, social behaviour, and language²⁹. This connection therefore implicates forceps minor in a wide range of cognitive skills, and damage to the tract has been associated with neuropsychiatric and neurological disorders, such as multiple sclerosis and depression^{30,31}.

Global and individual tract findings—largest associations

The two genes with the largest associations were *DCAKD*, globally and across numerous tracts as measured by higher MD, and *SCL35A4* across tracts measured by lower FA, with a peak in projection fibres, localised to forceps minor. *DCAKD* is a protein coding gene which is ubiquitously expressed in brain, among other tissues³². Previous evidence using mouse models indicates expression of this gene has a putative role in neurodevelopment³³, and is associated with a number of psychiatric and neurological disorders, including schizophrenia, autism spectrum disorder, and Parkinson's disease^{32,34,35}. Evidence for involvement in autism spectrum disorder comes from Butler et al.³⁵, who compiled a list of clinically relevant genes for the disorder, with *DCAKD* among the participating susceptibility genes. Expression of *DCAKD* was also found to be implicated in Parkinson's disease³², a disorder previously associated with lower white matter integrity in tracts within the temporal, parietal and occipital lobes³⁶.

SLC35A4 belongs to the *SLC35* family, members of which act as transporters of nucleotide sugars, and is known to be expressed in brain³⁷. There is limited knowledge about its specific function, although a recent review investigating the subcellular localisation and topology of *SLC35A4* demonstrated that it localizes mainly to the Golgi apparatus³⁷.

Disease-linked genes—lower FA and higher MD (decreased white matter integrity)

Four genes (*SRAI*, *UBE3C*, *SEC14L4*, *PLEKHM1*) were associated with lower FA within several individual tracts pertaining to projection and association fibres, as well as with higher global MD. *SRAI* encodes both non-coding and protein-coding RNAs, is implicated in the regulation of numerous nuclear receptor activities, such as metabolism and chromatin organisation, and is known to be expressed in the brain. Kotan et al.³⁸ posited that *SRAI* plays a role in the initiation of puberty in humans by finding that inactivating *SRAI* variants were associated with idiopathic hypogonadotropic hypogonadism (IHH) in three independent families. IHH is a rare genetic disorder caused by the inability of the hypothalamus to secrete gonadotropin-releasing hormones (GnRH) or by the inability of GnRH to act on pituitary gonadotropes³⁹. These previous results might link the association of *SRAI* with projection fibres, which connect the cerebral cortex to the spinal cord and brainstem, as well as to other centres of the brain (e.g., thalamus).

UBE3C contains ubiquitin-protein ligase (E3), an enzyme which accepts ubiquitin from E2 before transferring it to the target lysine; ubiquitin targets proteins for degradation via the proteasome. *UBE3C* is expressed in numerous tissues, including the brain, and has been previously associated with some neuropsychiatric-related phenotypes. For instance, Garriock et al.⁴⁰ performed a GWAS to determine the association between genetic variation and Citalopram response. Although not genome-wide significant, their top finding was a SNP in proximity to *UBE3C* and was found to be associated with antidepressant response and MDD remission (rs6966038,

$p = 4.65e-07$ and $p = 3.63E-07$, respectively)⁴⁰. Moreover, Filatova et al.⁴¹ studied the expression of genes within the ubiquitin-proteasome protein degradation system, which is implicated in Parkinson's disease, in mice with MPTP-induced pre-symptomatic and early symptomatic stages of Parkinson's disease. They found decreased expression in the striatum and the substantia nigra of mice, which may lead to a decrease in performance of the system. This may in turn lead to accumulation of abnormal and toxic proteins which guide neuronal cell death⁴¹.

The specific function of *SEC14L4* has not yet been determined, although the protein encoded by it is similar to a protein encoded by the *SEC14* gene in *Saccharomyces cerevisiae*, which is essential to the biogenesis of Golgi-derived transport vesicles. Curwin and McMaster⁴² found that mutations in several *SEC14* domain-containing proteins in humans may be implicated in neurodegeneration, although it is not clear what the role of *SEC14L4* is within this context. Lastly, *PLEKHM1* is important in bone resorption, may be involved in vesicular transport in the osteoclast, and is weakly expressed in the brain. Although mutations in this gene have been associated with numerous phenotypes^{43,44}, none were neuropsychiatric-related.

Development-linked genes—higher FA and lower MD (increased white matter integrity)

For two of the eight genes (*CPNE1*, *NMT1*) we found higher FA and lower MD, indicating increased white matter integrity, associated with increased expression level as quantified by the corresponding eQTL.

CPNE1, which is thought to regulate molecular events at the cell membrane and cytoplasm, has previously been found to mediate several neuronal differentiation processes by interacting with intracellular signalling molecules. *CPNE1* has also been found to be highly expressed during brain development, indicating that it might be implicated in earlier developmental stages of neuronal function⁴⁵. Furthermore, C2 domains of *CPNE1*, calcium-dependent phospholipid-binding motives, have been shown to be implicated in neurite outgrowth of hippocampal progenitor HiB5 cells, which are hippocampal cell lines derived from the hippocampal anlagen of E16 rat^{46,47}. We provide evidence that *CPNE1* expression is associated with two tracts within projection fibres (FA) and with regional association fibres (MD), which link the cortex to lower brain areas. In mouse and human models, these findings may be of use when investigating neurite outgrowth from the hippocampus, which is part of the limbic system, an area located beneath the cortex.

NMT1 (N-myristoyltransferase) catalyzes the transfer of myristate (a rare 14-carbon saturated fatty acid) from CoA to proteins, and is expressed in numerous tissues, including ubiquitously in the brain. It has been found that

NMT1 is required for early mouse development, mainly due to its role in early embryogenesis⁴⁸. Expression of this gene has also been implicated in human brain tumours⁴⁹ and tumour cell proliferation⁵⁰. In our study, we found *NMT1* to be associated with tracts within thalamic radiations and projection fibres (FA) and global MD.

General discussion

In our study, we employed a novel strategy of investigating a direct association between eQTL scores and white matter tracts to uncover a relationship between specific regulatory variants and brain connectivity. Together, our findings indicate that increases in expression of these genes may be implicated in several processes which may directly or indirectly alter white matter microstructure, each with localised, pronounced effects in specific tracts. Further, while some of the significant associations had connections with other brain-related traits, such as neurite outgrowth or psychiatric and neurological disorders, others did not. Interestingly, decreased white matter microstructure integrity, as marked by lower FA and higher MD, was associated with eQTL scores which regulate expression of genes implicated in neuropsychiatric and neurological disorders. Conversely, increased white matter integrity, as marked by higher FA and lower MD, was associated with *CPNE1* and *NMT1*, which are important in developmental processes such as neurite outgrowth. In addition, encouragingly, regions of the corpus callosum (i.e., the forceps minor), the largest and arguably most reliably measured white matter tract in the brain, was demonstrated to be associated with all 8 scores for FA, and 7 for MD. These findings together suggest that utilising this approach to associate eQTL scores with white matter microstructure may add to previous research which found associations between genes and these brain-related traits and disorders. These genes or eQTL for them might indirectly implicate brain connectivity through other processes in which they participate.

The current study has several strengths and some potential limitations. First, to our knowledge, this study is the first one to compute eQTL scores for specific gene transcripts and attempt to associate them with white matter tracts. Moreover, our analysis consisted of a population-based sample of $N > 14,000$ individuals recruited to the UKB, large enough to make our findings robust and generalisable to other samples within the same age range, background and ethnicity. Lastly, our findings revealed novel associations which were not previously found in GWAS (²⁸; GWAS of g measures run locally), indicating a potential to use such scores for further discovery analyses.

However, a potential limitation in this study is calculation of scores for data taken from whole blood.

Although we note previous evidence indicates that whole blood can be used as a proxy for brain eQTL, important for study of in vivo brain traits¹⁴.

In summary, our results suggest that expression of the genes discussed above alter white matter microstructure and could facilitate the manifestation of numerous brain-related traits. Uncovering specific markers leading to the formation, maintenance and pathology of white matter could enable downstream analyses to elucidate links between genetics and neuroimaging in neurological and psychiatric disorders, as well as other brain-related traits.

Acknowledgements

This study is supported by a Wellcome Trust Strategic Award "Stratifying Resilience and Depression Longitudinally" (STRADL) (Reference 104036/Z/14/Z) and by the Sackler Foundation. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award (STRADL; Reference as above). DMH is supported by a Sir Henry Wellcome Postdoctoral Fellowship (Reference 213674/Z/18/Z) and a 2018 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (Ref: 27404). HCW is supported by a JMAS SIM fellowship from the Royal College of Physicians of Edinburgh and by an ESAT College Fellowship from the University of Edinburgh. Part of the work was undertaken in The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE), part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1); funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and MRC is gratefully acknowledged. Age UK (The Disconnected Mind project) also provided support for the work undertaken at CCACE. We would also like to thank Professor Chris Haley for valuable feedback on previous versions of this manuscript. We have deposited the manuscript on the repository archive *bioRxiv*.

Author details

¹Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. ²Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. ³Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. ⁴Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. ⁵Centre for Cognitive Ageing and Cognitive Epidemiology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK

Conflict of interest

A.M.M. has previously received grant support from Pfizer, Lilly and Janssen. These studies are not connected to the current investigation. In the past three years, S.M.L. has received grant and personal fees from Janssen, and personal fees from Otsuka and Sunovion, outside the submitted work. The remaining authors declare that they have no conflict of interest.

Publisher's note

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

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41398-020-0724-y>).

Received: 18 June 2019 Revised: 9 January 2020 Accepted: 13 January 2020
Published online: 04 February 2020

References

- Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. B* **368**, 20120362 (2013).
- Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
- Hurd, Y. L. Differential messenger RNA expression of prodynorphin and proenkephalin in the human brain. *Neuroscience* **72**, 767–783 (1996).
- Hurd, Y. L. Subjects with major depression or bipolar disorder show reduction of prodynorphin mRNA expression in discrete nuclei of the amygdaloid complex. *Mol. Psychiatry* **7**, 75–81 (2002).
- Gandal, M. J. et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).
- De Jong, S. et al. Expression QTL analysis of top loci from GWAS meta-analysis highlights additional schizophrenia candidate genes. *Eur. J. Hum. Genet.* **20**, 1004 (2012).
- Luo, X. J. et al. Systematic integration of brain eQTL and GWAS identifies ZNF323 as a novel schizophrenia risk gene and suggests recent positive selection based on compensatory advantage on pulmonary function. *Schizophrenia Bull.* **41**, 1294–1308 (2015).
- Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24**, 408–415 (2008).
- Stranger, B. E. et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1**, e78 (2005).
- Hernandez, D. G. et al. Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTL for human traits in blood and brain. *Neurobiol. Dis.* **47**, 20–28 (2012).
- Ramasamy, A. et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418 (2014).
- Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481 (2016).
- Zou, F. et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet.* **8**, e1002707 (2012).
- McKenzie, M., Henders, A. K., Caracella, A., Wray, N. R. & Powell, J. E. Overlap of expression quantitative trait loci (eQTL) in human brain and blood. *BMC Med. Genomics* **7**, 31 (2014).
- Qi, T. et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).
- Kochunov, P. et al. Heritability of fractional anisotropy in human white matter: a comparison of Human Connectome Project and ENIGMA-DTI data. *Neuroimage* **111**, 300–311 (2015).
- Vuoksimaa, E. et al. Heritability of white matter microstructure in late middle age: A twin study of tract-based fractional anisotropy and absolute diffusivity indices. *Hum. Brain Mapp.* **38**, 2026–2036 (2017).
- Sprooten, E. et al. Common genetic variants and gene expression associated with white matter microstructure in the human brain. *Neuroimage* **97**, 252–261 (2014).
- Shen, X. et al. Subcortical volume and white matter integrity abnormalities in major depressive disorder: findings from UK Biobank imaging data. *Sci. Rep.* **7**, 5547 (2017).
- Whalley, H. C. et al. Prediction of depression in individuals at high familial risk of mood disorders using functional magnetic resonance imaging. *PLoS ONE* **8**, e57357 (2013).
- Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245 (2016).
- Westra, H. J. et al. Systematic identification of trans eQTL as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238 (2013).
- 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 (2018).
- Barbu, M. C. et al. Association of whole-genome and NETRIN1 signaling pathway-derived polygenic risk scores for major depressive disorder and white matter microstructure in the UK Biobank. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **4**, 91–100 (2019).

25. Hagenaars, S. P. et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Mol. Psychiatry* **21**, 1624 (2016).
26. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203 (2018).
27. Shen, E. H., Overly, C. C. & Jones, A. R. The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. *Trends Neurosci.* **35**, 711–714 (2012).
28. Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210 (2018).
29. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
30. Gobbi, C. et al. Forceps minor damage and co-occurrence of depression and fatigue in multiple sclerosis. *Mult. Scler.* **20**, 1633–1640 (2014).
31. Mamiya, P. C., Richards, T. L. & Kuhl, P. K. Right forceps minor and anterior thalamic radiation predict executive function skills in young bilingual adults. *Front. Psychol.* **9**, 118 (2018).
32. Latourelle, J. C., Dumitriu, A., Hadzi, T. C., Beach, T. G. & Myers, R. H. Evaluation of Parkinson disease risk variants as expression-QTLs. *PLoS ONE* **7**, e46199 (2012).
33. Gonzalez-Lozano, M. A. et al. Dynamics of the mouse brain cortical synaptic proteome during postnatal brain development. *Sci. Rep.* **6**, 35456 (2016).
34. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 sub-phenotypes. *Cell* **173**, 1705–1715 (2018).
35. Butler, M. G., Rafi, S. K. & Manzardo, A. M. High-resolution chromosome ideogram representation of currently recognized genes for autism spectrum disorders. *Int. J. Mol. Sci.* **16**, 6464–6495 (2015).
36. Auning, E. et al. White matter integrity and cognition in Parkinson's disease: a cross-sectional study. *BMJ Open* **4**, e003976 (2014).
37. Sosicka, P. et al. An insight into the orphan nucleotide sugar transporter SLC35A4. *Biochim. Biophys. Acta Mol. Cell Res.* **1864**, 825–838 (2017).
38. Kotan, L. D. et al. Idiopathic hypogonadotropic hypogonadism caused by inactivating mutations in SRA1. *J. Clin. Res. Pediatr. Endocrinol.* **8**, 125 (2016).
39. Bianco, S. D. & Kaiser, U. B. The genetic and molecular basis of idiopathic hypogonadotropic hypogonadism. *Nat. Rev. Endocrinol.* **5**, 569 (2009).
40. Garriock, H. A. et al. A genomewide association study of citalopram response in major depressive disorder. *Biol. Psychiatry* **67**, 133–138 (2010).
41. Filatova, E. V. et al. Expression analysis of genes of ubiquitin-proteasome protein degradation system in MPTP-induced mice models of early stages of Parkinson's disease. In *Doklady Biochemistry and Biophysics*. 116–118 (Pleiades Publishing, 2014)
42. Curwin, A. & McMaster, C. Structure and function of the enigmatic Sec14 domain-containing proteins and the etiology of human disease. *Future Lipidol.* **3**, 399–410 (2008).
43. Fujiwara, T. et al. PLEKHM1/DEF8/RAB7 complex regulates lysosome positioning and bone homeostasis. *JCI Insight* **1**, e86330 (2016).
44. McEwan, D. G. et al. PLEKHM1 regulates autophagosome-lysosome fusion through HOPS complex and LC3/GABARAP proteins. *Mol. Cell* **57**, 39–54 (2015).
45. Kim, T. H. et al. Copine1 regulates neural stem cell functions during brain development. *Biochem. Biophys. Res. Commun.* **495**, 168–173 (2018).
46. Park, N. et al. Copine1 enhances neuronal differentiation of the hippocampal progenitor HiB5 cells. *Mol. Cells* **34**, 549–554 (2012).
47. Park, N. et al. Copine1 C2 domains have a critical calcium-independent role in the neuronal differentiation of hippocampal progenitor HiB5 cells. *Biochem. Biophys. Res. Commun.* **454**, 228–233 (2014).
48. Deng, L. et al. NMT1 inhibition modulates breast cancer progression through stress-triggered JNK pathway. *Cell Death Dis.* **9**, 1143 (2018).
49. Lu, Y. et al. Expression of N-myristoyltransferase in human brain tumors. *Neurochem. Res.* **30**, 9–13 (2005).
50. Ducker, C. E., Upson, J. J., French, K. J. & Smith, C. D. Two N-myristoyltransferase isozymes play unique roles in protein myristoylation, proliferation, and apoptosis. *Mol. Cancer Res.* **3**, 463–476 (2005).
51. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668 (2018).
52. Smith, B. H. et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med. Genet.* **7**, 74 (2006).