Multi-omics analyses of cognitive traits and psychiatric disorders highlight brain-dependent mechanisms

Roxanna Korologou-Linden[1,2], Genevieve M. Leyden[1,2,3], Caroline L. Relton[1,2], Rebecca C. Richmond[1,2] and Tom G. Richardson[1,2,*]

[1]MRC Integrative Epidemiology Unit at the University of Bristol, University of Bristol, Bristol BS8 2BN, UK
[2]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK
[3]Bristol Medical School, Translational Health Sciences, Dorothy Hodgkin Building, University of Bristol, Bristol BS1 3NY, UK
*To whom correspondence should be addressed at: MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. Tel: +44(0)1173313370; Email: tom.g.richardson@bristol.ac.uk

Abstract

Integrating findings from genome-wide association studies with molecular datasets can help develop insight into the underlying functional mechanisms responsible for trait-associated genetic variants. We have applied the principles of Mendelian randomization to investigate whether brain-derived gene expression ($n = 1194$) may be responsible for mediating the effect of genetic variants on eight cognitive and psychological outcomes (attention-deficit hyperactivity disorder, Alzheimer's disease, bipolar disorder, depression, intelligence, insomnia, neuroticism and schizophrenia). Transcriptome-wide analyses identified 83 genes associated with at least one outcome ($P_{Bonferroni} < 6.72 \times 10^{-6}$), with multiple trait colocalization also implicating changes to brain-derived DNA methylation at nine of these loci. Comparing effects between outcomes identified the evidence of enrichment, which may reflect putative causal relationships, such as an inverse relationship between genetic liability towards schizophrenia risk and cognitive ability in later life. Repeating these analyses in whole blood ($n = 31\,684$), we replicated 58.2% of brain-derived effects (based on $P < 0.05$). Finally, we undertook phenome-wide evaluations at associated loci to investigate pleiotropic effects with 700 complex traits. This highlighted pleiotropic loci such as *FURIN* [initially implicated in schizophrenia risk ($P = 1.05 \times 10^{-7}$)], which had evidence of an effect on 28 other outcomes, as well as genes which may have a more specific role in disease pathogenesis [e.g. *SLC12A5* which only provided evidence of an effect on depression ($P = 7.13 \times 10^{-10}$)]. Our results support the utility of whole blood as a valuable proxy for future studies analysing molecular datasets, but also suggest that conducting analyses in a tissue-specific manner may be more comprehensive.

Introduction

Genome-wide association studies (GWAS) have discovered and replicated thousands of variants associated with cognitive and neurological traits and disease (1,2). However, findings from GWAS on their own do not provide insight into the biological processes that mediate the effect of these genetic variants. Developing our understanding of these molecular mechanisms is crucial in terms of disease prevention and treating disorders which affect normal brain function.

The majority of known trait-associated variants reside in the non-protein-coding regions of the human genome, with previous research implicating them in transcriptional regulatory mechanisms (3,4). This includes altering promoter and enhancer elements as well as enrichment in regions of the genome affecting gene regulation. With developments in microarray and sequencing technologies, data on genome-wide genotyping and gene expression from large samples are becoming increasingly available and have been used to identify genetic variants which robustly influence transcription (5,6). These variants are known as expression quantitative trait loci (eQTL) and have been shown to operate in a tissue-dependent manner (7). The GWAS variants responsible for cognitive and psychiatric trait variations are likely to exert their influence via gene regulation in the brain, given

this organ's role in the pathogenesis of these phenotypes (8). As such, using eQTL derived from brain tissue should be the most pertinent tissue type in terms of characterizing these effects. Furthermore, there are various cognitive traits and psychiatric disorders, which have been shown to be correlated either genetically or phenotypically [e.g. intelligence with ADHD (9) and Alzheimer's disease (10,11) as well as neuroticism with schizophrenia, bipolar disorder and depression (12)]. As such, exploring whether these traits share a common eQTL derived from the brain tissue may highlight shared genetic architecture underlying such correlations.

We previously integrated tissue and cell-type specific eQTL data with findings from GWAS to investigate putative molecular mechanisms using the principles of Mendelian randomization (MR) (13,14). As with conventional risk factors, gene expression may be prone to confounding and reverse causation, which MR is robust to, by assessing whether a genetically predicted exposure has an effect on a complex trait or disease outcome. This approach is complemented by techniques in genetic colocalization to mitigate the likelihood that detected associations are driven by linkage disequilibrium between separate, but correlated, causal variants (i.e. one responsible for gene expression and the other for

complex trait variation). As postulated previously, genetic colocalization at conditionally independent loci is necessary but not sufficient for causality (15).

In this study, we have applied a transcriptome-wide MR framework to discern whether the brain-derived gene expression may mediate the effect of genetic variants on eight different cognitive and psychiatric outcomes [attention-deficit hyperactivity disorder (ADHD), Alzheimer's disease, bipolar disorder, depression, intelligence, insomnia, neuroticism and schizophrenia]. We subsequently applied the same analysis using eQTL data from a large collection of transcriptomes derived from whole blood (based on findings from the eQTLGen project). This was to elucidate brain tissue-dependent signals which would not have been detected using whole blood eQTL despite the large increase in sample size (brain eQTL, $n = 1194$, compared with whole blood eQTL, $n = 31\,684$). Next, we integrated methylation quantitative trait loci (mQTL) data derived from brain tissue to discern whether there is evidence at the associated loci that DNA methylation may also reside along the causal pathway to cognitive or psychiatric outcomes along with the gene expression. Finally, to investigate pleiotropic effects, we have undertaken phenome-wide analyses by systematically applying MR to 700 different complex traits and disease.

Results

Uncovering transcriptome-wide signatures in brain tissue which putatively mediate genetic effects onto outcomes

We applied the principles of MR to identify genetically predicted effects between brain-derived gene expression and eight cognitive and psychological outcomes, using *cis*-acting single nucleotide polymorphisms (SNPs) (i.e. within 1 Mb distance of their associated gene) as instruments obtained from brain-derived eQTL datasets. There were 7443 genes whose expression could be instrumented by at least 1 *cis*-eQTL ($P < 5 \times 10^{-8}$) using the brain tissue. Analysing each of these genes with the eight cognitive and psychological outcomes identified 91 genetically predicted effects (across 83 genes with several genes linked with more than one outcome) which survived a $P < 6.72 \times 10^{-6}$ (i.e. 0.05/7443) as well as a genetic colocalization posterior probability of association (PPA) > 0.8 for single SNP analyses using the 'coloc' method (16) (Supplementary Material, Table S3). These included established loci in psychiatric genetics, such as *NEGR1* with depression ($P = 2.15 \times 10^{-9}$) as well as *RERE* and *FURIN* with schizophrenia ($P = 4.43 \times 10^{-6}$ and $P = 1.05 \times 10^{-7}$, respectively). There were also a host of loci which have not been linked as strongly to neuropsychiatric disorders in the literature. These included solute carrier family 12 member 5 (*SLC12A5*) ($P = 7.13 \times 10^{-10}$ with depression), TRIO binding protein (*TRIOBP*) ($P = 3.59 \times 10^{-8}$ with intelligence), *STX1B* and solute carrier family 25 member 12 (*SLC25A12*; both of which had a putative effect on neuroticism with $P = 1.26 \times 10^{-8}$ and $P = 6.42 \times 10^{-6}$, respectively). We include a table of full MR results (i.e. without thresholding for genetic colocalization) in the Supplementary Material, Table S4. We found that 38.7% of these results were supported by the evidence of colocalization. Performing a sensitivity analysis using an intelligence GWAS (i.e. based on fluid intelligence score) not adjusting for socioeconomic status (SES) produced similar results to those of Savage *et al.* (17) (Supplementary Material, Table S5). The results from our study were comparable to those detected by previous transcriptome-wide association studies. For example, similar to publications by Hall *et al.* (18) and Gusev *et al.* (3), we

identified genes such as chloride channel 3 (*CLCN3*), *GATAD2A* and *ELAC2* but also identified novel genes not highlighted by these studies such as *NEURL*, *TRIM37* and *SDCCAG8* using MR and genetic colocalization.

Subsequently, to assess the potential shared aetiology between the eight traits, we investigated how the genetically predicted effects of identified genes on all eight outcomes clustered based on Euclidean distance matrix computation (19). We did this by applying the R package 'dist' to compare the distances between our MR estimates for the same gene across the eight cognitive traits and psychiatric disorders. These results were therefore used to highlight the local genetic correlation which exists between the 83 loci identified in our analysis across each of the outcomes evaluated. A heatmap visualizing these results can be found in Figure 1.

This analysis highlighted clusters of genes whose shared effects amongst outcomes which may reflect causal relationships. For example, we observed evidence of enrichment for an inverse relationship between schizophrenia-associated genes and intelligence (Fig. 1), which may reflect the evidence postulating a causal relationship between schizophrenia genetic liability and lower cognitive ability in later life as presented by previous MR analyses (20). Likewise, there was enrichment of an inverse relationship between ADHD-associated genes and intelligence (Fig. 1), which has also been reported by a previous MR study (21). Therefore, these loci contain candidate genes where vertical pleiotropy may exist, such that their effect on schizophrenia liability through changes to gene expression may consequently have an influence on the measures of cognition.

Enrichment analyses provided evidence that the genetically predicted expression of the 83 identified genes were enriched for various regions of the brain compared with a background set of loci (Supplementary Material, Table S7). For example, the predicted expression of neuroticism-associated genes was enriched in the hippocampus ($P = 8.61 \times 10^{-7}$), amygdala ($P = 3.61 \times 10^{-5}$) and substantia nigra ($P = 4.36 \times 10^{-5}$). Genes associated with intelligence were enriched in the frontal cortex ($P = 0.0001$), cerebellum ($P = 0.0005$) and anterior cingulate cortex ($P = 0.0004$) (Supplementary Material, Fig. S3). There was also enrichment for the genes associated with bipolar disease in the hypothalamus ($P = 0.0005$), whereas the expression of insomnia-associated genes were over-represented in the cerebellum ($P = 0.0006$).

Discerning whether effects are brain tissue-dependent or identifiable using whole blood data

Applying our MR framework using whole blood eQTL data ($n = 31\,684$) identified 260 genetically predicted effects (across 234 loci), which survived multiple testing corrections [$P < 3.27 \times 10^{-6}$ based on a maximum of 15 301 genes (Alzheimer's disease) with at least one instrument in this tissue] (Supplementary Material, Table S8). Comparing effect estimates from our brain tissue analysis found that 53 of the 91 genetically predicted effects had a $P < 0.05$ in blood (Supplementary Material, Table S9). Therefore, 41.8% of these putative effects may have been overlooked by not using brain tissue based on this heuristic. This included several genes which appear to be predominantly expressed in brain tissue based on findings from the GTEx consortium, such as N-ethylmaleimide sensitive factor (*NSF*), protocadherin alpha 8 (*PCDHA8*) and *SLC12A5* (Supplementary Material, Figs S1–S3, respectively). Figure 2 provides an illustration of findings for intelligence as an outcome, where four loci [*PSME4*, kinesin family
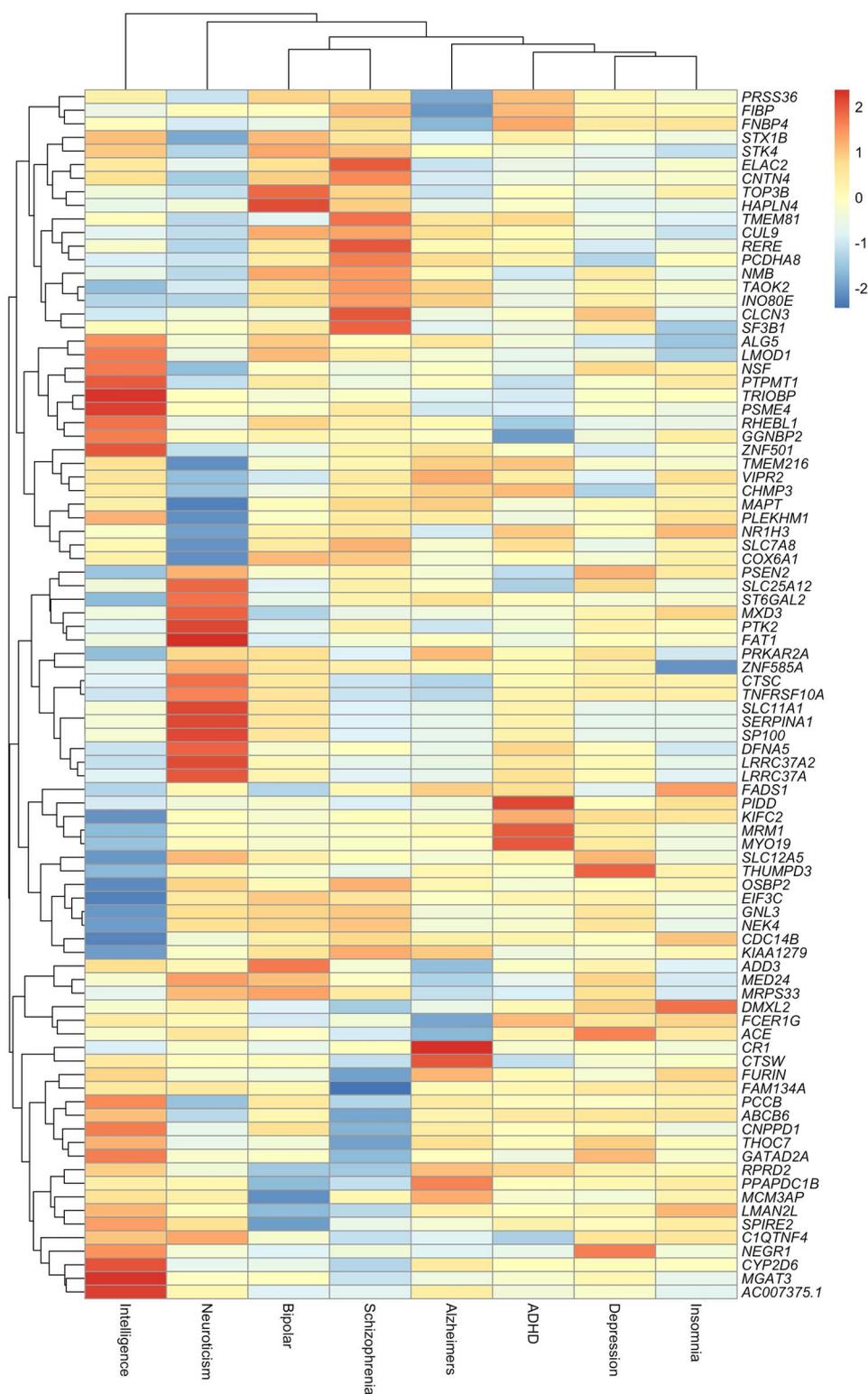
Figure 1. A heatmap illustrating MR effect estimates between the brain-derived gene expression for 83 genes and eight neurological and psychiatric traits and disease.

member C2 (*KIFC2*), *NSF* and *TRIOBP*] provided much stronger evidence of a genetically predicted effect using brain tissue than in whole blood despite the large difference in sample sizes ($n = 1194$ in brain tissue and $n = 31\,684$ in whole blood). Conversely, 37 (14%; i.e. 37/260) genetically predicted effects from the whole

blood analysis had an effect resulting in $P < 0.05$ in brain tissue, although these did not survive transcriptome-wide corrections (Supplementary Material, Table S10). However, this may be due to having lower power in the brain tissue-based analyses. As such, they may be worthwhile candidates for future studies to prioritize.
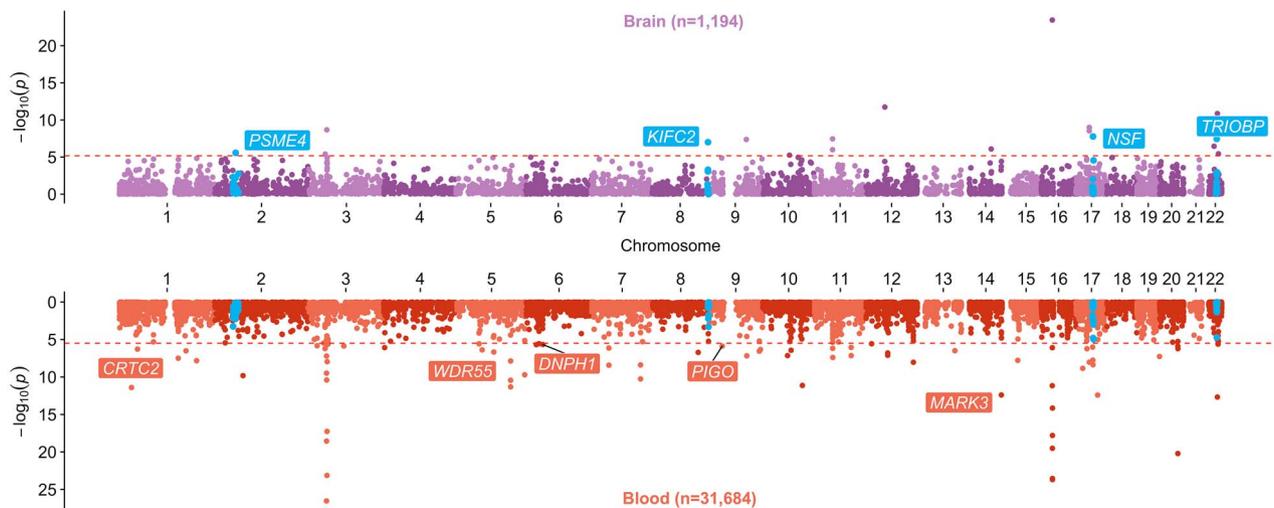
Figure 2. A bi-directional Manhattan plot depicting the transcriptome-wide effects for gene expression on intelligence. The top plot represents effects derived from brain tissue (purple, $n = 1194$), whereas the bottom plot illustrates the same analysis using whole blood data (red, $n = 31\,684$). Genes surviving multiple testing ($P < 6.72 \times 10^{-6}$ in brain, $P < 3.27 \times 10^{-6}$ in blood) based on single SNP analyses were additionally subjected to genetic colocalization (PPA > 0.8). Loci highlighted in blue survived multiple testing in brain-derived tissue but not in whole blood (despite the large difference in sample size). Loci which are labelled in red on the inverse plot represent candidates that were not identified in our brain tissue analysis but may survive multiple testing corrections in larger samples.
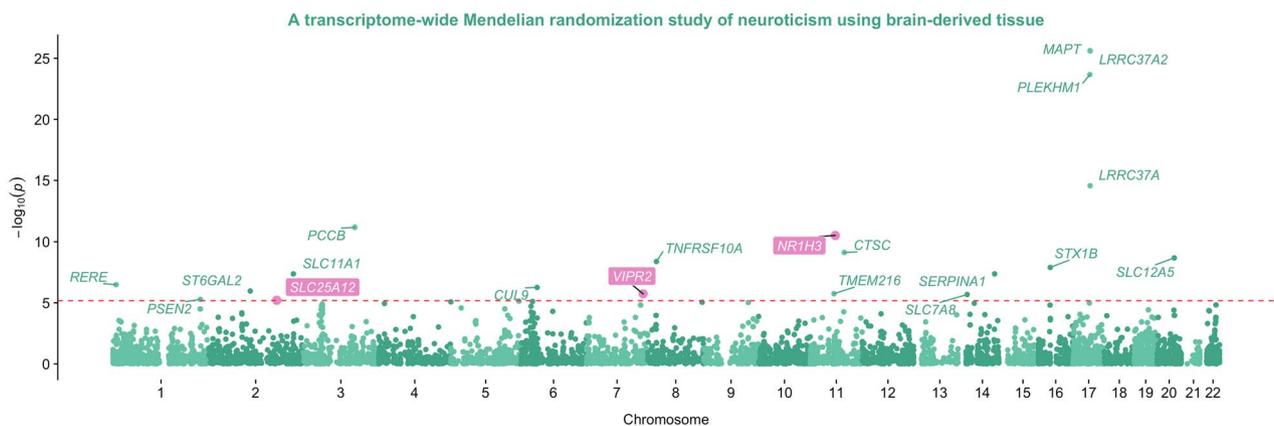
Figure 3. A Manhattan plot representing the transcriptome-wide associations between brain-derived expression and neuroticism. Genes surviving multiple testing ($P < 3.27 \times 10^{-6}$) based on single SNP analyses were additionally subjected to genetic colocalization (PPA > 0.8). Points highlighted in pink provided evidence from multiple trait colocalization analyses that brain-derived DNA methylation, as well as gene expression, share a causal variant at these loci with neuroticism susceptibility.

Elucidating epigenetic mechanisms which may also play a role in trait variation and disease susceptibility

For effects identified in the initial analyses using brain-derived eQTL data, we conducted follow-up analyses to distinguish whether DNA methylation levels are also implicated in the risk of the examined traits/disease outcomes. We undertook multiple trait colocalization using the R package 'moloc' (16) at each of the 83 loci identified in the brain tissue analysis to discern whether DNA methylation, gene expression and complex traits all shared a common causal variant. There was evidence that this was true at nine loci based on brain-derived DNA methylation data with a PPA > 0.8 (Supplementary Material, Table S11). Figure 3 depicts the results from the transcriptome-wide analysis of neuroticism, where highlighted loci provided evidence of a shared causal variant for DNA methylation, gene expression and neuroticism risk. These were *SLC25A12* (highest PPA = 0.96), *VIPR2* (highest PPA = 0.96) and *NR1H3* (highest PPA = 0.99).

A phenome-wide evaluation of genes to highlight pleiotropic effects

MR and genetic colocalization analyses were undertaken systematically on 700 outcomes using findings from large-scale GWAS consortia and the UK Biobank study (Supplementary Material, Table S12). Undertaking MR analyses at each of the 83 brain tissue genes and 700 complex traits from across the phenome identified 52 genes which had a genetically predicted effect on more than one outcome (Supplementary Material, Tables S13 and S14). There were established psychiatric loci which appear to be highly pleiotropic, such as *FURIN* (genetically predicted to influence 28 traits, Fig. 4A) and *MAPT* (predicted effects on 33 traits) based on the phenome-wide corrections of $P < 7.14 \times 10^{-5}$. These findings therefore suggest that they should be deprioritized as therapeutic targets for functional follow-up studies, as inhibiting their expression may result in unanticipated adverse effects.

In contrast, our analyses did not flag the potential adverse effects for various loci in this analysis. This included genes whose

genetically predicted expression was associated with schizophrenia, such as *CLCN3* ($P = 5.99 \times 10^{-7}$) and *ABCB6* ($P = 9.48 \times 10^{-7}$) and *SLC12A5*, which was associated with depression ($P = 7.30 \times 10^{-10}$, Fig. 4B), serine protease 36 (*PRSS36*) which was associated with Alzheimer's disease ($P = 7.30 \times 10^{-8}$) and neuromedin B (*NMB*) which was associated with bipolar disease ($P = 1.70 \times 10^{-6}$, Fig. 4C).

Discussion

We have applied an integrative approach using brain-derived molecular datasets to highlight biologically relevant genes which may putatively influence risk of eight neurological and psychiatric outcomes. This analysis provided evidence suggesting that genetically predicted gene expression in brain tissue may influence at least one of these traits at 83 genetic loci. Integrating DNA methylation data obtained from brain samples provided further evidence that epigenetic factors and gene expression may play a role in trait variation at nine of these loci. Furthermore, we repeated our integrative framework using gene expression data derived from whole blood to examine whether these putative effects would have been detected using this proxy tissue type. In our study, a large proportion of brain-derived associations (∼60%) survived a heuristic threshold of $P < 0.05$ in blood, corroborating the use of blood as a starting point for target identification. Conversely, despite the dramatically lower sample size for brain tissue (less than 30×), there were ∼40% of the genetically predicted brain tissue effects that were not detected in blood based on this threshold. This may suggest that these effects depend strongly on using data from a tissue-type which is biologically relevant to the outcomes assessed. However, it is important to consider that despite brain being the primary tissue of interest in this study, other tissues may be more pertinent in some instances. For example, lung and spleen tissues have previously been shown to be enriched amongst Alzheimer's disease loci (17). Finally, we investigated potential pleiotropic effects at these loci by assessing their association with 700 outcomes from large-scale GWAS and the UK Biobank study.

Tissue- and cell-dependent gene expressions have been previously shown to help elucidate the biological processes and functions of genes (22,23). There are various genes highlighted in our study which only provided evidence of a genetically predicted effect on outcomes when leveraging brain-derived gene expression. For instance, the genetically predicted effects between *KIFC2*, *NSF* and *TRIOBP* with intelligence appeared to depend on using expression data from the brain tissue, despite the roughly 30-fold increase in the sample size with whole blood. Furthermore, we identified other genes that provided much stronger evidence of association using brain tissue in comparison with whole blood. This included genes such as *SLC12A5*, which showed evidence of a genetically predicted effect on depression and neuroticism ($P = 7.13 \times 10^{-10}$ and $P = 2.10 \times 10^{-9}$ respectively), as well as *PCDHA8*, which provided evidence of genetically predicted effect on schizophrenia risk ($P = 6.04 \times 10^{-6}$). All these loci have been shown to have a role in neuronal maintenance and neurotransmitter release (see details in Table 1).

We also found evidence suggesting that both gene expression and DNA methylation may be involved along the causal pathway between the genetic variant and trait variation at nine of the 83 associated loci. These findings build upon previous research that has harnessed QTL datasets from the brain tissue (24,25), suggesting a coordinated system of effects between the regulatory elements (e.g. transcription factors) and gene transcription, which is consistent with causality. These loci included *SLC25A12*, where

a PPA of 0.96 suggested that genetic variation here influences proximal DNA methylation, *SLC25A12* expression and neuroticism risk. This gene encodes a calcium-binding solute carrier in the inner mitochondrial membrane, which is essential for energy homeostasis (26). Polymorphisms in the *SLC25A12* gene have been shown to strongly associate with autism in candidate gene studies (27–30). Furthermore, its expression has been associated with neurite outgrowth and is upregulated in the prefrontal cortex of autistic individuals (31). Although we did not study autism, neuroticism and autism are genetically (32) and phenotypically correlated (33,34).

Phenome-wide investigations of the genes identified using brain tissue provide indications of possible pleiotropic effects, which may suggest that therapeutically targetting them could be detrimental to other health outcomes. For example, *FURIN*, whose expression was linked with schizophrenia risk ($P = 1.05 \times 10^{-7}$), was also associated with traits such as high blood pressure and diastolic blood pressure but with the opposite direction of effect. In contrast, we did not find evidence of potential adverse effects for potentially novel targets, such as *SLC12A5* (depression, $P = 7.13 \times 10^{-10}$), *NMB* (bipolar disorder, $P = 1.70 \times 10^{-6}$) and *PRSS36* (Alzheimer's disease, $P = 7.30 \times 10^{-8}$), deeming them potentially promising and worthwhile druggable loci. Moreover, we identified evidence to support previously implicated targets, such as *CLCN3* (schizophrenia, $P = 5.99 \times 10^{-7}$), which have been reported to play a role in the control of glutamatergic transmission (35). As with all analyses in this study, however, these genes are merely examples to highlight potentially translatable opportunities from our findings (more examples are provided in Supplementary Material).

A previous study by Qi *et al*. (36) reported that the genetic effects at the top *cis*-eQTLs and mQTLs are highly correlated between independent brain and blood samples ($r^2 = 0.7$ and $r^2 = 0.78$, respectively), supporting the use of whole blood as a proxy for gene discovery efforts in brain-related traits. We integrated whole blood-derived expression data from eQTLGen with the GWAS summary statistics of examined traits to prioritize potential candidates which we may have been underpowered to detect in brain tissue. This analysis identified 37 candidate loci for future follow-up once larger samples of brain tissue are available. Whilst these findings support the utility of whole blood as proxy for more pertinent tissue-types, our analyses also uncovered evidence suggesting that various genetically predicted effects may still be overlooked by not using brain tissue data (∼40%). This therefore supports future endeavours aiming to derive molecular datasets using brain tissue in larger sample sizes. Subsequently, this will reduce false positives rates to which smaller sample sizes may be prone to as well as validate findings such as those in our study. That said, whole blood may still be a useful proxy when sufficient samples of brain tissue data are not accessible. Genes identified in this study such as *SLC12A5* and *NSF* are such examples where whole blood appeared to be a valid proxy in our study, given that findings from GTEx suggest that they are expressed much more in brain tissue compared with whole blood (Supplementary Material, Figs S1 and S2).

In terms of future work, we performed our analysis instrumenting gene expression with *cis*-variants as *trans*-effects may be more prone to horizontal pleiotropy. The small variance explained by individual *trans*-eQTL/mQTL necessitates larger sample sizes for their identification and use in MR analyses (37). Furthermore, the lack of full summary data for the mQTL data poses a limitation for the multiple trait colocalization analysis. Our findings therefore prioritize these loci for future evaluations by epigenetic

Table 1. Function and previous literature on genes mentioned in the discussion

Gene name	Putative biological function and previous transcriptome-wide studies	Current study
CLCN3	Studies (18,35) using expression data in adult brain tissue from the dorsolateral prefrontal cortex of the CMC (35) showed that genetic variants encoding *CLCN3* are upregulated by schizophrenia risk allele. *CLCN3* is a brain-expressed chloride ion channel, implicated in controlling fast excitatory glutamatergic transmission (65). $CLCN3^{-/-}$ mice show neurological symptoms such as blindness, motor coordination deficits and spontaneous hyperlocomotion as well progressive degeneration of the retina, hippocampus and ileal mucosa (66).	DNA methylation and expression of *CLCN3* associated with a higher risk of schizophrenia (β: 0.28, se = 0.06, $P = 5.99 \times 10^{-7}$ in brain tissue)
FURIN (paired basic amino acid cleaving enzyme)	Previous eQTL studies have shown the expression of *FURIN* to be associated with the risk of schizophrenia (67,68). One of these studies (68) integrated genetic associations from schizophrenia GWAS and brain eQTL from 193 normal human subjects, using different approaches such as Sherlock, summary MR, DAPPLE, Prix Fixe and NetWAS. The other study (67) performed this analysis using the summary MR package, only using brain tissue samples from the CMC (N = 467). A study showed that altering expression of *FURIN* changed neurodevelopment in zebrafish, and knockdown of *FURIN* in human progenitor cells resulted in abnormal migration (35). *FURIN* processes precursor proteins to mature forms, including brain-derived neurotrophic factor (*BDNF*), whose downregulation has been associated with a higher risk of schizophrenia (69).	Expression of *FURIN* associated with risk of schizophrenia (β: −0.25, se = 0.05, $P = 1.05 \times 10^{-7}$ in brain tissue)
KIFC2	*KIFC2* was implicated in genetically predicted intelligence through eQTL mapping in two different studies (17,70). The *KIFC* gene encodes C-terminal kinesin-related motor protein that has been found to be specifically expressed in adult neurons in mice and has been shown to be involved in microtubule-dependent retrograde axonal transport in dendrites (71).	Expression of *KIFC2* associated with intelligence (β: −0.05, se = 0.01, $P = 9.93 \times 10^{-8}$ in brain tissue)
NMB	*NMB* encodes a family of bombesin-like peptides which are mainly expressed in the hypothalamus, stomach and colon (72). A study using summary MR integrating their GWAS summary statistics (73) findings with eQTL data from the dorsolateral prefrontal cortex (35) and whole blood (6), showed that genetic variants encoding *NMB* are upregulated by the bipolar risk allele.	Expression of *NMB* associated with bipolar disease (β: 0.30, se = 0.06, $P = 1.70 \times 10^{-6}$ in brain tissue)
NSF	The most recent genome-wide meta-analysis for intelligence (17) implicated *NSF* in predicting intelligence using methods such as SNP-based genome-wide association test, gene-based genome-wide association test (using aggregate effect of all SNPs in a gene) and eQTL mapping. The *NSF* gene encodes ATPase N-ethylmaleimide-sensitive factor, which plays an important role in vesicle fusion events and in regulating neurotransmitter release kinetics (74).	Expression of *NSF* associated with intelligence (β: 0.05, se = 0.01, $P = 1.70 \times 10^{-8}$ in brain tissue)
PCDHA8	A study (16) which applied multiple trait colocalization to identify regulatory effects at GWAS risk loci for schizophrenia identified a shared causal variant affecting the risk of schizophrenia through DNA methylation and expression of *PCDHA8*. This study used methylation data from postmortem tissue of the dorsolateral prefrontal cortex of non-psychiatric control donors (N = 121), gene expression data from dorsolateral prefrontal cortex of patients in a case-control study collected by the CMC sample (48) and the schizophrenia GWAS by Ripke *et al.* (75). *PCDHA8* was associated with schizophrenia risk which is a member of the protocadherin family of genes that are involved in the establishment and maintenance of neuronal connections in the brain (76).	Expression of *PCDHA8* associated with the risk of schizophrenia (β: 0.11, se = 0.02, $P = 6.04 \times 10^{-6}$ in brain tissue)
PRSS36 (serine protease 36)	*PRSS36* has been implicated via eQTL association in the hippocampus, which is highly affected early in Alzheimer's disease. Furthermore, a summary-based tissue-specific transcriptome-wide association study (77) using a factor polygenic QTL analysis, jointly modelling gene expression across tissues and individuals, found brain-specific expression of *PRSS36* to be associated with Alzheimer's disease.	Expression of *PRSS36* associated with risk of Alzheimer's disease (β: −0.02, se = 0.003, $P = 7.30 \times 10^{-8}$ in brain tissue)

Continued

studies to determine whether a disruption to transcription binding sites or chromatin compaction has downstream implications for neurological trait variation. For ADHD, we may have lacked the power to identify many disease-associated loci due to the small

sample of the corresponding GWAS. Due to modest sample sizes for current molecular datasets, largely attributed to the cost of sequencing arrays, we are unable to robustly delineate the causal sequence of events between molecular traits and outcomes. For

Table 1. Continued

Gene name	Putative biological function and previous transcriptome-wide studies	Current study
SLC12A5	A previous integrative analysis using a different brain eQTL and mQTL dataset found that the expression of the *SLC12A5* colocalized the risk of neuroticism (78). Furthermore, it has been reported to increase the risk of other brain-related disorders, such as epilepsy (79). Changes in *SLC12A5* expression were not identified as conferring risk to MDD in a Sherlock analysis (80), integrating the genetic associations from a MDD GWAS (42) and brain eQTL dataset (46). The *SLC12A5* gene encodes a neuron-specific transmembrane cotransporter (81), which plays a role in mediating fast synaptic inhibition.	Depression (β: 0.03, se = 0.004, $P = 7.13 \times 10^{-10}$). Neuroticism (β: 0.04, se = 0.007, $P = 2.10 \times 10^{-9}$ in brain tissue)
SLC25A12	None of the integrative analyses (78,82) have identified methylation or expression of *SLC25A12* to affect risk of neuroticism. *SLC25A12* encodes a calcium-binding solute carrier in the inner mitochondrial membrane that is essential for energy homeostasis (26).	DNA methylation and expression of *SLC25A12* associated with neuroticism score (β: 0.06, se = 0.01, $P = 6.42 \times 10^{-6}$ in brain tissue)
TRIOBP-1	A GWAS study (17) which performed functional annotation of top hits for intelligence identified a SNP nearest to the *TRIOBP* gene. Furthermore, *TRIOBP* contained three exonic non-synonymous variants associated with intelligence. *TRIOBP* has also been identified in gene-based association analyses of brain volumetric phenotypes (83). *TRIOBP* encodes proteins with a role in neural tissue development and controlling cytoskeleton organization, cell motility and cell growth (84).	Expression associated with intelligence (β: 0.03, se = 0.01, $P = 3.59 \times 10^{-8}$ in brain tissue)

Abbreviations: MDD, major depressive disorder; qTL, quantitative trait locus.

example, it is plausible in our study that the outcome may have an effect on DNA methylation and/or gene regulation rather than the converse direction of effect as evaluated in our framework.

Furthermore, although the integration of various biological datasets in many situations can help prioritize genes responsible for disease risk, it is difficult to isolate the causal gene(s) due to the mechanistic co-regulation of genes, where the expression of two or more genes is correlated due to their regulation by the same eQTL (i.e. mechanistic co-regulation). Moreover, co-regulation may imply that the genes are involved in the same pathway, making it plausible that multiple genes play a causal role. Although genetic colocalization analyses can mitigate the likelihood of false positives due to linkage structure at the loci, functional studies are required to definitively rule out horizontal pleiotropy as a potential explanation for our results. Conducting phenome-wide association studies in other tissues may be useful in identifying pleiotropic signals not confined to brain tissue. Likewise, applying our framework using tissue-specific protein QTL should prove useful in terms of identifying evidence potentially overlooked by eQTL (38). Lastly, there are limitations in using MR, such as the presence of non-transmitted genetic effects [3, 4] known to violate the independence assumption of MR [5]. Future developments in family studies may aid in disentangling transmitted environmental effects from those of genetic origin.

Our findings emphasize the importance of leveraging molecular datasets derived from biologically relevant tissues to develop insight into the causal pathway for trait-associated variants. Our results provide a prioritized list of candidate genes which future studies can investigate to better characterize the biological mechanisms which influence the neurological traits and risk of psychiatric disorders.

Materials and Methods
Data sources
GWAS summary data

We obtained the beta coefficient (log odds ratio), standard error, effect allele, alternate allele, effect allele frequency and sample size from the summary statistics of eight independent GWAS for the following brain-related diseases/traits: Alzheimer's disease ($N_{cases} = 71\,880$), ADHD ($N_{cases} = 20\,183$) (39), bipolar disorder ($N_{cases} = 20\,129$) (40), depression ($N_{cases} = 170\,756$) (41,42), insomnia ($N_{cases} = 208\,716$) (43), intelligence ($N = 279\,930$) (17), neuroticism (score) ($N = 374\,317$) (44) and schizophrenia ($N_{cases} = 33\,426$) (40). As the GWAS summary statistics for intelligence by Savage *et al.* (17) have been conditioned on socioeconomic status (SES), we also include the effect estimates of identified genes from a GWAS of fluid intelligence score without adjustment for SES in Supplementary Material. Details on GWAS datasets are described in Supplementary Material, Table S1.

Brain tissue-derived quantitative trait loci data

We obtained eQTL summary data from a meta-analysis ($n_{eff} = 1194$) of GTEx brain ($n = \sim 233$) (45), Common Mind Consortium (CMC) ($n = 467$) (35) and Religious Orders Study and Memory Aging Project (ROSMAP) ($n = 494$) (46) undertaken by Qi *et al.* Only SNPs within 1 Mb distance from each probe ($n = 28\,538$) were available (cis-QTLs). Additionally, from Qi *et al.* (36), we extracted mQTL summary statistics from a meta-analysis of three brain-derived datasets: brain cortical region from ROSMAP study ($n_{ind} = 468$, $n_{probe} = 420\,103$, $n_{snp} = 5$ million) (36); fetal brain ($n_{ind} = 166$, $n_{probe} = 26\,840$, $n_{snp} = 0.3$ million) (47) and frontal cortex region ($n_{ind} = 526$, $n_{probe} = 138\,917$, $n_{snp} = 1.5$ million) (48) (Supplementary Material, Table S2). DNAm levels in all these studies were based on the Illumina Human Methylation 450 K array. In ROSMAP, only SNPs within 5 kb of each DNA methylation probe were available. In the Hannon *et al.* data, only SNPs within 500 kb distance from each probe and with $P < 1.0 \times 10^{-10}$ were available. In the Jaffe *et al.* data, only SNPs within 20 Kb distance from each probe and FDR < 0.1 were available. In the CMC sample, 209 participants had a diagnosis of schizophrenia, whilst 288 in ROSMAP had a diagnosis of Alzheimer's disease.

Whole blood-derived eQTL data

We obtained eQTL data from a large-scale meta-analysis (6) in non-transformed peripheral blood samples ($N = 31\,684$) from
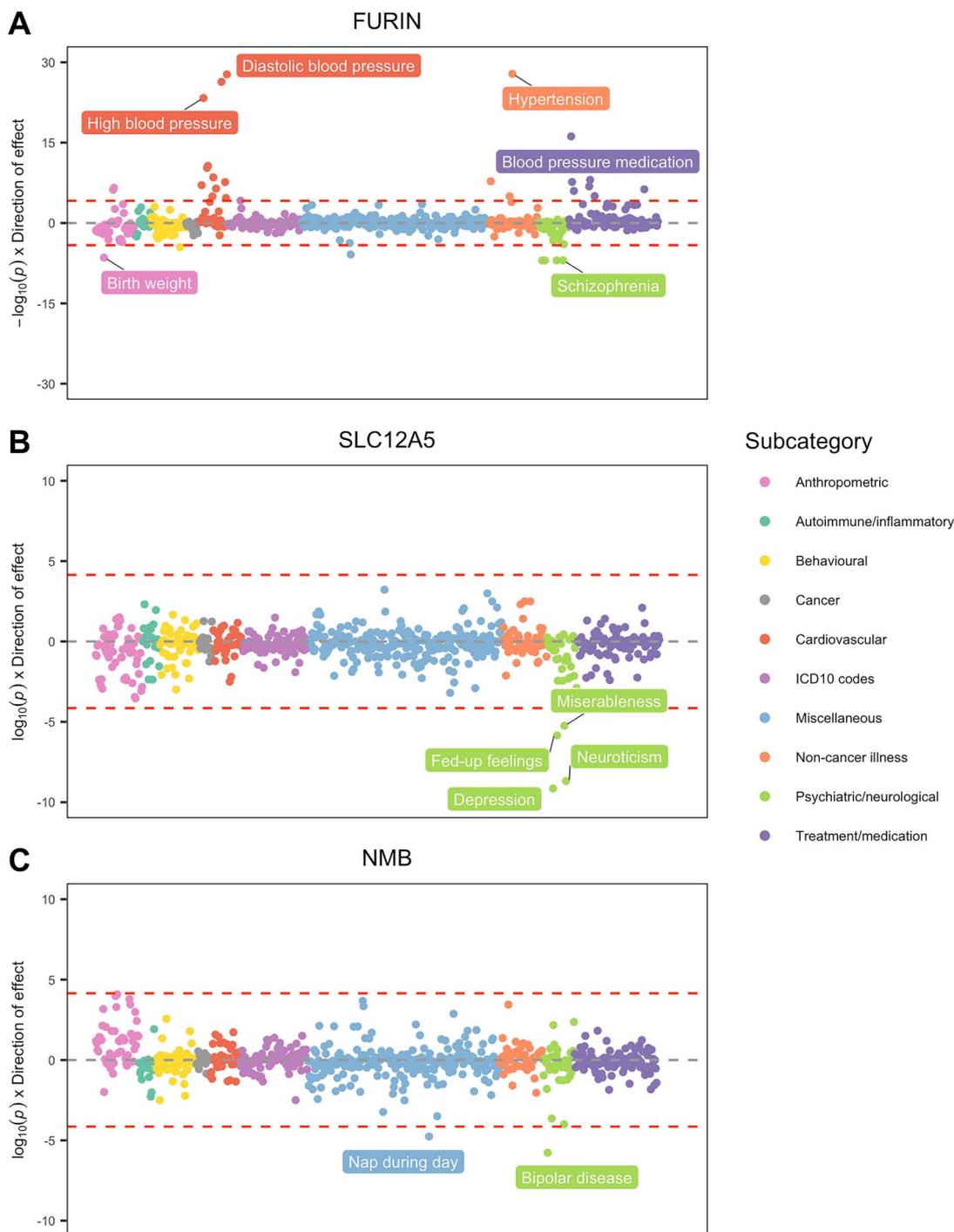
Figure 4. Bi-directional phenome-wide association plots depicting MR results between brain-derived gene expression and over 700 different traits and outcomes for (**A**) *FURIN*, (**B**) *SLC12A5* and (**C**) *NMB*. Points are coloured and clustered based on the subcategories of traits.

37 cohorts (Supplementary Material, Table S2). Data from the eQTLGen consortium include association summary statistics for 10 023 016 SNPs and 19 251 genes (resource accessed on 18 October 2018). *Cis*-eQTLs were defined based on a distance less than 1 Mb from their associated genes probe. Several of these cohorts included participants with cardiovascular disease or/and psychiatric disorders (Supplementary Material, Table S2). Further details are available at https://www.eqtlgen.org/.

Statistical analysis
Brain-derived gene expression analysis

Linkage disequilibrium clumping was undertaken with PLINK using $r^2 < 0.01$ and based on a reference panel of European individuals from the 1000 genomes phase 3 project (49). When we were only able to instrument a gene's expression using a single independent eQTL, the Wald ratio method was used to estimate the MR effects (50). When two or more independent eQTL were

available for MR analyses, we used the inverse variance weighted (IVW) method (51). Single SNP MR analyses may be prone to higher false discovery rates due to the linkage disequilibrium between the causal eQTL at a locus and a separate but correlated variant responsible for changes in trait variation. As such, results which survived multiple testing comparisons based on single SNP analyses were evaluated using genetic colocalization to discern whether the effects were attributed to the same underlying variant. This was assessed using the 'coloc' R package (52) where a PPA > 0.8 was used as the evidence of colocalization. Effects which survived multiple testing (Bonferroni-corrected threshold of $P < 0.05$/number of tests) were carried through to subsequent downstream analyses. We also compared the genes identified in our study for schizophrenia to those identified in two transcriptome-wide association studies (3,18) (Supplementary Material, Table S6).

Detecting shared aetiology between traits and brain region enrichment analysis

To examine potential shared aetiology between the eight traits, we investigated how the genetically predicted effects of identified genes for these traits clustered based on Euclidean distance matrix computation using the R package 'dist' (53). A heatmap to visualize these clusters was generated using the package 'pheatmap' (54). Enrichment analyses for clusters of genes associated with the same trait were evaluated using the MAGMA approach (55), which was undertaken as part of the FUMA platform (56). Genes of interest were tested against 19 283 protein-coding gene sets obtained from MsigDB (i.e. hallmark gene, sets, positional gene sets, curated gene sets, motif gene sets, computational gene sets, GO gene sets, oncogenic signatures and immunologic signatures) and WikiPathways, using hypergeometric tests (56). Using data from GTEx v7 (45), we assessed whether there was evidence of gene expression enrichment clustering in specific regions of the brain.

A comparison between brain-derived gene expression effects and those identified using whole blood

We repeated our analysis pipeline using eQTL derived from eQTLGen in whole blood to compare the effect estimates between brain and blood-derived analyses. This allowed to highlight effects which were much stronger in brain-derived tissue, particularly given the discrepancy in sample sizes ($n = 1194$ in brain and $n = 31\,684$ in blood). A threshold of $P < 0.05$ was used as a heuristic in this analysis to highlight associations with a strong evidence of differential expression in either brain or blood.

Multiple trait colocalization analysis using brain-derived DNA methylation effects

We used multiple trait colocalization to discern whether there was evidence that DNA methylation, gene expression and cognitive or psychological trait of interest are all influenced by the same causal variant at each loci investigated. A PPA > 0.8 was used as evidence of genetic colocalization, this time using the 'moloc' R package (16). Analyses were run multiple times to investigate the DNA methylation at CpG sites within 100 kb distance of gene coordinates. For loci where there were two or more independent eQTLs based on linkage disequilibrium clumping, we evaluated each signal in turn by conditioning out separate effects using GCTA-COJO (57).

Phenome-wide analysis

We undertook hypothesis-free analyses of genes identified in our primary analysis using brain eQTL data to elucidate pleiotropic effects. A Bonferroni-corrected threshold of $P < 7.14 \times 10^{-5}$ (i.e. 0.05/700 traits) was used as a heuristic to highlight putative pleiotropic effects for target genes, provided they were druggable based on previous evidence as described in the study by Leyden *et al.* (58). Briefly, this comprised of a curated list obtained from four data-driven drug discovery endeavours, which evaluated whether the product of protein-coding genes could be altered using targeted therapeutics (59–62). These findings were derived to flag the possible adverse side effects as well as the multi-purposing potential of therapeutic interventions of these genes.

Supplementary Material

Supplementary Material is available at *HMG* online.

Data Availability

All GWAS datasets on primary outcomes analyzed in this study are available publicly as described in Supplementary Material, Table S1. Analyses were conducted using R (version 3.3.1) unless stated otherwise. MR analyses were undertaken using the 'TwoSampleMR' package (63) where all the outcomes analyzed in our phenome-wide association analyses are available. Plots were generated using the 'ggplot2' package (64).

Acknowledgements

Funding

References

1. Hindorff, L.A., Sethupathy, P., Ramos, E.M., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci.*, **106**, 9362–9367.

2. Buniello, A., Macarthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E. *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D101.

3. Gusev, A., Mancuso, N., Won, H., Kousi, M., Finucane, H.K., Reshef, Y., Song, L., Safi, A., McCarroll, S., Neale, B.M. *et al.* (2018) Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat. Genet.*, **50**, 538–548.

4. Li, S., Luo, X.-J., Liu, J., Huo, Y. and Li, X. (2019) Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nat. Commun.*, **10**, 670.

5. Lloyd-Jones, L.R., Holloway, A., McRae, A., Yang, J., Small, K., Zhao, J., Zeng, B., Bakshi, A., Metspalu, A., Dermitzakis, M. *et al.* (2017) The genetic architecture of gene expression in peripheral blood. *Am. J. Hum. Genet.*, **100**, 228–237.

6. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E. *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, **45**, 1238–1243.

7. Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Arcelus, M.G., Sekowska, M. *et al.* (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science (80-)*, **325**, 1246–1250.

8. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L. *et al.* (2018) Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.*, **50**, 381–389.

9. Frazier, T.W., Demaree, H.A. and Youngstrom, E.A. (2004) Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, **18**, 543–555.

10. Anderson, E.L., Howe, L.D., Wade, K.H., Ben-Shlomo, Y., Hill, W.D., Deary, I.J., Sanderson, E.C., Zheng, J., Korologou-Linden, R., Stergiakouli, E. *et al.* (2020) Education, intelligence and Alzheimer's disease: evidence from a multivariable two-sample Mendelian randomization study. *Int. J. Epidemiol.*, **49**, 1163–1172.

11. Snowdon, D.A., Kemper, S.J., Mortimer, J.A., Greiner, L.H., Wekstein, D.R. and Markesbery, W.R. (1996) Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. *JAMA*, **275**, 528–532.

12. Okbay, A., Baselmans, B.M.L., De Neve, J.E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F.W., Linnér, R.K., Rietveld, C.A., Derringer, J. *et al.* (2016) Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.*, **48**, 624–633.

13. McGowan, L.M., Davey Smith, G., Gaunt, T.R. and Richardson, T.G. (2019) Integrating Mendelian randomization and multiple-trait colocalization to uncover cell-specific inflammatory drivers of autoimmune and atopic disease. *Hum. Mol. Genet.*, **28**, 3293–3300.

14. Taylor, K., Davey Smith, G., Relton, C.L., Gaunt, T.R. and Richardson, T.G. (2019) Prioritizing putative influential genes in cardiovascular disease susceptibility by applying tissue-specific Mendelian randomization. *Genome Med.*, **11**, 6.

15. Richardson, T.G., Haycock, P.C., Zheng, J., Timpson, N.J., Gaunt, T.R., Smith, G.D., Relton, C.L. and Hemani, G. (2018) Systematic Mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. *Hum. Mol. Genet.*, **27**, 3293–3304.

16. Giambartolomei, C., Liu, J.Z., Zhang, W., Hauberg, M., Shi, H., Boocock, J., Pickrell, J., Jaffe, A.E., Pasaniuc, B. and Roussos, P. (2018) A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics*, **34**, 2538–2545.

17. Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I. *et al.* (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.*, **50**, 912–919.

18. Hall, L.S., Medway, C.W., Pain, O., Pardiñas, A.F., Rees, E.G., Escott-Price, V., Pocklington, A., Bray, N.J., Holmans, P.A., Walters, J.T.R. *et al.* (2019) A transcriptome-wide association study implicates specific pre- and post-synaptic abnormalities in schizophrenia. *Hum. Mol. Genet.*, **29**, 159–167.

19. Danielsson, P.E. (1980) Euclidean distance mapping. *Comput. Graph. Image Process.*, **14**, 227–248.

20. Richardson, T.G., Harrison, S., Hemani, G. and Smith, G.D. (2019) An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife*, **8**, e43657.

21. Stergiakouli, E., Martin, J., Hamshere, M.L., Heron, J., St Pourcain, B., Timpson, N.J., Thapar, A. and Davey Smith, G. (2017) Association between polygenic risk scores for attention-deficit hyperactivity disorder and educational and cognitive outcomes in the general population. *Int. J. Epidemiol.*, **46**, dyw216.

22. Ramsköld, D., Wang, E.T., Burge, C.B. and Sandberg, R. (2009) An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.*, **5**, e1000598.

23. Kundu, K., Mann, A.L., Tardaguila, M., Watt, S., Ponstingl, H., Vasquez, L., Morrell, N.W., Stegle, O., Pastinen, T., Sawcer, S.J. *et al.* (2020) Genetic associations at regulatory phenotypes improve fine-mapping of causal variants for twelve immune-mediated diseases. *bioRxiv.*

24. Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J. *et al.* (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.*, **6**, e1000952.

25. Zhang, D., Cheng, L., Badner, J.A., Chen, C., Chen, Q., Luo, W., Craig, D.W., Redman, M., Gershon, E.S. and Liu, C. (2010) Genetic control of individual differences in gene-specific methylation in human brain. *Am. J. Hum. Genet.*, **86**, 411–419.

26. Del Arco, A. and Satrústegui, J. (1998) Molecular cloning of aralar, a new member of the mitochondrial carrier superfamily that binds calcium and is present in human muscle and brain. *J. Biol. Chem.*, **273**, 23327–23334.

27. Liu, J., Yang, A., Zhang, Q., Yang, G., Yang, W., Lei, H., Quan, J., Qu, F., Wang, M., Zhang, Z. *et al.* (2015) Association between genetic variants in *SLC25A12* and risk of autism spectrum disorders: an integrated meta-analysis. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **168B**, 236–246.

28. Turunen, J.A., Rehnstrom, K., Kilpinen, H., Kuokkanen, M., Kempas, E. and Ylisaukko-Oja, T. (2008) Mitochondrial aspartate/glutamate carrier *SLC25A12* gene is associated with autism. *Autism Res.*, **1**, 189–192.

29. Silverman, J.M., Buxbaum, J.D., Ramoz, N., Schmeidler, J., Reichenberg, A., Hollander, E., Angelo, G., Smith, C.J. and Kryzak, L.A. (2008) Autism-related routines and rituals associated with a mitochondrial aspartate/glutamate carrier *SLC25A12* polymorphism. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **147**, 408–410.

30. Segurado, R., Conroy, J., Meally, E., Fitzgerald, M., Gill, M. and Gallagher, L. (2005) Confirmation of association between autism and the mitochondrial aspartate/glutamate carrier *SLC25A12* gene on chromosome 2q31. *Am. J. Psychiatry*, **162**, 2182–2184.

31. Lepagnol-Bestel, A.M., Maussion, G., Boda, B., Cardona, A., Iwayama, Y., Delezoide, A.L., Moalic, J.M., Muller, D., Dean, B., Yoshikawa, T. *et al.* (2008) *SLC25A12* expression is associated with neurite outgrowth and is upregulated in the

prefrontal cortex of autistic subjects. *Mol. Psychiatry*, **13**, 385–397.

32. Park, S.H., Guastella, A.J., Lynskey, M., Agrawal, A., Constantino, J.N., Medland, S.E., Song, Y.J.C., Martin, N.G. and Colodro-Conde, L. (2017) Neuroticism and the overlap between autistic and ADHD traits: findings from a population sample of young adult Australian twins. *Twin Res. Hum. Genet.*, **20**, 319–329.

33. Wakabayashi, A., Baron-Cohen, S. and Wheelwright, S. (2006) Are autistic traits an independent personality dimension? A study of the Autism-spectrum quotient (AQ) and the NEO-PI-R. *Personal. Individ. Differ.*, **41**, 873–883.

34. Schwartzman, B.C., Wood, J.J. and Kapp, S.K. (2016) Can the five factor model of personality account for the variability of autism symptom expression? Multivariate approaches to behavioral phenotyping in adult autism spectrum disorder. *J. Autism Dev. Disord.*, **453**, 253–272.

35. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R. *et al.* (2016) Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.*, **19**, 1442–1453.

36. Qi, T., Wu, Y., Zeng, J., Zhang, F., Xue, A., Jiang, L., Zhu, Z., Kemper, K., Yengo, L., Zheng, Z. *et al.* (2018) Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.*, **9**, 2282.

37. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M. *et al.* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.*, **48**, 481–487.

38. Buccitelli, C. and Selbach, M. (2020) mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.*, **21**, 630–644.

39. Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M. *et al.* (2019) Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.*, **51**, 63–75.

40. Ruderfer, D.M., Ripke, S., McQuillin, A., Boocock, J., Stahl, E.A., Pavlides, J.M.W., Mullins, N., Charney, A.W., Ori, A.P.S., Loohuis, L.M.O. *et al.* (2018) Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell*, **173**, 1705–1715.

41. Howard, D.M., Adams, M.J., Shirali, M., Clarke, T., Marioni, R.E., Davies, G., Coleman, J.R.I., Alloza, C., Shen, X., Barbu, M.C. *et al.* (2018) Genome-wide association study of depression phenotypes in UK biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.*, **9**, 1470.

42. Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F. *et al.* (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.*, **50**, 668–681.

43. Jansen, P.R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A.R., de Leeuw, C.A., Benjamins, J.S., Muñoz-Manchado, A.B., Nagel, M. *et al.* (2019) Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.*, **51**, 394–403.

44. Elsworth, B.L., Mitchell, R., Raistrick, C.A., Paternoster, L., Hemani, G. and Gaunt, T.R. (2017) MRC IEU UK Biobank GWAS pipeline version 1. https://doi.org/10.5523/bris.2fahpksont1zi26xosyamqo8rr.

45. Aguet, F., Ardlie, K.G., Cummings, B.B., Gelfand, E.T., Getz, G., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., Karczewski, K.J. *et al.* (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.

46. Ng, B., White, C.C., Klein, H.U., Sieberts, S.K., McCabe, C., Patrick, E., Xu, J., Yu, L., Gaiteri, C., Bennett, D.A. *et al.* (2017) An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.*, **20**, 1418–1426.

47. Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T.M., Troakes, C., Turecki, G., O'Donovan, M.C., Schalkwyk, L.C. *et al.* (2015) Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.*, **19**, 48–54.

48. Jaffe, A.E., Gao, Y., Deep-Soboslay, A., Tao, R., Hyde, T.M., Weinberger, D.R. and Kleinman, J.E. (2015) Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.*, **19**, 40–47.

49. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.

50. Wald, A. (1940) The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.*, **11**, 284–300.

51. Burgess, S., Butterworth, A. and Thompson, S.G. (2013) Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.*, **37**, 658–665.

52. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C. and Plagnol, V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**.

53. R Core Team (2018) R: A language and environment for statistical computing. In *R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

54. Kolde, R. (2019) *Pheatmap: Pretty Heatmaps. R Package Version 1.0.12. Pheatmap: Pretty Heatmaps. R Package Version 1.0.12.*

55. de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.*, **11**, e1004219.

56. Watanabe, K., Taskesen, E., Van Bochoven, A. and Posthuma, D. (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.*, **8**.

57. Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M.R., McGrath, J.J., Visscher, P.M., Wray, N.R. *et al.* (2018) Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.*, **9**, 224.

58. Leyden, G., Gaunt, T.R. and Richardson, T.G. (2020) A factorial Mendelian randomization study to systematically prioritize genetic targets for the treatment of cardiovascular disease. *medRxiv*.

59. Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I. *et al.* (2017) A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.*, **16**, 19–34.

60. Nelson, M.R., Tipney, H., Painter, J.L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P.C., Li, M.J., Wang, J. *et al.* (2015) The support of human genetic evidence for approved drug indications. *Nat. Genet.*, **47**, 856–860.

61. Oprea, T.I., Bologa, C.G., Brunak, S., Campbell, A., Gan, G.N., Gaulton, A., Gomez, S.M., Guha, R., Hersey, A., Holmes, J. *et al.* (2018) Unexplored therapeutic opportunities in the human genome. *Nat. Rev. Drug Discov.*, **17**, 317–332.

62. Finan, C., Gaulton, A., Kruger, F.A., Lumbers, R.T., Shah, T., Engmann, J., Galver, L., Kelley, R., Karlsson, A., Santos, R. *et al.* (2017)

The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.*, **9**, eaag1166.

63. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R. *et al.* (2018) The MR-base platform supports systematic causal inference across the human phenome. *Elife*, **7**, e34408.

64. Wickham, H. (2016) ggplot2: elegant graphics for data analysis. ggplot2: elegant graphics for data analysis.

65. Guzman, R.E., Alekov, A.K., Filippov, M., Hegermann, J. and Fahlke, C. (2014) Involvement of ClC-3 chloride/proton exchangers in controlling glutamatergic synaptic strength in cultured hippocampal neurons. *Front. Cell. Neurosci.*, **8**, 143.

66. Yoshikawa, M., Uchida, S., Ezaki, J., Rai, T., Hayama, A., Kobayashi, K., Kida, Y., Noda, M., Koike, M., Uchiyama, Y. *et al.* (2002) CLC-3 deficiency leads to phenotypes similar to human neuronal ceroid lipofuscinosis. *Genes Cells*, **7**, 597–605.

67. Hauberg, M.E., Zhang, W., Giambartolomei, C., Franzén, O., Morris, D.L., Vyse, T.J., Ruusalepp, A., Fromer, M., Sieberts, S.K., Johnson, J.S. *et al.* (2017) Large-scale identification of common trait and disease variants affecting gene expression. *Am. J. Hum. Genet.*, **100**, 885–894.

68. Ma, C., Gu, C., Huo, Y., Li, X. and Luo, X.J. (2018) The integrated landscape of causal genes and pathways in schizophrenia. *Transl. Psychiatry*, **8**, 67.

69. Yuan, Q., Yang, F., Xiao, Y., Tan, S., Husain, N., Ren, M., Hu, Z., Martinowich, K., Ng, J.S., Kim, P.J. *et al.* (2016) Regulation of brain-derived neurotrophic factor exocytosis and gamma-aminobutyric acidergic interneuron synapse by the schizophrenia susceptibility gene dysbindin-1. *Biol. Psychiatry*, **80**, 312–322.

70. Rietveld, C.A., Esko, T., Davies, G., Pers, T.H., Turley, P., Benyamin, B., Chabris, C.F., Emilsson, V., Johnson, A.D., Lee, J.J. *et al.* (2014) Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 13790–13794.

71. Hanlon, D.W., Yang, Z. and Goldstein, L.S.B. (1997) Characterization of *KIFC2*, a neuronal kinesin superfamily member in mouse. *Neuron*, **18**, 439–451.

72. Krane, I.M., Naylor, S.L., Helin-Davis, D., Chin, W.W. and Spindel, E.R. (1988) Molecular cloning of cDNAs encoding the human bombesin-like peptide neuromedin B. Chromosomal localization and comparison to cDNAs encoding its amphibian homolog ranatensin. *J. Biol. Chem.*, **263**, 13317–13323.

73. Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A. *et al.* (2019) Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.*, **51**, 793–803.

74. Söllner, T. and Rothman, J.E. (1994) Neurotransmission: harnessing fusion machinery at the synapse. *Trends Neurosci.*, **17**, 344–348.

75. Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H. *et al.* (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.

76. Wu, Q. and Maniatis, T. (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell*, **97**, 779–790.

77. Park, Y., Sarkar, A., Bhutani, K. and Kellis, M. (2017) Multi-tissue polygenic models for transcriptome-wide association studies. *bioRxiv*.

78. Hatcher, C., Relton, C.L., Gaunt, T.R. and Richardson, T.G. (2019) Leveraging brain cortex-derived molecular data to elucidate epigenetic and transcriptomic drivers of complex traits and disease. *Transl. Psychiatry*, **9**, 105.

79. Stodberg, T., McTague, A., Ruiz, A.J., Hirata, H., Zhen, J., Long, P., Farabella, I., Meyer, E., Kawahara, A., Vassallo, G. *et al.* (2015) Mutations in *SLC12A5* in epilepsy of infancy with migrating focal seizures. *Nat. Commun.*, **6**, 8038.

80. Zhong, J., Li, S., Zeng, W., Li, X., Gu, C., Liu, J. and Luo, X.J. (2019) Integration of GWAS and brain eQTL identifies FLOT1 as a risk gene for major depressive disorder. *Neuropsychopharmacology*, **44**, 1542–1551.

81. Blaesse, P., Airaksinen, M.S., Rivera, C. and Kaila, K. (2009) Cation-chloride cotransporters and neuronal function. *Neuron*, **61**, 820–838.

82. Fan, Q., Wang, W., Hao, J., He, A., Wen, Y., Guo, X., Wu, C., Ning, Y., Wang, X., Wang, S. *et al.* (2017) Integrating genome-wide association study and expression quantitative trait loci data identifies multiple genes and gene set associated with neuroticism. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry*, **78**, 149–152.

83. Zhao, B., Luo, T., Li, T., Li, Y., Zhang, J., Shan, Y., Wang, X., Yang, L., Zhou, F., Zhu, Z. *et al.* (2019) Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat. Genet.*, **51**, 1637–1644.

84. Seipel, K., O'Brien, S.P., Lannotti, E., Medley, Q.G. and Streuli, M. (2001) Tara, a novel F-actin binding protein, associates with the Trio guanine nucleotide exchange factor and regulates actin cytoskeletal organization. *J. Cell Sci.*, **114**, 389–399.