

RESEARCH ARTICLE

Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome

Denis A. Baird^{1*}, Jimmy Z. Liu², Jie Zheng¹, Solveig K. Sieberts³, Thanneer Perumal³, Benjamin Elsworth¹, Tom G. Richardson¹, Chia-Yen Chen², Minerva M. Carrasquillo⁴, Mariet Allen⁴, Joseph S. Reddy⁵, Philip L. De Jager^{6,7}, Nilufer Ertekin-Taner^{4,8}, Lara M. Mangravite³, Ben Logsdon³, Karol Estrada^{2,9}, Philip C. Haycock¹, Gibran Hemani¹, Heiko Runz², George Davey Smith^{1,10}, Tom R. Gaunt^{1,10*}, AMP-AD eQTL working group¹¹

1 MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, University of Bristol, Bristol, United Kingdom, **2** Translational Biology, Research and Development, Cambridge, Massachusetts, United States of America, **3** Sage Bionetworks, Seattle, Washington, United States of America, **4** Department of Neuroscience, Mayo Clinic Florida, Jacksonville, Florida, United States of America, **5** Department of Health Sciences Research, Mayo Clinic Florida, Jacksonville, Florida, United States of America, **6** Centre for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Medical Centre, New York, New York, United States of America, **7** Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Irving Medical Centre, New York, New York, United States of America, **8** Department of Neurology, Mayo Clinic Florida, Jacksonville, Florida, United States of America, **9** BioMarin Pharmaceuticals, San Rafael, California, United States of America, **10** NIHR Bristol Biomedical Research Centre, Oakfield House, University of Bristol, Bristol, United Kingdom

¶ Membership of the AMP-AD-eQTL Working Group is listed in the Acknowledgements section.
* Denis.Baird@bristol.ac.uk (DAB); Tom.Gaunt@bristol.ac.uk (TRG)



OPEN ACCESS

Citation: Baird DA, Liu JZ, Zheng J, Sieberts SK, Perumal T, Elsworth B, et al. (2021) Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome. *PLoS Genet* 17(1): e1009224. <https://doi.org/10.1371/journal.pgen.1009224>

Editor: Amanda J. Myers, University of Miami, Miller School of Medicine, UNITED STATES

Received: July 9, 2020

Accepted: October 26, 2020

Published: January 8, 2021

Copyright: © 2021 Baird et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: We deposited the full result files for our MR analysis on GitHub (available at [denisbrd/braineQTLMR_datasets](https://github.com/denisbrd/braineQTLMR_datasets), first release at DOI: [10.5281/zenodo.3778433](https://doi.org/10.5281/zenodo.3778433)).

Funding: DAB is funded by Biogen. PCH is supported by CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). JZ is a Vice-Chancellor's research fellow in the MRC Integrative Epidemiology Unit. GH is supported by the Wellcome Trust and Royal Society [208806/Z/17/Z], University of Bristol. TRG and GDS receive

Abstract

Discovering drugs that efficiently treat brain diseases has been challenging. Genetic variants that modulate the expression of potential drug targets can be utilized to assess the efficacy of therapeutic interventions. We therefore employed Mendelian Randomization (MR) on gene expression measured in brain tissue to identify drug targets involved in neurological and psychiatric diseases. We conducted a two-sample MR using cis-acting brain-derived expression quantitative trait loci (eQTLs) from the Accelerating Medicines Partnership for Alzheimer's Disease consortium (AMP-AD) and the CommonMind Consortium (CMC) meta-analysis study (n = 1,286) as genetic instruments to predict the effects of 7,137 genes on 12 neurological and psychiatric disorders. We conducted Bayesian colocalization analysis on the top MR findings (using $P < 6 \times 10^{-7}$ as evidence threshold, Bonferroni-corrected for 80,557 MR tests) to confirm sharing of the same causal variants between gene expression and trait in each genomic region. We then intersected the colocalized genes with known monogenic disease genes recorded in Online Mendelian Inheritance in Man (OMIM) and with genes annotated as drug targets in the Open Targets platform to identify promising drug targets. 80 eQTLs showed MR evidence of a causal effect, from which we prioritised 47 genes based on colocalization with the trait. We causally linked the expression of 23 genes with schizophrenia and a single gene each with anorexia, bipolar disorder and major depressive disorder within the psychiatric diseases and 9 genes with Alzheimer's disease, 6

funding from the UK Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/4 (TRG), MC_UU_00011/1 (GDS)). Mayo Clinic Florida work was supported by National Institute on Aging [U01AG046139 to NE-T]. This study was supported by the NIHR Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: JZL, C-YC and HR are employees and shareholders in Biogen. KE is an employee at BioMarin Pharmaceuticals. DAB is employed on a grant funded by Biogen. TRG and GDS received funding from Biogen for the work described here. The other authors have no competing interests to declare.

genes with Parkinson's disease, 4 genes with multiple sclerosis and two genes with amyotrophic lateral sclerosis within the neurological diseases we tested. From these we identified five genes (*ACE*, *GPNMB*, *KCNQ5*, *RERE* and *SUOX*) as attractive drug targets that may warrant follow-up in functional studies and clinical trials, demonstrating the value of this study design for discovering drug targets in neuropsychiatric diseases.

Author summary

Genetic association studies have been successful in identifying many genetic variants associated with disease risk, but it has been far more challenging to determine the genes through which these act. This is important, because such genes may encode effective drug targets for these diseases. We used Mendelian randomization (MR) and colocalization, two methods which in combination exploit these genetic variants to estimate the causal effects of individual genes. We applied this approach to 12 neurological and psychiatric diseases using data from the AMP-AD and CMC brain expression quantitative locus dataset, which is large enough to provide robust evidence for the relationship between genetic variants and gene expression. We found a causal relationship between the change in expression of 47 genes and increased disease risk across the 12 diseases we tested. As drug targets with human genetic evidence are far more likely to be approved in clinical trials, these findings provide a valuable list of potential therapeutic targets, including the *ACE*, *GPNMB*, *KCNQ5*, *RERE* and *SUOX* genes.

Introduction

There is a compelling need to improve the processes to discover new drugs for human disease, which is time consuming and expensive, with high rates of attrition [1]. Attrition rates have been particularly high for therapies against brain-related diseases. For instance, a survey of 413 Alzheimer's disease trials reported a failure rate of 99.6% [2]. Neurological and psychiatric diseases are highly heritable [3]. Genome-wide association studies (GWAS) have identified numerous genetic loci associated with disease risk [4,5], and can therefore provide evidence to identify molecular pathways involved in disease that can be intervened on by drugs. Indeed, drug targets with human genetic evidence are more than twice as likely to be approved than those without [6,7].

One approach to nominate potential drug targets through genetics is to integrate GWAS associations with gene expression quantitative trait loci (eQTLs) to select the most likely functional variants and the genes controlled by these variants. Recent studies have adopted such an approach through a combination of two sample Mendelian randomization (MR) and colocalization methods to infer whether a causal relationship exists between gene expression and disease outcomes. For example, two sample MR studies have systematically explored causal relationships between the transcriptome [8,9,10], methylome [11] and proteome [12] and complex traits. Zhu et al performed MR between blood eQTLs and five complex traits and found potential causal effects for 126 genes in total, including 25 novel genes previously unreported in the literature [9]. Richardson et al [11] and Zheng et al [12] performed MR on a wider phenome-wide scale and found putative causal effects for 1,148 methylation quantitative trait loci (mQTLs) on 139 complex traits and 111 protein quantitative trait loci (pQTLs) on 225 complex traits respectively. After causal variants have been identified, a follow up MR

phenome-wide association analysis [13] can then be conducted to assess a gene's suitability as a drug target. Specifically, it can be assessed whether associations to clinical endpoints match the intended indications, or whether there might be a risk for adverse events [14].

A major challenge for two sample MR of gene expression with brain diseases is to obtain tissue-specific expression datasets at large enough sample sizes to be powered for obtaining robust eQTL associations to instrument genes. In this study, we utilized summary statistics from a recent cerebral cortex eQTL meta-analysis ($n = 1,286$) that combined RNA-seq data from the Accelerating Medicines Partnership for Alzheimer's Disease consortium (AMP-AD) and the CommonMind Consortium (CMC) [15] (henceforth referred to as AMP-CMC). We extracted 7,689 eQTL instruments from this study and conducted a transcriptome-wide two sample MR and colocalization study between brain eQTLs and 12 neurological and psychiatric conditions. By identifying shared genetic effects between brain-specific gene expression and Central Nervous System (CNS) disease, our study establishes causal roles for several genes across the conditions tested that could be starting points for the development of efficient drugs.

Results

Selection of instrumental variables for Mendelian Randomization from AMP-CMC

AMP-CMC combines RNA-seq data from 1,286 human cerebral cortex samples from the Accelerating Medicines Partnership for Alzheimer's Disease (AMP-AD, composed of the ROSMAP and MAYO cohorts) and the CommonMind consortia (CMC, composed of the MSSM-Penn_Pitt cohorts) [15]. We first assessed how eQTLs used as instruments from this cohort compared to alternative brain eQTL datasets available for public download, specifically a meta-analysis across AMP, CMC and GTEx on pre-frontal cortex tissue (brain cortex meta-analysis), and a meta-analysis across 10 different brain tissues available in GTEx (GTEx brain meta-analysis) [10]. Gene expression effects for these brain eQTL datasets agreed well with our study ($r = 0.72$, 95% CI [0.71, 0.73] brain cortex meta-analysis, $r = 0.70$, 95% CI [0.68, 0.71] for GTEx brain meta-analysis). Of the 6,790 harmonised brain cortex meta-analysis effects, 3,922 (57.8%) effects achieved a genome-wide significance level used to select instruments in this study ($P < 5 \times 10^{-8}$). An additional 2,224 effects (32.8%) reached nominal significance ($P < 0.05$) and 644 effects (9.5%) showed no evidence of association ($P > 0.05$). Of the 6,628 harmonised effects from the GTEx brain meta-analysis, 1,642 (24.8%) effects reached $P < 5 \times 10^{-8}$, an additional 3,256 (49.1%) reached $P < 0.05$, and 1,730 (26.1%) showed no evidence of association ($P > 0.05$). Of the nominally significant effects, only 49 (0.8%) and 71 (1.4%), respectively, showed opposite directionality of effect relative to our study (S1 Table).

Blood eQTLs are sometimes used as instruments when examining brain-related disorders, but this approach can be problematic for genes that are differentially expressed between brain and blood. To better understand differences, we compared AMP-CMC instruments against the currently largest public blood eQTL resource, eQTLGen [16]. As expected, the correlation between blood and brain across all gene expression effects was modest ($r = 0.48$, 95% CI [0.46, 0.50]). However, of the 5,820 harmonised eQTLGen effects, 4,133 (71.0%) effects reached $P < 5 \times 10^{-8}$ and an additional 984 effects (16.9%) nominal significance ($P < 0.05$). Interestingly, a substantial proportion of eQTLs (1,028 or 21.2% of the 4,836 eQTLGen effects at $P < 0.05$) showed opposite direction of gene expression effect between blood and brain (S1 Table), the reasons for which will require clarification in future studies. Correlations of AMP-CMC eQTLs to eQTLs from other GTEx tissues ranged from 0.80 (for brain cortex GTEx brain tissues) to 0.51 (for EBV-transformed lymphocyte cells), with variable agreement on

directionality (S2 Table). Focusing analyses on the subset of eQTLs identified by and reported as of identical directionality in other studies further improved specificity of our results (S1 and S2 Tables).

Mendelian randomization identifies shared genetic effects of brain eQTLs with neurological and psychiatric disease

We conducted MR to determine the shared genetic effects between gene expression (using genetically predicted transcript levels) and 12 neurological and psychiatric diseases (see flow chart in S1 Fig for overview of study design). Our MR analysis included a total of 80,557 Wald ratio tests across 7,137 genes. Of these, 80 Wald ratio effects across 5 out of the 7 psychiatric diseases tested and 4 out of the 5 neurological diseases tested showed evidence of MR at the multiple correction threshold of $P < 6 \times 10^{-7}$ (S3 and S4 Tables). The Wald ratio estimates the log odds change in disease risk per standard deviation (SD) change in gene expression relative to the risk allele for the instrumenting SNP. A positive Wald ratio implies a causal relationship between increased expression of the gene and increased disease risk, whereas a negative Wald ratio implies a causal relationship between decreased expression of the gene and increased disease risk. For the psychiatric diseases, we discovered a relationship between 38 genes and schizophrenia, 2 genes and attention deficient hyperactivity disorder, and a single gene each for anorexia, bipolar disorder and major depressive disorder risk. For the neurological diseases we detected a relationship between 16 genes and Parkinson's disease, 15 genes and Alzheimer's disease, 4 genes and multiple sclerosis and 2 genes and amyotrophic lateral sclerosis risk. 47 of these 80 signals showed evidence of colocalization between gene expression and disease ($PP_4 > 70\%$) indicating that both gene expression and disease trait share the same casual variant in the region (i.e. eQTL instrument is not coincidentally associated with disease SNP due to Linkage Disequilibrium (LD) patterns in the region). We prioritized these co-localizing eQTLs as pointing to the most likely causal genes and substrate for further analyses (Tables 1 and 2). One third (17 genes in total) of the 47 genes co-occurred within the same region of the genome (i.e. instrumented eQTLs used for the different genes were in LD and colocalized with disease, making it difficult to disentangle the causal gene in this region).

Shared genetic effects between brain gene expression and psychiatric diseases. MR and colocalization analyses identified 23 potentially causal genes with evidence of a shared genetic effect between gene expression (eQTL) and schizophrenia risk (Table 1). Of these, 12 genes showed a positive direction of Wald ratio effect, indicating a relationship between increased gene expression and increased schizophrenia risk (*SF3B1*, *IN080E*, *CNTN4*, *NMB*, *FES*, *GOLGA2P7*, *CLCN3*, *RERE*, *ZNF823*, *ZC3H7B*, *KCNQ5*, and *NAT8*). 11 genes showed a negative direction of Wald ratio effect indicating a relationship between decreased gene expression and increased schizophrenia risk (*FTCDNL1*, *FURIN*, *AC243562.2*, *AC105749.1*, *PCCB*, *PTPRU*, *THOC7*, *LINC02210*, *GATAD2A*, *FAM85B* and *FAM86B3P*). Notably, 7 out of the 23 genes reside in close genomic proximity, challenging attribution of causality for schizophrenia risk (Fig 1). Instruments for *AC243562.2* and *GOLGA2P7* are in high LD with each other ($r^2 = 0.91$) and in moderate LD with eQTLs impacting *NMB* ($r^2 \approx 0.55$ with each instrument). Instruments for the nearby *FURIN* and *FES* genes are in moderate LD with each other ($r^2 = 0.67$). Likewise, eQTLs for *FAM86B3P* and *FAM85B* are in high LD ($r^2 = 0.83$), limiting resolution of MR analyses. For the remaining 6 psychiatric diseases, a positive Wald ratio effect was observed for *RHEBL* expression and bipolar disorder, and negative Wald ratio effects for *SUOX* expression and anorexia risk, and for *PTPN1* expression and major depressive disorder risk.

Table 1. Genes we related with psychiatric traits in the MR and colocalization analysis. Wald ratio (WR) estimate with standard error (SE) and p-value, coloc (posterior probability of colocalization between gene expression and outcome). WR represents the genetically predicted log odds change in risk per standard deviation (SD) change in gene expression.

outcome	exposure gene	SNP	cytogenic	WR	SE	P	coloc
anorexia	<i>SUOX</i>	rs1081975	12q13.2	-0.366	0.072	3.00x10 ⁻⁷	93.4
bipolar disorder	<i>RHEBL1</i>	rs7969091	12q13.12	0.400	0.075	8.63x10 ⁻⁸	98.2
major depressive disorder	<i>PTPN1</i>	rs718050	20q13.13	-0.195	0.037	1.92x10 ⁻⁷	98.4
schizophrenia	<i>FURIN</i>	rs4702	15q26.1	-0.237	0.034	2.55x10 ⁻¹²	100.0
	<i>ZNF823</i>	rs72986630	19p13.2	0.238	0.044	4.90x10 ⁻⁸	100.0
	<i>THOC7</i>	rs832187	3p14.1	-0.163	0.029	2.57x10 ⁻⁸	99.3
	<i>PTPRU</i>	rs1498232	1p35.2	-0.256	0.042	1.22x10 ⁻⁹	99.0
	<i>FAM85B</i>	rs2980436	8p23.1	-0.126	0.023	7.47x10 ⁻⁸	98.9
	<i>KCNQ5</i>	rs2492966	6q13	0.241	0.045	1.05x10 ⁻⁷	98.6
	<i>LINC00222</i>	rs9398171	6q21	-0.253	0.046	3.05x10 ⁻⁸	98.6
	<i>CLCN3</i>	rs10520163	4q33	0.248	0.043	9.19x10 ⁻⁹	97.9
	<i>RERE</i>	rs2708630	1p36.23	0.193	0.034	1.54x10 ⁻⁸	96.7
	<i>SF3B1</i>	rs2564383	2q33.1	0.265	0.040	4.82x10 ⁻¹¹	95.6
	<i>FTCDNL1</i>	rs281794	2q33.1	-0.167	0.023	7.75x10 ⁻¹³	95.2
	<i>FAM86B3P</i>	rs2945253	8p23.1	-0.075	0.015	3.54x10 ⁻⁷	94.5
	<i>ZC3H7B</i>	rs11090045	22q13.2	0.273	0.051	9.77x10 ⁻⁸	93.4
	<i>INO80E</i>	rs3814880	16p11.2	0.096	0.016	6.45x10 ⁻¹⁰	92.7
	<i>GATAD2A</i>	rs2905435	19p13.11	-0.137	0.025	4.11x10 ⁻⁸	92.4
	<i>NAT8</i>	rs10179134	2p13.1	0.145	0.028	2.16x10 ⁻⁷	91.4
	<i>AC105749.1</i>	rs11706780	3p22.2	-0.245	0.039	4.17x10 ⁻¹⁰	90.7
	<i>PCCB</i>	rs527888	3q22.3	-0.175	0.029	8.79x10 ⁻¹⁰	88.8
	<i>AC243562.2</i>	rs12905223	15q25.2	-0.113	0.018	4.82x10 ⁻¹⁰	87.5
	<i>CNTN4</i>	rs6796313	3p26.3	0.313	0.051	6.76x10 ⁻¹⁰	81.6
<i>NMB</i>	rs56864281	15q25.3	0.307	0.051	1.38x10 ⁻⁹	78.8	
<i>GOLGA2P7</i>	rs2019611	15q25.2	0.110	0.019	2.64x10 ⁻⁹	75.2	
<i>FES</i>	rs6224	15q26.1	0.255	0.042	1.58x10 ⁻⁹	70.5	

<https://doi.org/10.1371/journal.pgen.1009224.t001>

Shared genetic effects between gene expression and neurological diseases. MR and colocalization analyses further identified 21 potentially causal genes with evidence of a shared genetic effect between gene expression and five neurological diseases tested (Table 2). We detected a positive Wald ratio effect between the expression of three genes (*CR1*, *CCDC6*, *TSPAN14*), and a negative Wald ratio effect between the expression of 6 genes (*AC012146.1*, *PRSS36*, *ZNF646*, *KAT8*, *ACE*, *CCNT2-AS1*) and the risk for Alzheimer's disease. Of these, instruments for *ZNF646* and *KAT8* are in high LD ($r^2 = 0.85$), and for *PRSS36* in relatively low LD ($r^2 \approx 0.2$ with each instrument) (Fig 2). We further found a positive Wald ratio effect between the expression of two genes (*GPNMB*, *STX4*) and Parkinson's disease risk, and a negative Wald ratio effect between the expression of four genes (*KLHL7-DT*, *HSD3B7*, *GRN*, *AC135050.3*) and Parkinson's disease risk. Instruments for *STX4*, *AC135050.3* and *HSD3B7* are in high LD ($r^2 > 0.85$), and for *KLHL7-DT* and *GPNMB* are in moderate LD ($r^2 = 0.64$) (Fig 2). Moreover, we identified Wald ratio effects between four genes (*IQCB*, *TTC34*, *MPV17L2* and *SLC30A7*) and multiple sclerosis risk, and effects between two genes (*SCFD1*, *G2E3*) and amyotrophic lateral sclerosis risk. *SCFD1* and *G2E3* instruments are in high LD ($r^2 = 0.99$).

Identification of pleiotropic effects across neurological and psychiatric diseases. To assess whether the neurological and psychiatric diseases analysed in this study share pleiotropic relationships, we next queried for evidence of common Wald ratio effects at a less stringent

Table 2. Genes we related with neurological traits detected in the MR and colocalization analysis. WR with SE and p-value, coloc (posterior probability of colocalization between gene expression and outcome). WR represents the genetically predicted log odds change in risk per SD change in gene expression.

outcome	exposure gene	SNP	cytogenic	WR	SE	P	coloc
Alzheimer's disease	<i>CR1</i>	rs679515	1q32.2	0.274	0.028	2.15x10 ⁻²³	99.6
	<i>CCDC6</i>	rs1171830	10q21.2	0.222	0.039	1.22x10 ⁻⁸	99.5
	<i>ACE</i>	rs4295	17q23.3	-0.153	0.028	3.37x10 ⁻⁸	99.0
	<i>TSPAN14</i>	rs7097656	10q23.1	0.131	0.025	2.64x10 ⁻⁷	98.5
	<i>KAT8</i>	rs12597511	16p11.2	-0.133	0.023	6.18x10 ⁻⁹	95.5
	<i>ZNF646</i>	rs8050894	16p11.2	-0.216	0.037	3.53x10 ⁻⁹	93.7
	<i>CCNT2-AS1</i>	rs766271	2q21.3	-0.151	0.030	5.06x10 ⁻⁷	91.7
	<i>PRSS36</i>	rs55667375	16p11.2	-0.135	0.022	1.66x10 ⁻⁹	83.6
	<i>AC012146.1</i>	rs56377155	17p13.2	-0.144	0.023	5.62x10 ⁻¹⁰	77.0
amyotrophic lateral sclerosis	<i>SCFD1</i>	rs229184	14q12	0.141	0.028	5.14x10 ⁻⁷	92.6
	<i>G2E3</i>	rs229243	14q12	-0.201	0.040	5.94x10 ⁻⁷	92.2
multiple sclerosis	<i>MPV17L2</i>	rs62120364	19p13.11	-0.477	0.087	4.34x10 ⁻⁸	99.9
	<i>TTC34</i>	rs4310388	1p36.32	-0.487	0.066	2.26x10 ⁻¹³	99.8
	<i>IQCB1</i>	rs6438665	3q13.33	0.118	0.022	8.08x10 ⁻⁸	90.9
	<i>SLC30A7</i>	rs6681932	1p21.2	-0.357	0.067	9.85x10 ⁻⁸	82.9
Parkinson's disease	<i>GRN</i>	rs5848	17q21.31	-0.242	0.047	2.18x10 ⁻⁷	98.4
	<i>STX4</i>	rs7184567	16p11.2	0.318	0.051	6.13x10 ⁻¹⁰	95.9
	<i>GPNMB</i>	rs858239	7p15.3	0.121	0.017	2.37x10 ⁻¹³	95.5
	<i>HSD3B7</i>	rs7196726	16p11.2	-0.301	0.051	2.71x10 ⁻⁹	89.8
	<i>KLHL7-DT</i>	rs73272053	7p15.3	-0.295	0.042	2.69x10 ⁻¹²	77.2
	<i>AC135050.3</i>	rs35713203	16p11.2	-0.235	0.046	3.74x10 ⁻⁷	72.2

<https://doi.org/10.1371/journal.pgen.1009224.t002>

significance threshold of $P < 0.05$ for our 47 most likely causal candidate genes. 141 of the 551 Wald ratio estimates showed a relationship with a second disease at this cut-off. Of these, nine genes also passed colocalization ($PP_4 > 70\%$) with this alternative disease. This included six of the schizophrenia genes (*GOLGA2P7*, *NMB*, *AC105749.1*, *FES*, *FURIN*, *FTCDNL1*), both of the amyotrophic lateral sclerosis genes (*SCFD1* and *G2E3*), and one of the genes related with Parkinson's disease risk (*GRN*) (Table 3). For instance, expression levels of *G2E3* and *SCFD1* not only appear to causally effect the risk for amyotrophic lateral sclerosis but also attention deficient disorder, albeit with alternative directionalities. Wald ratio effects with identical directionality for two diseases were observed for *GRN* with Parkinson's disease and Alzheimer's disease; for *AC105749.1*, *GOLGA2P7* and *NMB* with schizophrenia and bipolar disorder; and *FES* and *FURIN* with schizophrenia and major depressive disorder. Permutation testing suggested that the likelihood to observe these shared genetic effects by chance was very low (empirical p-value $< 1 \times 10^{-4}$; Methods).

Follow up MR analysis on selected instruments to assess pleiotropy, tissue specificity and reverse causation

For the 47 signals identified from our primary MR analysis we conducted a series of follow-up analyses to assess the validity of our instruments.

Assessment of pleiotropy phenome-wide: enrichment for association with brain-related traits with no evidence of horizontal pleiotropy. It is possible that undetected horizontal pleiotropy may explain the MR relationship between gene expression and disease observed in our study. This occurs when the instrument is associated with risk factors that exert their effect on disease through distinct pathways, which is in contrast to vertical pleiotropy, where the

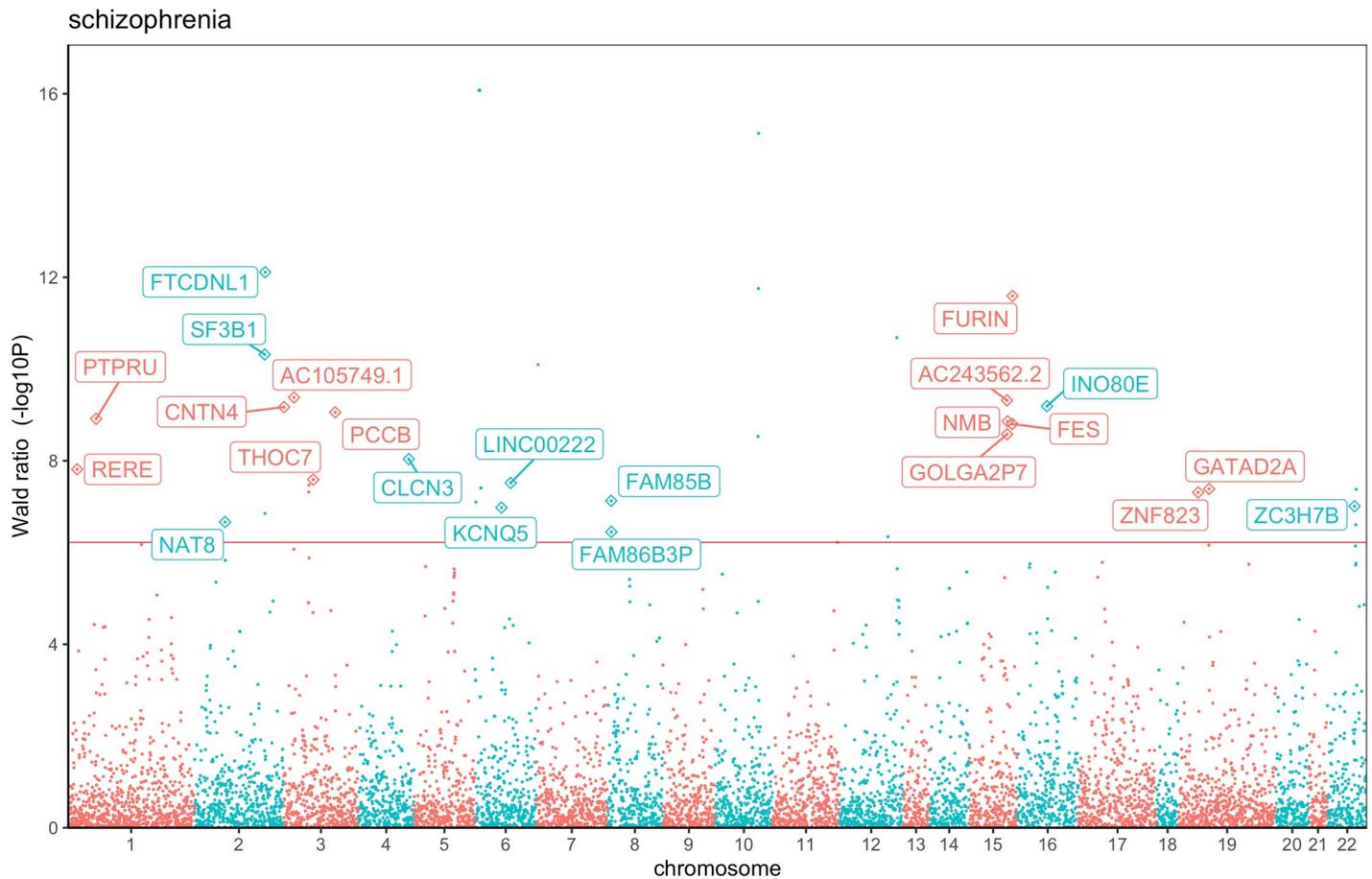


Fig 1. Manhattan plot showing the MR relationships between gene expression changes and schizophrenia risk. Chromosomal position is on the x-axis and p-value for the Wald ratio estimate ($-\log_{10}$ scaled) is on the y-axis. The prioritised genes which passed the multiple correction threshold (solid red line) and colocalization analysis ($PP_4 > 70\%$) are annotated as diamonds and labelled with the gene name. *FAM86B3P* and *FAM85B* genes on chromosome 8 and *AC243562.2*, *GOLGA2P7* and *NMB* genes and *FURIN* and *FES* genes on chromosome 15 are located close together (in LD with each other), challenging attribution of the causal gene at these loci.

<https://doi.org/10.1371/journal.pgen.1009224.g001>

instrument is associated with risk factors that act through the same pathway, and does not violate the assumptions of MR [17,18]. We therefore conducted a phenome-wide assessment on the 47 selected instruments in order to assess the extent of pleiotropy and risk of violating MR assumptions. For this, a total of 9,441 Wald ratio tests were computed across 219 disease and risk factor traits (see Methods). 133 Wald ratios showed evidence of a relationship at the multiple testing correction threshold ($P < 6 \times 10^{-6}$) and an additional 1,037 Wald ratios showed weaker evidence of a relationship ($P < 0.05$). Of these MR effects, 103 showed evidence of colocalization between gene expression and the selected trait ($PP_4 > 70\%$), which consisted of 64 from the 133 Wald ratios passing the threshold of $P < 6 \times 10^{-6}$ and an additional 39 from the 1,037 Wald ratios passing the threshold of $P < 0.05$. 1,549 (16%) of the Wald ratios pertained to brain-related traits (defined as neurological, psychiatric, education, personality and behaviour categories), of which 41 colocalised: 29 Wald ratios at $P < 6 \times 10^{-6}$ ($29/1549 = 1.87\%$) and an additional 12 Wald ratios at $P < 0.05$ ($41/1549 = 2.65\%$). 7,892 (84%) of the Wald ratios corresponded to other non-brain related traits of which 62 colocalised: 35 Wald ratios at $P < 6 \times 10^{-6}$ ($35/7892 = 0.44\%$) and an additional 27 Wald ratios at the $P < 0.05$ threshold ($62/7892 = 0.79\%$). Overall, the phenome-wide MR analysis showed evidence for enrichment of colocalization between gene expression and brain-related traits relative to colocalization with

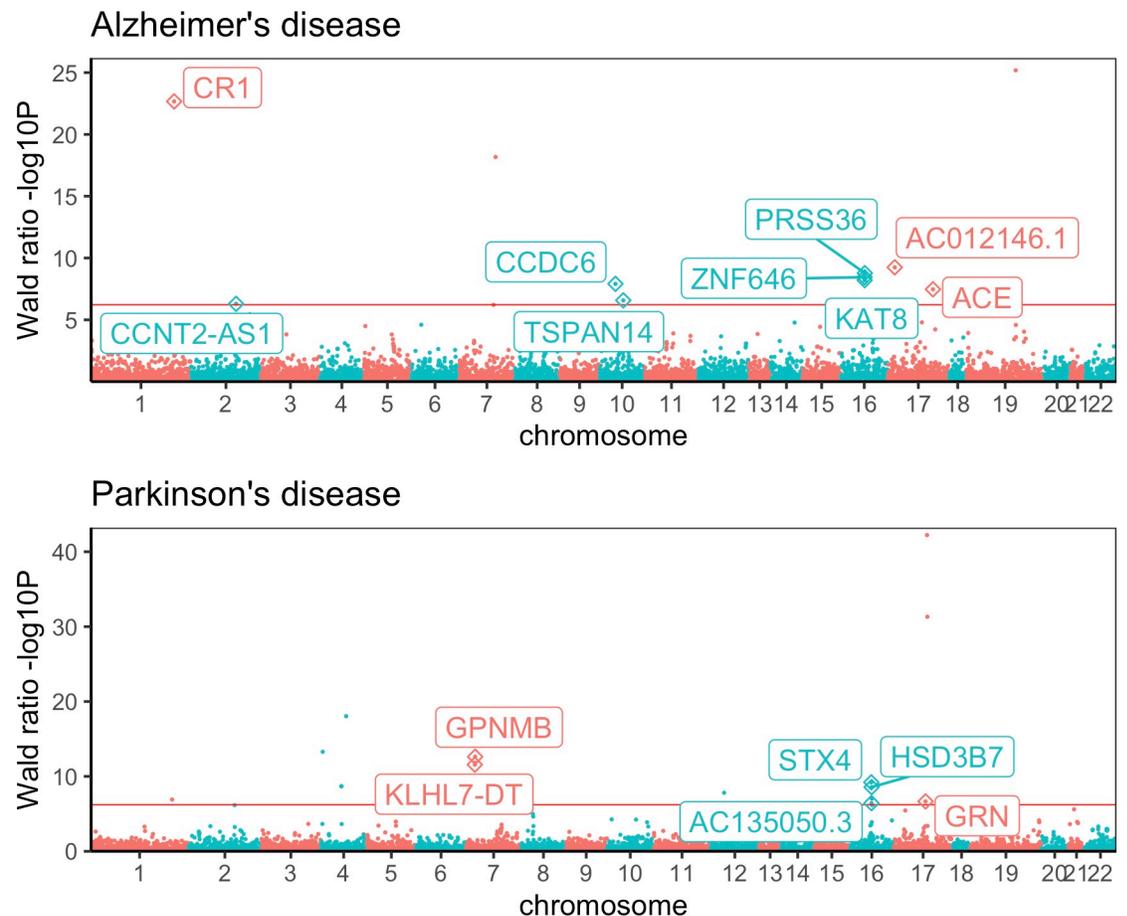


Fig 2. Manhattan plots showing the MR relationships between gene expression changes and Alzheimer's disease and Parkinson's disease outcomes. Chromosomal position is on the x-axis and p-value for the Wald ratio estimate ($-\log_{10}$ scaled) is on the y-axis. The prioritised genes which passed the multiple correction threshold (solid red line) and colocalization analysis ($PP_4 > 70\%$) are annotated as diamonds and labelled with the gene name. For Alzheimer's disease the *ZNF646*, *KAT8* and *PRSS36* genes on chromosome 16 are in LD, and for Parkinson's disease the *GPNMB* and *KLHL7-DT* genes on chromosome 7 and the *STX4*, *AC135050.3* and *HSD3B7* on chromosome 16 are in LD, challenging attribution of the causal gene at these loci.

<https://doi.org/10.1371/journal.pgen.1009224.g002>

other traits (Fisher's exact test p-values = $p < 0.0001$ for both the $P < 6 \times 10^{-6}$ and $P < 0.05$ thresholds), which suggests that much of the pleiotropy present is likely to be vertical (i.e. potentially acting through the same brain related pathway). Colocalization was detected with an additional non-brain trait for 20 of the 47 genes at the $P < 6 \times 10^{-6}$ and another 4 genes at the $P < 0.05$ threshold (S5 Table). Although this shows that non-brain specific effects are present none of the traits appear to open horizontal pleiotropic pathways which would violate the MR. We found evidence of non-brain specific effects (Wald Ratio $P < 0.05$ and coloc $PP_4 > 70\%$) for two of the genes (*FES* and *FURIN*) we had previously identified as having pleiotropy effects across the primary outcomes in our main MR. Increased *FES* expression colocalised with increased birth weight, increased ovarian cancer risk and decreased blood pressure along with decreased risk of hypertension and other related cardiovascular events and increased *FURIN* colocalised with decreased birth weight and increased angina risk (S5 Table). Our phenome-wide analysis suggests there is a limited scope for horizontal pleiotropy to affect our MR results. However, it is challenging to distinguish horizontal from vertical pleiotropy without an intricate and complete understanding of the pathways, therefore latent pleiotropic factors could still be confounding certain MR estimates.

Table 3. Genes which shared MR and colocalization evidence across the psychiatric and neurological outcomes we tested. The outcome detected in our main MR is highlighted in bold. Amyotrophic lateral sclerosis (ALS), attention deficient disorder (ADHD), Parkinson's disease (PD), Alzheimer's disease (AD), schizophrenia (Sz), multiple sclerosis (MS), bipolar disorder (BD), and major depressive disorder (MDD).

gene	SNP	cytogenic	outcome	WR	SE	P	coloc
<i>G2E3</i>	rs229243	14q12	ALS	-0.201	0.040	5.94x10 ⁻⁷	92.2
<i>G2E3</i>	rs229243	14q12	ADHD	0.136	0.040	6.47x10 ⁻⁴	71.4
<i>SCFD1</i>	rs229184	14q12	ALS	0.141	0.028	5.14x10 ⁻⁷	92.6
<i>SCFD1</i>	rs229184	14q12	ADHD	-0.092	0.027	8.16x10 ⁻⁴	70.6
<i>GRN</i>	rs5848	17q21.31	PD	-0.242	0.047	2.18x10 ⁻⁷	98.4
<i>GRN</i>	rs5848	17q12.31	AD	-0.145	0.034	1.60x10 ⁻⁵	97.9
<i>AC105749.1</i>	rs11706780	3p22.2	Sz	-0.245	0.039	4.17x10 ⁻¹⁰	90.7
<i>AC105749.1</i>	rs11706780	3p22.2	BD	-0.370	0.085	1.33x10 ⁻⁵	97.9
<i>FTCDNL1</i>	rs281794	2q33.1	Sz	-0.167	0.023	7.75x10 ⁻¹³	95.2
<i>FTCDNL1</i>	rs281794	2q33.1	MS	0.158	0.047	7.10x10 ⁻⁴	98.3
<i>GOLGA2P7</i>	rs2019611	15q25.2	Sz	0.110	0.019	2.64x10 ⁻⁹	75.2
<i>GOLGA2P7</i>	rs2019611	15q25.2	BD	0.130	0.038	6.87x10 ⁻⁴	73.0
<i>NMB</i>	rs56864281	15q25.3	Sz	0.307	0.051	1.38x10 ⁻⁹	78.8
<i>NMB</i>	rs56864281	15q25.3	BD	0.332	0.112	2.97x10 ⁻³	86.9
<i>FES</i>	rs6224	15q25.2	Sz	0.255	0.042	1.58x10 ⁻⁹	70.5
<i>FES</i>	rs6224	15q25.2	MDD	0.120	0.031	1.42x10 ⁻⁴	76.8
<i>FURIN</i>	rs4702	15q26.1	Sz	-0.237	0.034	2.55x10 ⁻¹²	100.0
<i>FURIN</i>	rs4702	15q26.1	MDD	-0.096	0.025	1.34x10 ⁻⁴	78.4

<https://doi.org/10.1371/journal.pgen.1009224.t003>

Molecular pleiotropy is pervasive. Another potential source of horizontal pleiotropy could be through instrument association with the expression of multiple cis or trans genes (molecular pleiotropy). We observed in our main MR analysis that several of the genes resided within the same genomic region, indicating potential for molecular pleiotropy. However, our MR analysis will miss genes where the eQTL could not be included as an instrument due to the SNP not being robustly associated enough ($P < 5 \times 10^{-8}$) or not finding the SNP in the outcome. Therefore, in order to more extensively evaluate pleiotropy within the cis-region, the 47 instruments used for the genes identified in our main MR were looked up across the gene transcripts available in AMP-CMC (without employing a p-value threshold). A total of 1,482 SNP-gene expression effect estimates were extracted. Of these 143 SNPs were associated with gene expression effects at a multiple testing correction threshold of $P < 3.4 \times 10^{-5}$, from which 87 also passed colocalization ($PP_4 > 70\%$) with the primary outcome. 23 of the 47 instruments showed evidence of colocalization between the primary outcome and more than one gene (S6 Table). In terms of the degree of molecular pleiotropy this consisted of 11 instruments associated with 3 genes, 10 instruments with 2 genes, and 2 instruments with 5 different genes. For the other 24 instruments which colocalised with the expression of a single gene only, there is less ambiguity regarding the causal variant and therefore potential genetic pathways involved which strengthen these as drug target candidates.

Evidence of tissue specific effects. We investigated whether our top MR findings would be generalisable to other brain tissues by looking up the instruments within the 13 brain tissues available in GTEx version 7. 44 of our 47 instruments were also present in GTEx, all SNPs except one (rs9398171-*LINC00222* instrument) had effect estimates available across every brain tissue. 167 tissue effects from 566 effects available in total, across 27 of the 44 instruments, showed evidence of eQTL association at a multiple-testing corrected threshold ($P < 9 \times 10^{-5}$, $\alpha = 0.05$) in the GTEx tissues (S7 Table). All these associations except one (rs7184567-*STX4* association in Cerebellum tissue) had the same direction of effect as our

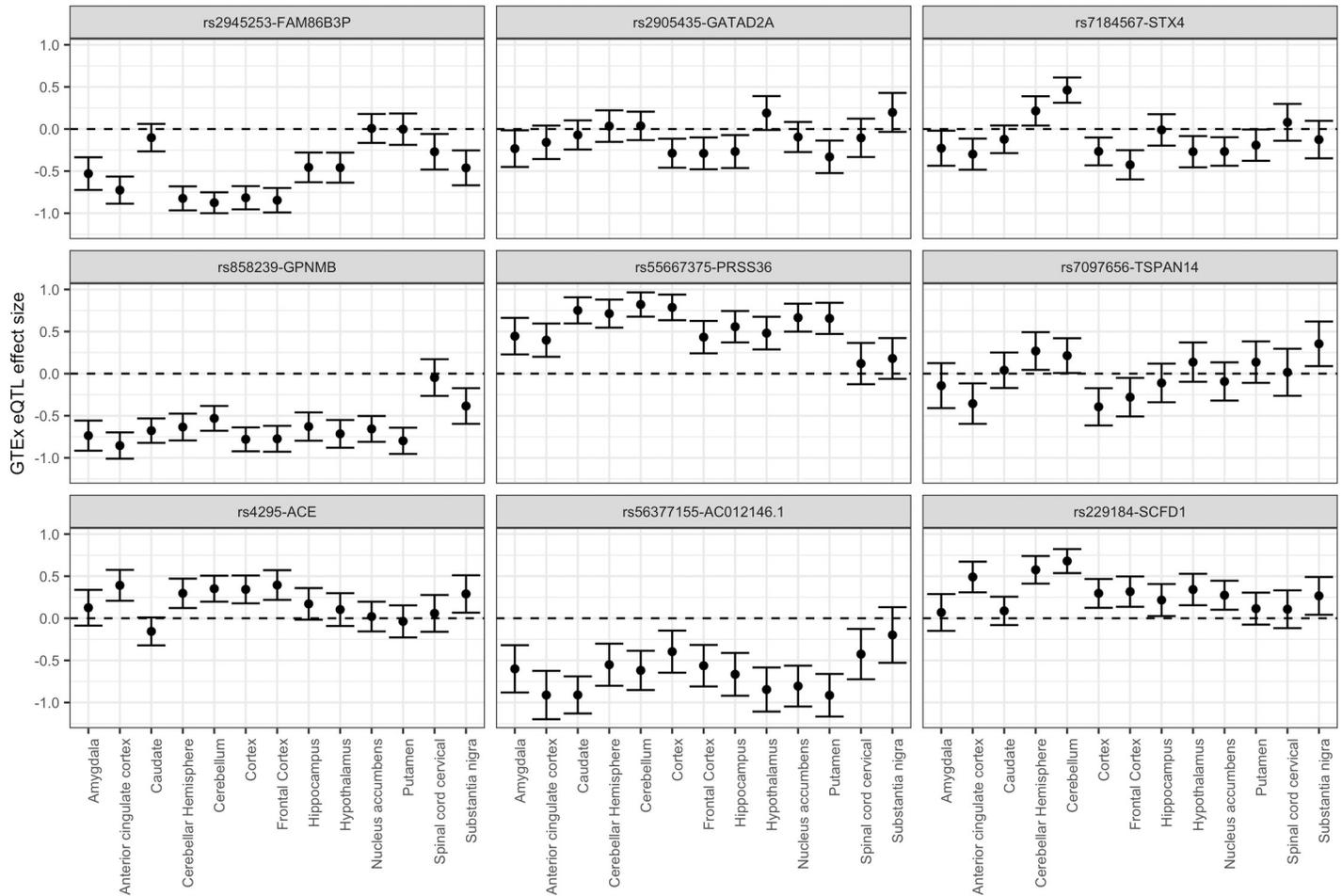


Fig 3. Error bar plot (95% confidence interval around the eQTL effect is displayed) comparing gene expression effects across the 13 brain tissues available in GTEx version 7. 9 of our 47 prioritised genes showed statistical evidence of effect size heterogeneity across the brain tissues. *STX4* and *TSPAN14* genes showed evidence of differing direction of gene expression effect for certain brain tissues.

<https://doi.org/10.1371/journal.pgen.1009224.g003>

instrument. As we would expect, the overall correlation between the eQTL effects in our study and GTEx (considering all 13 brain tissues grouped together) was moderately high ($r = 0.74$, 95% CI = [0.71, 0.78]). Correlations were also computed within the separate tissues and were of a similar magnitude. 9 of the instruments showed evidence of eQTL effect size heterogeneity (Cochran’s Q test; $P < 0.001$) across GTEx tissues, which consisted of rs2945253-*FAM86B3P* and rs2905435-*GATA2DA* associations with schizophrenia, rs7184567-*STX4* and rs858239-*GPNMB* associations with Parkinson’s disease; rs55667375-*PRSS36*, rs7097656-*TSPAN14*, rs4295-*ACE* and rs56377155-*AC012146.1* associations with Alzheimer’s disease; and the rs229184-*SCFD1* association with amyotrophic lateral sclerosis (S8 Table). On inspection of the error bar plots (Fig 3), we found two of the instruments (rs7184567-*STX4* and rs7097656-*TSPAN14*) displayed a difference in direction of effect between certain brain tissues. rs7184567-*STX4* eQTL showed evidence ($P < 0.05$) of increased expression in the cerebellar hemisphere and cerebellum tissues and decreased expression in the amygdala, anterior cingulate cortex, cortex, frontal cortex, hypothalamus, nucleus accumbens, and putamen tissues. rs7097656-*TSPAN14* eQTL showed evidence ($P < 0.05$) of increased expression in the cerebellar hemisphere, cerebellum, and substantia nigra tissues, and decreased expression in the

anterior cingulate cortex, cortex and frontal cortex tissues. As an additional analysis, we then looked up our brain instruments in eQTLGen to determine if blood tissue would have been a reasonable proxy for the effects we observed. 42 of our brain instruments were found in the eQTLGen study. Correlation of the estimates between our brain study and eQTLGen was moderate ($r = 0.56$, 95% CI = [0.30,0.73]). 18 of the 42 instruments were genome-wide significant ($P < 5 \times 10^{-8}$) in eQTLGen and showed evidence of colocalization ($PP_4 > 70\%$) with the same gene's expression measured in our brain study (S9 Table). All 18 of the instruments apart from two, the rs229184-*SCFD1* and rs7184567-*STX4* associations, had the same direction of expression effect as our brain instrument. We also re-performed MR analysis instrumenting on eQTLs available within eQTLGen in order to establish whether additional genes could be reliably discovered using blood tissue as a proxy. A total of 41,157 independent SNPs across 16,190 genes were available as instruments at the $P < 5 \times 10^{-8}$ cut-off in eQTLGen. eQTLGen allowed for the instrumentation of many other genes with 10,754 of the 16,190 genes (66.4%) not intersecting with the genes we were able to instrument in the AMP-CMC MR analysis (S10 Table). 449 instruments ($n = 417$ genes) at the $P < 6 \times 10^{-7}$ and 30,245 instruments ($n = 11,818$ genes) at the $P < 0.05$ cut-off showed evidence of a MR effect from the 373,176 MR tests conducted in total across the 12 primary outcomes tested (full eQTLGen MR results available at GitHub repository denisbrd/braineQTLMR_datasets). We then matched the MR effect estimates for these eQTLGen instruments with the AMP-CMC MR estimates ($n = 7,689$ instruments across 7,136 genes) and found that a large proportion of the effects had discrepant directions: at the $P < 6 \times 10^{-7}$ cut-off, 50 from the 110 MR effects shared between the studies (45.5%) were in the opposite direction, and at the $P < 0.05$ cut-off 4,068 from 10,494 shared MR effects (38.9%) were in the opposite direction.

No evidence of reverse causation. As the genetic instruments were selected from a meta-analysis which contained post-mortem samples collected from the brain tissue of Alzheimer's disease patients, it is possible that some of the gene expression changes are a consequence of Alzheimer's disease pathogenesis rather than being cause of a neurological disease. We therefore examined whether our MR effects were in the correct direction using reverse MR and Steiger filtering. In reverse MR, we assessed the effect of the outcome on gene expression, using SNPs associated with the primary outcome at $P < 5 \times 10^{-8}$ as instruments. A total of 13,282 SNPs was retrieved from the primary outcome GWASs, which reduced to 180 independent SNP instruments after LD clumping. These consisted of 83 SNPs associated with schizophrenia, 28 with Alzheimer's disease, 27 with Parkinson's disease, 26 with multiple sclerosis, 6 with amyotrophic lateral sclerosis, 5 with major depressive disorder, 4 with bipolar disorder and one with anorexia nervosa. After harmonisation with SNP effects retrieved from eQTLGen, it was possible to conduct reverse MR for 41 of the 47 genes identified in the primary MR. Little evidence of MR relationships in the reverse direction (from outcome to gene expression) was detected for any of the genes (S11 Table). Steiger filtering was performed on the brain eQTL dataset from our study, and little evidence of reverse causation was detected from this analysis as well (S12 Table).

MR outcomes inform drug discovery

We next assessed whether the genes prioritized by MR might serve as promising targets for drug discovery. For this, we assessed each of the 47 genes causally linked to neurologic or psychiatric disease using the following three criteria: (i) evidence for an "allelic series" with coding variants reported to cause Mendelian diseases with neurological or behavioural phenotypes; (ii) evidence for being the target of a marketed or clinically tested drug; and (iii) a satisfying safety profile in our phenome-wide MR analyses. Of the 47 genes, five (*RERE*, *KCNQ5*, *SUOX*,

ACE and *GPNMB*) fulfilled our criteria as strong candidates for drug development (Tables S13 and 4). Our MR results suggest that inhibition of *RERE* and *KCNQ5* could be used to treat schizophrenia and *GPNMB* to treat Parkinson's disease whereas promotion of *SUOX* could be used to treat Anorexia and *ACE* to treat Alzheimer's disease. For instance, mutations in three genes our MR analyses linked to schizophrenia (*RERE*, *KCNQ5*) or anorexia (*SUOX*), respectively, are reported to cause monogenic diseases. Loss-of *RERE* function results in an early-onset neurodevelopmental disorder with anomalies in the brain, eye and heart (OMIM 605226) [19,20], and haploinsufficiency for *RERE* variants with assumed milder impact are believed to contribute to neurological abnormalities in adults [21]. Mutations in *KCNQ5* cause an autosomal-dominant mental retardation syndrome (OMIM 607357). *KCNQ5* is part of the transmembrane voltage-gated potassium channel gene family involved in the regulation of the conduction of electrical signals across nerve and muscle cells [22,23]. Lehman et al. identified both gain (increased nerve cell excitability) and loss of function (decreased nerve cell excitability) autosomal dominant mutations in the *KCNQ5* gene after conducting exome sequencing on families with intellectual disability and epilepsy [24]. *SUOX* encodes the sulphite oxidase enzyme responsible for breaking down sulfuric amino acids accumulation of which is neurotoxic. Autosomal-recessive deficiency for this enzyme results in Isolated Sulfite Oxidase Deficiency (ISOD) which presents as a neurodevelopmental disorder with motor and behavioural signs (OMIM 272300).

Three of the 47 genes had been considered earlier for drug development and are listed as targets for approved medicines or drugs tested in clinical trials. For instance, Dalfampridine is a *KCNQ5* inhibitor that has been approved to improve mobility in multiple sclerosis patients, and *KCNQ5* inhibitors [25] have been explored also for other neurological conditions including amyotrophic lateral sclerosis, Gullian-Barre Syndrome, unipolar depression and cerebral palsy (Open Target annotations). Similarly, *GPNMB* inhibitors have been explored for various types of cancer, for instance the monoclonal antibody glematubab (CDX-011) in Phase 2 trials for melanoma and breast cancer [26,27]. And *ACE* inhibitors are a well-established approved class of drugs for the treatment of hypertension, heart failure and diabetes mellitus.

We conducted an MR-PheWAS analysis to assess whether targeting the five genes that showed either an "allelic series" or were previously reported as drug targets would be safe (Table 4). Substantial new toxicity signals were identified for none of the five genes. However, consistent with the known pharmacological effect of *ACE* inhibitors to lower blood pressure, increased *ACE* expression colocalised with increased hypertension risk which might limit exploration of this target to treat or prevent Alzheimer's disease. Also, an increase in *SUOX* expression as it might be desirable to treat anorexia, colocalised with reduced hypothyroidism risk and additionally an elevated risk for rheumatoid arthritis and hypertension, warranting further safety investigations in follow up studies. Interestingly, expression of *RERE* also colocalized with heel bone mineral density, suggesting that in addition to schizophrenia *RERE* inhibitors (decreasing expression of *RERE*) might potentially be beneficial in osteoporosis (increasing heel bone BMD) and eventually other non-CNS traits, which will need to be clarified in future studies.

Discussion

In this study, we used a combination of two sample MR and colocalization analysis to infer potential causal relationships between expression of genes measured in brain tissues and 12 neurological and psychiatric conditions. MR was used to mitigate the confounding that may arise when assessing associations between gene expression and disease, and colocalization was used to verify that the eQTL instrument used in the MR was not coincidentally associated with

Table 4. Genes we prioritised as drug targets that had MR and colocalization evidence with other indications. Gene expression-disease relationship identified in the main MR (instrument-gene-indication): Alzheimer's disease (AD), Schizophrenia (Sz), anorexia (AN), suggested direction of drug action according to MR effect (inhibition or promotion of gene expression), other indications detected with the same gene in our MR-PheWAS (other indication), Wald ratio (WR), SE and P-value of MR effect and colocalization probability PP_4 (coloc).

instrument-gene-indication	drug mechanism	other indication	WR	SE	P	coloc
rs4295-ACE-AD	promotion (negative MR effect)	Non-cancer illness code self-reported: hypertension	0.024	0.003	3.57×10^{-15}	99.0
rs2708630-RERE-Sz	inhibition (positive MR effect)	Heel bone mineral density (BMD) T-score automated	-0.080	0.010	2.17×10^{-16}	97.1
		Diastolic blood pressure automated reading	0.054	0.008	2.15×10^{-12}	96.6
		Systolic blood pressure automated reading	0.050	0.008	1.58×10^{-10}	96.7
		Age at menarche	0.081	0.021	1.50×10^{-4}	97.8
rs1081975-SUOX-AN	promotion (negative MR effect)	Non-cancer illness code self-reported: hypothyroidism/myxoedema	-0.008	0.001	1.20×10^{-10}	87.0
		Rheumatoid arthritis	0.199	0.033	2.23×10^{-9}	97.9
		Non-cancer illness code self-reported: hypertension	0.011	0.003	3.06×10^{-5}	74.9

<https://doi.org/10.1371/journal.pgen.1009224.t004>

both traits (i.e. eliminate the possibility of the MR effect arising due to alternative causal variants in LD). This two-step approach to causal gene identification is important since expression of many genes with MR effects does not colocalize with association signals for the outcome trait. For instance, a previous two sample MR study [9] showed that many of the genes detected in their MR did not subsequently pass the HEIDI test (around 65%) used to distinguish between genuine pleiotropy and association through LD. In our current study, 47 eQTLs showed evidence of a shared genetic effect with neurological or psychiatric disease outcomes through MR and colocalization, suggesting a functional role for the genes in disease pathogenesis. 16 of these eQTLs had only a suggestive association ($5 \times 10^{-8} < P < 5 \times 10^{-5}$) in the respective outcome GWAS, which suggests that integrating eQTL evidence may help substantiate GWAS findings.

For eQTLs prioritized by MR, we investigated whether there was evidence for shared effects across the different neurological or psychiatric outcomes analysed, opening the possibility for identifying intervention points for novel drugs with cross-indication potential. No evidence for sharing of genetic effects at our multiple testing corrected P value threshold was detected. However, as we expected the genes we identified in our main MR to share outcome effects more often than a random gene we also looked for evidence of shared relationships at a less stringent threshold of $P < 0.05$, and found some weaker evidence of shared MR effects between schizophrenia, amyotrophic lateral sclerosis, and Parkinson's disease genes. *GRN* emerged as a particular promising target from our analyses since it was prioritised in our main MR analysis for Parkinson's disease, but also showed a suggestive relationship with Alzheimer's disease risk. *GRN* encodes the progranulin protein, that is highly active in brain cells and plays an important role in the survival of neurons. Autosomal dominant mutations in this gene have been found to result in frontotemporal dementia via haploinsufficiency, whereby the shortage of progranulin causes a build-up of TDP-43 aggregates in the brain resulting in premature cell death [28]. Mutations in this gene have also been identified in patients diagnosed with Alzheimer's disease [29]. A whole genome-wide sequencing study [30] discovered evidence of enrichment of rare deleterious variants in the *GRN* gene amongst Parkinson's disease cases relative to European controls.

We performed a systematic comparison of our instrument effects against other eQTL datasets in order to evaluate the generalisability of our MR findings. We found that a significant proportion of our instruments showed the inverse direction of effect when compared against blood tissue effects in eQTLGen and across a range of other non-brain related tissue effects in GTEx. We then performed the comparison on effect sizes for the instruments belonging to our top 47 gene-expression disease MR relationships. Heterogeneity tests revealed evidence of differing eQTL effect sizes for *FAM86B3P*, *STX4*, *PRSS36*, *GPNMB*, *TSPAN14*, *SCFD1*, *ACE*, *GATAD2A* and *AC012146.1* genes across the 13 brain tissues in GTEx. In particular, *STX4* and *TSPAN14* showed opposing directions of expression effects for certain tissues, which could result in drawing differing MR conclusions regarding direction of effect. Under half of the brain instruments for our top findings were also suitable instruments in eQTLGen, which indicates that disease relevant loci would have been missed if the study had relied on using a blood tissue as a proxy. In the MR analysis we conducted on eQTLGen instruments, we were able to detect many additional genes related to our diseases not picked up in our brain eQTL analysis, however the shared MR effects with our brain study tended to have a different causal direction. This argues that future MR studies should be conducted in disease relevant tissues to ensure the correct direction of causality. When there is uncertainty regarding the correct tissue, then it is important to examine instrument effects across a range of applicable tissues rather than relying on a single tissue. Studies which seek to enhance statistical power through combining eQTL effects across tissues, such as the MeCs meta-analysis method across brain tissues [10], will require the assumption of homogeneity of effect across tissues for all loci to be checked beforehand.

We examined the potential for our top MR findings to inform drug discovery and identified three genes (*RERE*, *KCNQ5*, *SUOX*) in which high-impact mutations are reported to cause Mendelian CNS diseases, suggestive of function-phenotype effects observed for “allelic series” genes [1]. Moreover, three genes (*ACE*, *GPMNB*, *KCNQ5*) have already been explored as drug targets in clinical trials or are the targets of marketed drugs, proposing existing therapies that could be explored for CNS conditions.

ACE inhibitors were found to have strong evidence of a causal relationship with increased Alzheimer’s disease risk in a previous MR study conducted to investigate the use of antihypertensive drugs to treat the disease [31]. However, mice models have demonstrated that increased expression of *ACE* helps to prevent cognitive decline and to protect against amyloid Beta-protein deposits [32], which if they build up in a brain are a risk factor for the development of Alzheimer’s disease. Therefore, we recommend that although we found that promotion of ACE could help to treat Alzheimer’s disease, due to the risks associated with the consequent increased blood pressure this action would require further functional studies to evaluate safety.

Our MR suggested that *GPNMB* inhibitors could be re-purposed to treat Parkinson’s disease. The role of *GPNMB* is already well-established in tumour progression, where over-expression of the gene suppresses the immune response to tumour growth [33]. Evidence also exists that supports a similar pathogenic role for this gene in Parkinson’s disease. Both *GPNMB* gene expression and protein levels have been shown to be elevated within brain regions of Parkinson’s disease patients, and functional follow-up in these same studies indicated that *GPNMB* may act through suppression of the inflammatory response [34,35]. Our *GPNMB* MR finding was consistent with a transcriptome-wide association study (TWAS) conducted by Li et al [36] on the Nalls et al [37] Parkinson’s disease GWAS, who also found that increased *GPNMB* expression in dorsolateral prefrontal cortex and peripheral monocytes was associated with increased disease risk.

MR further suggests that *KCNQ5* inhibitors could be re-purposed to treat schizophrenia. Potassium channel blockers are known to be effective in the treatment of multiple sclerosis [25], and several studies suggest that genes within this family would make strong candidates for the treatment of symptoms associated with schizophrenia [38]. For example, a SNP within the *KCNH7* gene was demonstrated to be involved in modulating the treatment response of schizophrenics to the antipsychotic drug risperidone, therefore the gene was recommended as a novel drug target by this study [39].

We decided to be relatively conservative with our cut-off used to select our instruments ($P < 5 \times 10^{-8}$) and selection of our top MR findings to avoid false positive MR findings. Although, two sample MR studies will frequently select SNPs at the $P < 5 \times 10^{-8}$ as instruments, less stringent thresholds have been used for eQTL studies. For example, to investigate brain tissue-specific gene expression effects on schizophrenia, a previous MR study [9] conducted on a brain tissue eQTL dataset ($n = 134$) [40] used a less stringent P value threshold of $P < 1.6 \times 10^{-3}$. Therefore, it is possible that we have missed reporting some genes that have a true causal effect. In contrast, for our colocalization analysis, we decided to use a relatively liberal probability cut-off of $PP_4 > 70\%$ to avoid wrongly eliminating potentially interesting and important loci (i.e. reduce false negative rate). For example, in the original method publication [41] considered a cut-off of $PP_4 > 75\%$ to be strong evidence of colocalization and other studies have used $PP_4 > 80\%$ as the cut-off to determine colocalization between traits [12,42]. 6 from 47 of our prioritised genes colocalized with a PP_4 of between 70% to 80%: *NMB*, *GOLGA2P7*, *FES* genes with schizophrenia, *KLHL7-DT* and *AC135050.3* genes with Parkinson's Disease, and the *AC012146.1* gene with Alzheimer's disease. These findings should therefore be considered as having more borderline evidence of colocalization. It is also possible that regions may fail to colocalize due to lack of statistical power and multiple causal variants in the region, which may obscure association peaks that would colocalize after conditional analysis is conducted across the region.

A limitation of our MR study, which is true of MR studies on QTLs in general, is the sparsity of SNP instruments available within most of the genomic regions. For the MR analysis, only a small fraction of genes could be instrumented by more than one SNP (of which the vast majority could only be instrumented by two SNPs), therefore the type of MR and sensitivity analysis that could be conducted was limited. Due to the lack of single SNP instruments within genes it was not possible to achieve substantially stronger instrumentation via the Inverse Variance Weighted (IVW) method (where Wald ratio effects are pooled across the same gene transcript as a weighted average to obtain an MR effect, full IVW results available at GitHub repository [denisbrd/braineQTLMR_datasets](https://github.com/denisbrd/braineQTLMR_datasets)) and to conduct many of the sensitivity tests to assess the impact of heterogeneity, invalid instruments and pleiotropy on the MR inference [43]. For example, for around half of our top findings' instruments, we detected shared association with other eQTLs for genes in the cis-region, making it difficult to prioritise the causal gene at this locus (i.e. imperfect drug target specificity) and also raising the possibility that these genes could be confounded via horizontal pleiotropy if genes acted through different molecular pathways. Reassuringly, we identified no such issues with the genes which we prioritised as drug targets as there was no evidence of shared association with the expression of other genes in their genomic region.

In conclusion, our study demonstrates the utility of using transcriptome-wide MR and colocalization analysis in brain tissue to identify drug targets for psychiatric and neurological diseases. However, it is important to note that our MR inference was based on outcomes which captured the incident risk of the disease, which relates to factors associated with onset of disease. As such targeting our findings with drugs may be useful in prevention of disease but it is not clear whether this may be effective in preventing progression of the disease as well.

GWASs are increasingly becoming available on disease progression traits which can be used in future MR studies to confirm whether susceptibility loci are also involved in disease progression and can be targeted by drugs to slow this process. Furthermore, as the SNP coverage and sample size of eQTL datasets available in brain tissue increases, this will allow for better instrumentation of existing genes and for more genes to be instrumented, improving the scope of MR analysis (to include trans-effects for example) that can be carried out and potential to discover further loci. Other brain QTLs (such as methylation, splicing, protein and cell-type QTLs) can also provide valuable information on underlying biology, therefore applying a multi-omic approach to future target identification efforts will be a highly fruitful avenue of MR research.

Materials and methods

Primary MR analysis: Two sample MR between brain eQTLs and neurological and psychiatric disease traits

Primary outcome selection. We selected 12 traits that were either neurological or psychiatric conditions as the primary outcomes (S14 Table) for this study. The 7 psychiatric traits included: anorexia nervosa [44], attention deficit disorder [45], autism spectrum disorder [46], bipolar disorder [47], major depressive disorder [48], obsessive compulsive disorder [49], and schizophrenia [50]. The 5 neurological traits included: Alzheimer's disease, amyotrophic lateral sclerosis [51], frontotemporal dementia [52], multiple sclerosis [53], and Parkinson's disease.

For Alzheimer's disease, we combined the summary statistics from the Lambert et al GWAS [54] (17008 cases, 37154 controls) with a GWAS by proxy of Alzheimer's disease family history in the UK Biobank (56773 proxy-cases and 379370 controls). Similarly, for Parkinson's disease, we combined summary statistics from Nalls et al GWAS [37] (13708 cases, 95282 controls) and a GWAS by proxy of Parkinson's disease family history in the UK Biobank (16909 proxy-cases, 388689 controls). For the UK Biobank analyses, we defined proxy-cases as participants who answered yes to whether they have a biological father (UK Biobank field 20107), mother (UK Biobank field 20110) or sibling (UK Biobank field 20111) who suffered from Alzheimer's disease/dementia (or Parkinson's disease). Participants who answered "Do not know" or "Do not wish to answer" were excluded from analyses. All other individuals were included as controls. We further excluded participants of non-European ancestry [55]. Genome-wide association analyses were performed using BOLT-LMM [56]. To enable meta-analysis with the previous GWAS, we multiplied the log(OR)-scale effect sizes and standard errors by a factor of two [57], and then performed meta-analysis using a fixed-effects inverse-variance weighted approach.

Instrument selection and preparation. We obtained our genetic instruments from a cis-eQTL meta-analysis performed on cohorts from AMP-CMC. This dataset consisted of a total of 6,937,060 SNPs across 19,281 gene transcripts measured in 1,286 pre-frontal brain cortex tissues. We removed 66,788 SNPs within the Major Histocompatibility (MHC) region (defined as SNPs between 24Mb and 36Mb in chromosome 6 hg19 build) from the dataset due to the extensive LD structure and complexity of this region. We then selected 1,054,129 genome-wide significant SNPs using $P < 5 \times 10^{-8}$ as a threshold and conducted LD clumping on this set of SNPs to obtain a set of independent eQTL instruments for each gene transcript. A total of 7,689 SNPs across 7,137 genes remained as instruments after clumping. LD clumping was performed using the `clump_data` function provided by the TwoSampleMR R package [43] on the default settings where all SNPs present in the 1000 Genomes EUR population with $r^2 > 0.001$ within a 10Mb window of the top hit were removed.

We standardised all eQTL SNP effects ($\beta_j \sim N(0,1)$) to represent a per SD change in gene expression to allow for comparison of the magnitude of the MR effect sizes across all genes for the different diseases. Standardised effect sizes and standard errors (SE) were derived using the following formulae [9,43]:

$$\beta_{standardised} = \frac{Z_j}{\sqrt{2 * MAF_j(1 - MAF_j) * (N_j + Z_j^2)}}$$

$$SE_{standardised} = \frac{1}{\sqrt{2 * MAF_j(1 - MAF_j) * (N_j + Z_j^2)}}$$

where the z score Z_j is obtained for SNP_j by dividing the raw SNP effect by the SE of the effect, and the variance of SNP_j genotypes is derived from the study's allele frequencies using $2 * MAF_j(1 - MAF_j)$, N_j is the sample size for the SNP (reported sample size for the study was used if SNP-specific sample size was not available). To determine instrument strength we calculated the F-statistic for each instrument using the Cragg-Donald statistic [58,59]:

$$PVE = \beta^2 / (\beta^2 + se^2 * n)$$

where PVE is the proportion of variance explained by the eQTL instrument, β and se is the effect size and standard error of the instrument, n is the sample size of the instrument ($n = 1286$ for our brain eQTL).

$$F \text{ statistic} = PVE * (n - 1 - k) / (1 - PVE * k)$$

where k is the number of instruments used in the MR estimate ($k = 1$ for single SNP MR) and n is the instrument sample size if $k = 1$ or average of the sample size over k instruments if $k > 1$.

Data harmonisation and two sample MR analysis. We extracted the selected instruments from the outcome GWASs, harmonised the genetic association between the instrument and outcome GWAS so that they reflected the same allele, and then conducted MR analyses on the harmonised data. We estimated the effect of gene expression on the primary outcomes using a single SNP instrument (Wald ratio method). All these steps were implemented using the TwoSampleMR R package [43] version 0.4 maintained by MR-Base.

Selection of top MR findings. A Bonferroni correction threshold of $P = 6 \times 10^{-7}$ at $\alpha = 0.05$ based on $n = 80,557$ MR tests was used to prioritise gene-outcome associations for further follow up. Some of the prioritised genes were within the same genomic region. For these cis-genes, in order to visualise the LD between instruments, regional association plots were generated in LocusZoom (locuszoom.org) [60] using 1000 Genomes EUR reference panel to compute the LD relative to the Wald ratio estimate for the gene with the lowest p-value in the region. LocusZoom plots are available for viewing on GitHub repository (denisbrd/brain-eQTLMR_datasets; DOI: [10.5281/zenodo.3778433](https://doi.org/10.5281/zenodo.3778433)).

Colocalization analysis between gene expression and traits

Colocalization analysis was conducted on the identified gene-outcome MR associations to confirm that the gene expression exposure and outcome shared the same causal variant within the region of interest (rather than the variant being shared coincidentally due to correlation through LD). Colocalization compares the probability of five hypotheses across a region: neither gene expression or the outcome are associated with genetic variants in the region (H_0), only gene expression is associated with a genetic variant in the region (H_1), only the outcome is associated with a genetic variant in the region (H_2), gene expression and outcome are both

associated with the region, but with different causal variants (H_3), gene expression and outcome are associated with the same causal variant (H_4).

We extracted the summary association statistics for all SNPs residing within a 1Mb region surrounding the instrument (i.e. instrument \pm 500kb) from the eQTL dataset, and then extracted the summary association statistics for the same SNPs from the outcome GWAS. We then used Bayesian coloc.abf function in R to conduct colocalization analyses on the SNPs common to both datasets [41]. Default priors were used: the expected proportion of SNPs associated with gene expression ($p_1 = 5 \times 10^{-4}$), associated with the outcome ($p_2 = 5 \times 10^{-4}$) and associated with both gene expression and the outcome ($p_3 = 5 \times 10^{-5}$). We considered the traits to colocalize if the posterior probability (PP_4) for H_4 was greater than 70%.

Instrument comparison between our study and other eQTL datasets

In order to determine the extent of similarity which would be observed from conducting the same MR analysis in other eQTL datasets, we compared the gene expression effect estimates for our instruments (from the AMP-CMC brain eQTL dataset) to the effect estimates for the same SNP in other eQTL datasets. In order to compare against brain tissue, we downloaded the full summary association statistics for the two brain eQTL dataset used in Qi publication [10] from their SMR website (<https://cnsgenomics.com/software/smr/#DataResource>): AMP, CMC and GTEx meta-analysis on brain cortex tissue ($n = 1194$) and the meta-analysis across ten different brain tissues available in GTEx ($n = 233$). In order to compare against blood tissue, we acquired the full summary statistics for eQTLGen [16] ($n = 31,684$) from their website (<https://www.eqtlgen.org>). In order to conduct a cross-tissue comparison, we downloaded the full association summary statistics available across all 48 tissues in GTEx version 7 (<https://gtexportal.org/home/>) [61]. We then extracted our instruments from each of these eQTL datasets, matching by both dbSNP rsid for the SNP and Ensembl ENSG gene identifier. We calculated the Pearson's correlation between the eQTL effect sizes in our study and the other studies across the instruments. We also determined the number of times our instruments had a gene expression effect which was significant at $P < 5 \times 10^{-8}$ (cut-off used for instruments in our study) and $P < 0.05$ level in the other eQTL datasets as well as the number of times our instruments disagreed with the direction of gene expression effect in other datasets (provided $P < 0.05$ for the gene expression effect in the other dataset).

Follow up MR analysis on selected instruments

The following analyses were conducted on the selected instruments identified in the main MR analysis (i.e. for the potentially causal genes which passed the Bonferroni correction).

Assessment of pleiotropy phenome-wide. We conducted a MR phenome-wide association analysis (MR-PheWAS) on 219 disease outcomes and risk factors, previously collated by Zheng et al [62] (S15 Table) to assess the potential for pleiotropic relationships across traits. Colocalization was performed on the MR effects which had $P < 0.05$

Using the MR-PheWAS dataset, we also examined whether the proportion of Wald ratio effects that had a shared effect ($P < 0.05$) in our primary MR could have been through random chance alone. We did this through permutation testing, in which we generated a distribution based on repeatedly randomly selecting ($n = 10,000$ iterations) the same number of non-brain related traits ($n = 12$ selected from 182 in total) from the MR-PheWAS dataset. In each iteration, we determined the number and proportion of effects which reached $P < 0.05$. From this distribution, we could then estimate the 95% empirical confidence interval around the proportion and p-value (number of times each iteration had a proportion that as least as extreme as our test statistic (proportion = $141/551 = 0.26$). We checked the distribution and it was

approximately Normal: descriptive parameters were as follows: number of effects at $P < 0.05$: min = 17, mean = 61.9, max = 140, proportion of effects at $P < 0.05$, min = 0.037, mean = 0.12, max = 0.27, empirical 95% CI for proportion = 0.12, [0.07,0.17].

Assessment of molecular pleiotropy. We assessed the number of different genes whose expression was associated with each of our instruments (i.e. molecular pleiotropy). We retrieved the summary association statistics across the gene transcripts available in the AMP-CMC study and determined the genes which were associated with the instrument at Bonferroni P -value $\alpha = 0.05$. We then conducted colocalization analysis between the gene expression and the outcome detected in our primary MR.

Assessment of tissue-specific effects. We assessed consistency between our selected instruments and brain tissue eQTLs from the Genotype-Tissue Expression (GTEx) project version 7 [61]. The 13 brain tissues available in GTEx were examined: Amygdala ($n = 88$), Anterior cingulate cortex BA24 ($n = 109$), Caudate basal ganglia ($n = 144$), Cerebellar Hemisphere ($n = 125$), Cerebellum ($n = 154$), Cortex ($n = 136$), Frontal Cortex BA9 ($n = 109$), Hippocampus ($n = 111$), Hypothalamus ($n = 108$), Nucleus accumbens basal ganglia ($n = 130$), Putamen basal gangli ($n = 111$), Spinal cord cervical c-1 ($n = 83$), and the Substantia nigra ($n = 80$). We extracted our selected instruments from these datasets, and we determined the number of GTEx eQTLs across the brain tissues that showed evidence of association (using a Bonferroni corrected $\alpha = 0.05$) and had the same direction of effect as our instruments (from AMP-CMC study). We also computed the Pearson's pairwise correlation between our gene expression effects and GTEx effects in each tissue. In order to assess whether eQTL effects showed statistical evidence of varying across different brain tissues, a meta-analysis was performed across the 13 brain GTEx tissues using the GWAMA meta-analysis package [63]. We defined evidence for heterogeneity across tissues as $P < 0.001$ based on Cochran's Q test. We also examined if using our selected instruments in blood tissue could proxy for brain tissue effects in an MR analysis. To do this summary association statistics were retrieved from eQTLGen (<https://www.eqtlgen.org>) [16]. We performed colocalization analysis between the gene expression effects measured in brain (AMP-CMC) and blood (eQTLGen) and for the regions which passed colocalization analysis ($PP_4 > 70\%$), we determined the number of blood effects which were genome-wide significant ($P < 5 \times 10^{-8}$) and had the same direction of effect as the brain instrument. We also computed the Pearson's pairwise correlation between brain and blood effects across our selected instruments.

Assessment of directionality of MR effect. A potential issue with the MR analysis could be that the gene expression changes are a consequence of genetic predisposition to disease rather than a cause of the disease. Directionality can be determined using either bi-directional MR [64] or Steiger filtering [65,66]. For a bi-directional MR, a reverse MR is conducted where the SNPs associated with the outcome are used as instruments, therefore evidence of a relationship would imply an incorrect direction of effect from the outcome to the exposure. The Steiger filtering method compares the effect estimates for the instrument-exposure association (ρ_{gx}) and instrument-outcome association (ρ_{gy}) in order to infer the direction of causality. The Steiger test computes a p -value (p_{steiger}) which is the probability of obtaining a difference between these two correlations ($\rho_{gx} - \rho_{gy}$) at least as extreme as the one observed in the study under the null hypothesis. The p_{steiger} statistic indicates the strength of evidence that the relationship is in the correct direction from exposure to outcome. Since we did not have access to full summary association statistics for in the AMP-CMC study (i.e. only captured cis-acting regions), it was not possible to conduct bi-directional MR in our brain tissue dataset, but we could perform the Steiger filtering test. Steiger filtering was conducted on our instruments using the `mr_steiger` function in the `TwoSampleMR` package, and those labelled as not having passed the `mr_steiger` filter reported if present. We performed bi-directional MR using

eQTLGen as the outcome dataset in the reverse MR. For the reverse MR step, we first defined instruments for outcomes by selecting genome-wide significant ($P < 5 \times 10^{-8}$) SNPs from our primary neurological and psychiatric GWASs, and performed LD clumping on the selected SNPs. We then looked up these SNPs in the genes available within eQTLGen, harmonised the effects and performed MR analyses to assess the effect of the neurological and psychiatric traits on gene expression using IVW linear regression, MR Egger and the weighted median estimator via the TwoSampleMR R package.

Evaluation of existing knowledge to assess potential drug targets

We carried out lookups in two public databases to assess the functional relevance and potential druggability of the genes identified in the main MR analysis: (i) Online Mendelian Inheritance in Man (OMIM) [67] to search for human monogenic disorders relating to the gene. Gene names were searched for in Online Mendelian Inheritance in Man (<https://omim.org>) online and the allelic variants section was manually reviewed for evidence of mutations which caused abnormal neurological or behavioural phenotypes and (ii) Open Targets [68] to find potential drug target-indication pairs for the gene. Ensembl ENSG gene name was searched in the Open Targets Genetics database online (<https://genetics.opentargets.org>) and clinical trials information was collected for genes which were annotated as being potentially druggable, which we defined as the target having entered the clinical trial stage and having at least completed Phase 0.

Supporting information

S1 Fig. Flow chart providing overview of instrument selection, MR analysis and drug target prioritization steps used in this study. Wald ratio (WR) and coloc prob (posterior probability of colocalization). Outcome GWAS datasets used in the MR analysis are reported in oval boxes and main findings reported in square boxes.
(PDF)

S1 Table. Summary association statistics comparison for instruments from our study (AMP-CMC) which had a different direction of gene expression effect to other eQTL datasets (Qi brain eQTL studies and across GTEx version 7 tissues).
(XLSX)

S2 Table. Correlation analysis results between gene expression effects for our instruments (AMP-CMC) and different tissue types present in GTEx version 7.
(XLSX)

S3 Table. Descriptive summary of the number of GWAS associations, instrumented genes and top MR findings across our primary outcomes.
(XLSX)

S4 Table. Harmonised SNP effects, Wald ratio statistics and colocalization results for the genes detected in the main MR analysis.
(XLSX)

S5 Table. MR and colocalization results for the prioritised genes found to be related with other non-brain related traits in our MR-PheWAS.
(XLSX)

S6 Table. eQTL effects and colocalization results for our prioritised gene instruments found to be associated with multiple cis-genes in the genomic region. Gene identified in the main MR analysis is highlighted in bold. Colocalization was also performed between the other

cis-gene(s) and the outcome identified in the main MR analysis for our prioritised gene.
(XLSX)

S7 Table. Number of different tissues in GTEx version 7 our selected instruments shared gene expression (eQTL) association with. Number of tissues in which gene expression effect was available in GTEx (tissue.count) for our instrument, number of tissues where the gene expression effect passed the Bonferroni correction threshold ($P < 9 \times 10^{-5}$), flag to indicate whether our instrument association was reproduced in GTEx tissue which we defined as Bonferroni significant tissue associations which had the same direction effect as our instrument (reproduce.tissue).
(XLSX)

S8 Table. Heterogeneity results from the meta-analysis of eQTL effects across the 13 brain tissues available in GTEx version 7. Analysis was conducted on our prioritised gene instruments.
(XLSX)

S9 Table. Summary association statistics from our comparison between our prioritised gene instruments and blood eQTLs from eQTLGen.
(XLSX)

S10 Table. Comparison between instruments derived in eQTLGen and AMP+CMC studies. AMPAD.shared (0/1 flag if gene is also instrumented in AMP+AD + CMC MR analysis) and AMP.AD.tophit.instrument (rsid of instrument used in AMP_AD + CMC, if multiple instruments available for the gene then rsid with the lowest p-value is reported).
(XLSX)

S11 Table. MR results from our reverse MR conducted between primary outcomes and gene expression in eQTLGen.
(XLSX)

S12 Table. Steiger filtering analysis results conducted on our prioritised gene instruments.
(XLSX)

S13 Table. Descriptive table detailing the findings from our drug target evidence follow up on our prioritised genes.
(XLSX)

S14 Table. Descriptive statistics of the studies which were selected as the primary outcomes for our MR analysis.
(XLSX)

S15 Table. Descriptive statistics for the risk factors and disease outcomes tested in our MR-PheWas.
(XLSX)

Acknowledgments

We gratefully acknowledge all studies and databases that have made their GWAS summary data available for this study: arcOGEN (Arthritis Research UK Osteoarthritis Genetics), BCAC (the Breast Cancer Association Consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDIoGRAM (Coronary ARtery DIsease Genome wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), DIAGRAM (DIabetes Genetics Replication And Meta-analysis), EAGLE (EARly Genetics and Lifecourse

Epidemiology Consortium), EAGLE Eczema (EARly Genetics and Lifecourse Epidemiology Eczema Consortium), EGG (Early Growth Genetics Consortium), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (GENetic Factors for OSteoporosis Consortium), GIANT (Genetic Investigation of ANthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics Consortium), GliomaScan (cohort-based genome-wide association study of glioma), GPC (Genetics of Personality Consortium), GUGC (Global Urate and Gout consortium), HaemGen (haematological and platelet traits genetics consortium), IGAP (International Genomics of Alzheimer's Project), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetic Consortium), ISGC (International Stroke Genetics Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MDACC (MD Anderson Cancer Center), MESA (Multi-Ethnic Study of Atherosclerosis), Neale's lab (a team of researchers from Dr Benjamin Neale's group, who made the UK Biobank GWAS summary statistics available), OCAC (Ovarian Cancer Association Consortium), IPSCSG (the International PSC study group), NHGRI-EBI GWAS catalog (National Human Genome Research Institute and European Bioinformatics Institute Catalog of published genome-wide association studies), PanScan (Pancreatic Cancer Cohort Consortium), PGC (Psychiatric Genomics Consortium), Project MinE consortium, ReproGen (Reproductive ageing Genetics consortium), SSGAC (Social Science Genetics Association Consortium), TAG (Tobacco and Genetics Consortium), TRICL (Transdisciplinary Research in Cancer of the Lung consortium) and UK Biobank. We would also like to gratefully acknowledge the members involved in the AMP-AD-eQTL Working Group: Devika Agarwal, David Airey, Mariet Allen, Sandeep Amberkar, Noam Beckmann, Rosa Canet-Aviles, John Calley, Minerva Carrasquillo, Kristen Dang, Jacob Degner, Philip De Jager, Derek Drake, Philip Ebert, James Eddy, Ayla Ergun, Nilufer Ertekin-Taner, Karol Estrada, Chris Gaiteri, Yushi Liu, Ben Logsdon, Winston Hide, Toby Johnson, Lara Mangravite, Jennifer Mollon, Sara Mostafavi, Thanner Perumal, Suzana Petanceska, Towfique Raj, Joseph Reddy, Janina Ried, James E Scherschel, Xue Wang and Aliza Wingo.

Author Contributions

Conceptualization: Denis A. Baird, Jimmy Z. Liu, Karol Estrada, Tom R. Gaunt.

Data curation: Denis A. Baird.

Formal analysis: Denis A. Baird.

Funding acquisition: George Davey Smith, Tom R. Gaunt.

Investigation: Denis A. Baird.

Methodology: Denis A. Baird, Jie Zheng, Benjamin Elsworth.

Project administration: Denis A. Baird, Tom R. Gaunt.

Resources: Solveig K. Sieberts, Thanneer Perumal, Benjamin Elsworth, Minerva M. Carrasquillo, Mariet Allen, Joseph S. Reddy, Philip L. De Jager, Nilufer Ertekin-Taner, Lara M. Mangravite, Ben Logsdon, Gibran Hemani.

Software: Benjamin Elsworth, Gibran Hemani.

Supervision: Jimmy Z. Liu, Karol Estrada, Heiko Runz, Tom R. Gaunt.

Visualization: Denis A. Baird.

Writing – original draft: Denis A. Baird.

Writing – review & editing: Denis A. Baird, Jimmy Z. Liu, Jie Zheng, Tom G. Richardson, Chia-Yen Chen, Karol Estrada, Philip C. Haycock, Gibran Hemani, Heiko Runz, George Davey Smith, Tom R. Gaunt.

References

1. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov.* 2013; 12: 581. <https://doi.org/10.1038/nrd4051> PMID: 23868113
2. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther.* 2014; 6: 37–37. <https://doi.org/10.1186/alzrt269> PMID: 25024750
3. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. *Science.* 2018; 360: eaap8757. <https://doi.org/10.1126/science.aap8757> PMID: 29930110
4. Horwitz T, Lam K, Chen Y, Xia Y, Liu C. A decade in psychiatric GWAS research. *Mol Psychiatry.* 2019; 24: 378–389. <https://doi.org/10.1038/s41380-018-0055-z> PMID: 29942042
5. Tan M-S, Jiang T, Tan L, Yu J-T. Genome-wide association studies in neurology. *Ann Transl Med.* 2014; 2: 124–124. <https://doi.org/10.3978/j.issn.2305-5839.2014.11.12> PMID: 25568877
6. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet.* 2015; 47: 856–860. <https://doi.org/10.1038/ng.3314> PMID: 26121088
7. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *bioRxiv.* 2019; 513945. <https://doi.org/10.1371/journal.pgen.1008489> PMID: 31830040
8. Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zheng Z, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun.* 2018; 9: 918. <https://doi.org/10.1038/s41467-018-03371-0> PMID: 29500431
9. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016; 48: 481–487. <https://doi.org/10.1038/ng.3538> PMID: 27019110
10. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun.* 2018; 9: 2282. <https://doi.org/10.1038/s41467-018-04558-1> PMID: 29891976
11. Richardson TG, Haycock PC, Zheng J, Timpson NJ, Gaunt TR, Davey Smith G, et al. Systematic Mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. *Hum Mol Genet.* 2018; 27: 3293–3304. <https://doi.org/10.1093/hmg/ddy210> PMID: 29893838
12. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurler MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet.* 2020; 52: 1122–1131. <https://doi.org/10.1038/s41588-020-0682-6> PMID: 32895551
13. Millard LAC, Davies NM, Timpson NJ, Tilling K, Flach PA, Smith GD. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. *Sci Rep.* 2015; 5: 16645. <https://doi.org/10.1038/srep16645> PMID: 26568383
14. Diogo D, Tian C, Franklin CS, Alanne-Kinnunen M, March M, Spencer CCA, et al. Phenome-wide association studies across large population cohorts support drug target validation. *Nat Commun.* 2018; 9: 4285. <https://doi.org/10.1038/s41467-018-06540-3> PMID: 30327483
15. Sieberts SK, Perumal T, Carrasquillo MM, Allen M, Reddy JS, Hoffman GE, et al. Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions. *bioRxiv.* 2019; 638544. <https://doi.org/10.1101/638544>
16. Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv.* 2018; 447367. <https://doi.org/10.1101/447367>
17. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet.* 2018; 27: R195–R208. <https://doi.org/10.1093/hmg/ddy163> PMID: 29771313
18. Paaby AB, Rockman MV. The many faces of pleiotropy. *Trends Genet TIG.* 2013; 29: 66–73. <https://doi.org/10.1016/j.tig.2012.10.010> PMID: 23140989

19. Jordan VK, Fregeau B, Ge X, Giordano J, Wapner RJ, Balci TB, et al. Genotype-phenotype correlations in individuals with pathogenic RERE variants. *Hum Mutat.* 2018; 39: 666–675. <https://doi.org/10.1002/humu.23400> PMID: 29330883
20. Bosch DGM, Boonstra FN, de Leeuw N, Pfundt R, Nillesen WM, de Ligt J, et al. Novel genetic causes for cerebral visual impairment. *Eur J Hum Genet EJHG.* 2016; 24: 660–665. <https://doi.org/10.1038/ejhg.2015.186> PMID: 26350515
21. Fregeau B, Kim BJ, Hernández-García A, Jordan VK, Cho MT, Schnur RE, et al. De Novo Mutations of RERE Cause a Genetic Syndrome with Features that Overlap Those Associated with Proximal 1p36 Deletions. *Am J Hum Genet.* 2016; 98: 963–970. <https://doi.org/10.1016/j.ajhg.2016.03.002> PMID: 27087320
22. Lerche C, Scherer CR, Seeböhm G, Derst C, Wei AD, Busch AE, et al. Molecular cloning and functional expression of KCNQ5, a potassium channel subunit that may contribute to neuronal M-current diversity. *J Biol Chem.* 2000; 275: 22395–22400. <https://doi.org/10.1074/jbc.M002378200> PMID: 10787416
23. Schroeder BC, Hechenberger M, Weinreich F, Kubisch C, Jentsch TJ. KCNQ5, a novel potassium channel broadly expressed in brain, mediates M-type currents. *J Biol Chem.* 2000; 275: 24089–24095. <https://doi.org/10.1074/jbc.M003245200> PMID: 10816588
24. Lehman A, Thouta S, Mancini GMS, Naidu S, van Slegtenhorst M, McWalter K, et al. Loss-of-Function and Gain-of-Function Mutations in KCNQ5 Cause Intellectual Disability or Epileptic Encephalopathy. *Am J Hum Genet.* 2017; 101: 65–74. <https://doi.org/10.1016/j.ajhg.2017.05.016> PMID: 28669405
25. Dunn J, Blight A. Dalfampridine: a brief review of its mechanism of action and efficacy as a treatment to improve walking in patients with multiple sclerosis. *Curr Med Res Opin.* 2011; 27: 1415–1423. <https://doi.org/10.1185/03007995.2011.583229> PMID: 21595605
26. Ott PA, Hamid O, Pavlick AC, Kluger H, Kim KB, Boasberg PD, et al. Phase I/II study of the antibody-drug conjugate glembatumumab vedotin in patients with advanced melanoma. *J Clin Oncol Off J Am Soc Clin Oncol.* 2014; 32: 3659–3666. <https://doi.org/10.1200/JCO.2013.54.8115> PMID: 25267741
27. Bendell J, Saleh M, Rose AAN, Siegel PM, Hart L, Sirpal S, et al. Phase I/II study of the antibody-drug conjugate glembatumumab vedotin in patients with locally advanced or metastatic breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol.* 2014; 32: 3619–3625. <https://doi.org/10.1200/JCO.2013.52.5683> PMID: 25267761
28. Baker M, Mackenzie IR, Pickering-Brown SM, Gass J, Rademakers R, Lindholm C, et al. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature.* 2006; 442: 916–919. <https://doi.org/10.1038/nature05016> PMID: 16862116
29. Piaceri I, Imperiale D, Ghidoni E, Atzori C, Bagnoli S, Ferrari C, et al. Novel GRN Mutations in Alzheimer's Disease and Frontotemporal Lobar Degeneration. *J Alzheimers Dis JAD.* 2018; 62: 1683–1689. <https://doi.org/10.3233/JAD-170989> PMID: 29614680
30. Ibanez L, Dube U, Davis AA, Fernandez MV, Budde J, Cooper B, et al. Pleiotropic Effects of Variants in Dementia Genes in Parkinson Disease. *Front Neurosci.* 2018; 12: 230. <https://doi.org/10.3389/fnins.2018.00230> PMID: 29692703
31. Walker VM, Kehoe PG, Martin RM, Davies NM. Repurposing antihypertensive drugs for the prevention of Alzheimer's disease: a Mendelian randomization study. *Int J Epidemiol.* 2019. <https://doi.org/10.1093/ije/dyz155> PMID: 31335937
32. Koronyo-Hamaoui M, Shah K, Koronyo Y, Bernstein E, Giani JF, Janjulia T, et al. ACE overexpression in myelomonocytic cells: effect on a mouse model of Alzheimer's disease. *Curr Hypertens Rep.* 2014; 16: 444–444. <https://doi.org/10.1007/s11906-014-0444-x> PMID: 24792094
33. Maric G, Rose AA, Annis MG, Siegel PM. Glycoprotein non-metastatic b (GPNMB): A metastatic mediator and emerging therapeutic target in cancer. *Oncotargets Ther.* 2013; 6: 839–852. <https://doi.org/10.2147/OTT.S44906> PMID: 23874106
34. Neal ML, Boyle AM, Budge KM, Safadi FF, Richardson JR. The glycoprotein GPNMB attenuates astrocyte inflammatory responses through the CD44 receptor. *J Neuroinflammation.* 2018; 15: 73. <https://doi.org/10.1186/s12974-018-1100-1> PMID: 29519253
35. Moloney EB, Moskites A, Ferrari EJ, Isacson O, Hallett PJ. The glycoprotein GPNMB is selectively elevated in the substantia nigra of Parkinson's disease patients and increases after lysosomal stress. *Neurobiol Dis.* 2018; 120: 1–11. <https://doi.org/10.1016/j.nbd.2018.08.013> PMID: 30149180
36. Li YI, Wong G, Humphrey J, Raj T. Prioritizing Parkinson's disease genes using population-scale transcriptomic data. *Nat Commun.* 2019; 10: 994–994. <https://doi.org/10.1038/s41467-019-08912-9> PMID: 30824768
37. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet.* 2014; 46: 989–993. <https://doi.org/10.1038/ng.3043> PMID: 25064009

38. Imbrici P, Camerino DC, Tricarico D. Major channels involved in neuropsychiatric disorders and therapeutic perspectives. *Front Genet.* 2013; 4: 76–76. <https://doi.org/10.3389/fgene.2013.00076> PMID: 23675382
39. Wang X, Su Y, Yan H, Huang Z, Huang Y, Yue W. Association Study of KCNH7 Polymorphisms and Individual Responses to Risperidone Treatment in Schizophrenia. *Front Psychiatry.* 2019; 10: 633. <https://doi.org/10.3389/fpsy.2019.00633> PMID: 31543842
40. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci.* 2014; 17: 1418–1428. <https://doi.org/10.1038/nn.3801> PMID: 25174004
41. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014; 10: e1004383. <https://doi.org/10.1371/journal.pgen.1004383> PMID: 24830394
42. Hatcher C, Relton CL, Gaunt TR, Richardson TG. Leveraging brain cortex-derived molecular data to elucidate epigenetic and transcriptomic drivers of complex traits and disease. *Transl Psychiatry.* 2019; 9: 105. <https://doi.org/10.1038/s41398-019-0437-2> PMID: 30820025
43. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife.* 2018; 7. <https://doi.org/10.7554/eLife.34408> PMID: 29846171
44. Duncan L, Yilmaz Z, Gaspar H, Walters R, Goldstein J, Anttila V, et al. Significant Locus and Metabolic Genetic Correlations Revealed in Genome-Wide Association Study of Anorexia Nervosa. *Am J Psychiatry.* 2017; 174: 850–858. <https://doi.org/10.1176/appi.ajp.2017.16121402> PMID: 28494655
45. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet.* 2019; 51: 63–75. <https://doi.org/10.1038/s41588-018-0269-7> PMID: 30478444
46. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism.* 2017; 8: 21. <https://doi.org/10.1186/s13229-017-0137-9> PMID: 28540026
47. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet.* 2011; 43: 977–983. <https://doi.org/10.1038/ng.943> PMID: 21926972
48. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet.* 2018; 50: 668–681. <https://doi.org/10.1038/s41588-018-0090-3> PMID: 29700475
49. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry.* 2018; 23: 1181–1188. <https://doi.org/10.1038/mp.2017.154> PMID: 28761083
50. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014; 511: 421–427. <https://doi.org/10.1038/nature13595> PMID: 25056061
51. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron.* 2018; 97: 1268–1283.e6. <https://doi.org/10.1016/j.neuron.2018.02.027> PMID: 29566793
52. Ferrari R, Hernandez DG, Nalls MA, Rohrer JD, Ramasamy A, Kwok JBJ, et al. Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol.* 2014; 13: 686–699. [https://doi.org/10.1016/S1474-4422\(14\)70065-1](https://doi.org/10.1016/S1474-4422(14)70065-1) PMID: 24943344
53. Sawcer S, Hellenthal G, Pirinen M, Spencer CCA, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* 2011; 476: 214–219. <https://doi.org/10.1038/nature10251> PMID: 21833088
54. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013; 45: 1452–1458. <https://doi.org/10.1038/ng.2802> PMID: 24162737
55. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018; 562: 203–209. <https://doi.org/10.1038/s41586-018-0579-z> PMID: 30305743
56. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet.* 2015; 47: 284–290. <https://doi.org/10.1038/ng.3190> PMID: 25642633

57. Liu JZ, Erlich Y, Pickrell JK. Case-control association mapping by proxy using family history of disease. *Nat Genet.* 2017; 49: 325–331. <https://doi.org/10.1038/ng.3766> PMID: 28092683
58. Zeng P, Wang T, Zheng J, Zhou X. Causal association of type 2 diabetes with amyotrophic lateral sclerosis: new evidence from Mendelian randomization using GWAS summary statistics. *BMC Med.* 2019; 17: 225. <https://doi.org/10.1186/s12916-019-1448-9> PMID: 31796040
59. Cragg JG, Donald SG. Testing Identifiability and Specification in Instrumental Variable Models. *Econom Theory.* 1993; 9: 222–240.
60. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinforma Oxf Engl.* 2010; 26: 2336–2337. <https://doi.org/10.1093/bioinformatics/btq419> PMID: 20634204
61. GTEx Consortium, Aguet F, Brown AA, Castel SE, Davis JR, He Y, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017; 550: 204. <https://doi.org/10.1038/nature24277> PMID: 29022597
62. Zheng J, Haberland V, Baird D, Walker V, Haycock P, Gutteridge A, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *bioRxiv.* 2019; 627398. <https://doi.org/10.1101/627398>
63. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics.* 2010; 11: 288. <https://doi.org/10.1186/1471-2105-11-288> PMID: 20509871
64. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014; 23: R89–R98. <https://doi.org/10.1093/hmg/ddu328> PMID: 25064373
65. Hemani G, Tilling K, Davey Smith G. Correction: Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* 2017; 13: e1007149. <https://doi.org/10.1371/journal.pgen.1007149> PMID: 29287073
66. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* 2017; 13: e1007081. <https://doi.org/10.1371/journal.pgen.1007081> PMID: 29149188
67. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005; 33: D514–D517. <https://doi.org/10.1093/nar/gki033> PMID: 15608251
68. Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, et al. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* 2017; 45: D985–D994. <https://doi.org/10.1093/nar/gkw1055> PMID: 27899665