

Identification of Brain Expression Quantitative Trait Loci Associated with Schizophrenia and Affective Disorders in Normal Brain Tissue

Oneil G. Bhalala PhD^{1,†}, Artika P. Nath MSc^{1,2}, UK Brain Expression Consortium*, Michael Inouye PhD^{1,3} & Christopher R. Sibley PhD^{1,4,5,†}

¹Centre for Systems Genomics, School of BioSciences, The University of Melbourne, Parkville, Victoria 3010, Australia

²Department of Microbiology and Immunology, The University of Melbourne, Parkville, Victoria 3010, Australia

³Department of Pathology, The University of Melbourne, Parkville 3010, Victoria, Australia

⁴Department of Molecular Neuroscience, University College London Institute of Neurology, Russell Square House, Russell Square, London WC1B 5EH, United Kingdom

⁵Department of Medicine, Division of Brain Sciences, Imperial College London, Burlington Danes, DuCane Road, London W12 0NN, United Kingdom

*See Supplementary File for affiliations.

†Corresponding Authors:

Oneil G. Bhalala, PhD
Centre for Systems Genomics
Building 184, Royal Parade
The University of Melbourne
Parkville, Victoria 3010
Australia
+61 3 9035 8659
obhalala@student.unimelb.edu.au

Christopher R. Sibley, PhD
Department of Medicine
Division of Brain Sciences
Burlington Danes, DuCane Road
Imperial College London
London, W12 0NN
United Kingdom
+44 207 5947019
c.sibley@imperial.ac.uk

Abstract

Schizophrenia and the affective disorders, here comprising bipolar disorder and major depressive disorder, are psychiatric illnesses that lead to significant morbidity and mortality worldwide. Whilst understanding of their pathobiology remains limited, large case-control studies have recently identified single nucleotide polymorphisms (SNPs) associated with these disorders. However, discerning the functional effects of these SNPs has been difficult as the associated causal genes are unknown. Here we evaluated whether schizophrenia and affective disorder associated-SNPs are correlated with gene expression within human brain tissue. Specifically, to identify expression quantitative trait loci (eQTLs), we leveraged disorder-associated SNPs identified from six Psychiatric Genomics Consortium and CONVERGE Consortium studies with gene expression levels in post-mortem, neurologically-normal tissue from two independent human brain tissue expression datasets (UK Brain Expression Consortium (UKBEC) and Genotype-Tissue Expression (GTEx)). We identified 6 188 and 16 720 *cis*-acting SNPs exceeding genome-wide significance ($p < 5 \times 10^{-8}$) in the UKBEC and GTEx datasets, respectively. 1 288 *cis*-eQTLs were significant in a meta-analysis leveraging overlapping brain regions and were associated with expression of 15 genes, including three non-coding RNAs. One *cis*-eQTL, rs16969968, results in a functionally disruptive missense mutation in *CHRNA5*, a schizophrenia-implicated gene. Meta-analysis identified 297 *trans*-eQTLs associated with 24 genes that were significant in a region-specific manner. Importantly, comparing across tissues, we find that blood eQTLs largely do not capture brain *cis*-eQTLs. This study identifies putatively causal genes whose expression in region-specific brain tissue may contribute to the risk of schizophrenia and affective disorders.

Introduction

Schizophrenia and affective disorders, comprising bipolar disorder and major depressive disorder, constitute a significant global burden of disease. Worldwide it is estimated that more than 21 million individuals are living with schizophrenia, 60 million with bipolar disorder and over 400 million with major depressive disorder¹. The consequences are staggering as these three diseases, which usually emerge in early-adulthood, accounted for over 90 million disability-adjusted life years in 2010².

Ineffective management, which contributes to the enormous disease burden, is largely due to our lack of understanding about the pathobiology underlying these disorders. Family and twin studies have estimated heritability to be between 70% and 80% for schizophrenia and bipolar disorder and up to 40% for major depressive disorder³. This has prompted the establishment of genome-wide association studies (GWAS) to identify genetic variants associated with these disorders. Recent and large-scale GWAS organized by the Psychiatric Genomics Consortium (PGC) and CONVERGE Consortium have been published for schizophrenia⁴, bipolar disorder^{5, 6}, major depressive disorder^{7, 8} as well as for a multiple-disorder analysis⁹. Results from these studies have suggested that schizophrenia, bipolar disorder and major depressive disorder may share common genetic architecture^{10, 11}.

While GWAS have identified numerous loci associated with human diseases^{12, 13}, understanding their roles in disease mechanism remains limited. Expression quantitative trait loci (eQTLs) are genetic variants that affect gene expression levels and may offer insights into mechanisms contributing to health and disease¹⁴⁻¹⁶. eQTLs can act in *cis*, meaning that the variant is associated expression of a gene within 1Mb, or in *trans*, with the variant located outside of this window. These two categories of eQTLs highlight the different avenues by

which eQTLs may be exerting their influence in cellular processes¹⁷. Studies have leveraged GWAS data, particularly that of single nucleotide polymorphisms (SNPs), to identify eQTLs^{18, 19}. Moreover, it is becoming evident that disease-associated variants are enriched for eQTLs^{9, 20}. Hence, eQTL studies based on disease-based GWAS may offer insights into important molecular processes underlying pathobiology.

An important consideration in eQTL analyses is the source of tissue used for gene expression data. While identification of eQTLs is increased by studying associations in multiple tissues^{21, 22}, other studies have found that analyses in disease-related tissues are enriched for disease-associated eQTLs^{23, 24}, especially *trans*²⁵⁻²⁷. Recent availability of datasets detailing gene expression in various regions of human brain have now allowed for eQTL analyses in nervous tissue²⁸⁻³⁰.

In this study, we sought to identify eQTLs associated with schizophrenia and affective disorders in neurologically-normal post-mortem brain tissue. By leveraging gene expression data from UK Brain Expression Consortium (UKBEC)²⁹ and Genotype-Tissue Expression (GTEx)³⁰ with SNP data available from PGC, we identified *cis*-eQTLs that were pervasive across various brain regions. In contrast, *trans*-eQTLs were only detected in a region-specific manner. Results reported here may help prioritize future studies of GWAS SNPs associated with these disorders.

Materials and Methods

Collation of analysis-SNPs from GWAS data

Publically available GWAS data from six recent studies related to schizophrenia, bipolar disorder and major depressive disorder were included in this analysis (Figure 1). Data were

downloaded from the PGC website (<https://www.med.unc.edu/pgc>). SNPs were collated from the following studies: PGC-SCZ2⁴ for schizophrenia, PGC-BIP⁵ and PGC-MooDs⁶ for bipolar disorder, PGC-MDD⁷ and CONVERGE⁸ for major depressive disorder, and PGC-Cross Disorder Analysis⁹ for multiple disorders (schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder and attention-deficit hyperactivity disorder). Of note, PGC-MooDs analysis included samples from the PGC-BIP study; PGC-Cross Disorders analysis included samples from the PGC-SCZ2, PGC-BIP and PGC-MDD studies. We included overlapping studies to maximize the number of disorder-associated SNPs in our analysis as there may be loci identified in one study but not in another.

SNPs with a study p-value $< 5 \times 10^{-5}$ were included in the analysis. For SNPs from all studies except PGC-SCZ2, we also obtained SNPs that are in moderate-high linkage disequilibrium (LD, $R^2 \geq 0.5$) with the study-SNPs³¹ using the web-based tool rAggr (<http://raggr.usc.edu/>). LD-analysis settings were as follows: CEU population from 1000 Genomes Phase 3 October 2014 release, build hg19, minimum minor allele frequency = 0.001, maximum distance = 500kb, maximum Mendelian errors = 1, Hardy-Weinberg p-value cutoff = 0, and minimum genotype = 75%. Study-SNPs from PGC-BIP, PGC-MDD and PGC-Cross Disorder Analysis were lifted from hg18 to hg19 prior to LD-analysis. PGC-SCZ2 had nearly 46,000 study SNPs with $p < 5 \times 10^{-5}$, so additional SNPs in LD with these were not obtained. Supplementary Table 1 indicates the number of study-SNPs and analysis-SNPs (study-SNPs and those SNPs in LD with them) per study.

Brain expression and genotype data

Genotype and gene expression data were obtained from the UK Brain Expression Consortium (UKBEC)²⁹. These data contained samples from 134 neuropathologically-free individuals

from the following brain regions: cerebellum (n=96), frontal cortex (n=97), hippocampus (n=96), medulla (n=95), occipital cortex (n=98), putamen (n=99), substantia nigra (n=73), temporal cortex (n=86), thalamus (n=95), and white matter (n=97). Processing of genotype is previously described²⁹. Briefly, these data included ~5.88 million imputed (1000 Genome, March 2012 release) and typed SNPs. Raw expression data from Affymetrix Human Exon 1.0 ST microarrays were processed as described previously with minor modifications. Specifically, all ~5 million probes were initially re-mapped to Ensembl v75 annotations using Biomart. Only probe sets containing three or more probes free of the ‘polymorphism-in-probe’ problem were used for subsequent analysis³². Following RMA normalisation and background filtering, we calculated gene-level estimates by taking the winsorised mean of all probe sets for a given gene. Prior to eQTL detection, these were covariate-corrected for sex, cause of death, post-mortem interval and RIN.

Data from Genotype-Tissue Expression (GTEx)³⁰ were also included in our study. The following brain regions were analysed: anterior cingulate cortex (BA24) (n=72), caudate (n=100), cerebellar hemisphere (n=89), cerebellum (n=103), cortex (n=96), frontal cortex (BA9) (n=92), hippocampus (n=81), hypothalamus (n=81), nucleus accumbens (n=93), and putamen (n=82). Whole blood tissue (n=338) was also analysed, serving as a comparison to brain tissues. Genotype and expression processing is described at <http://gtexportal.org>. Briefly, dataset contained ~11.55 million imputed (1000 Genomes, August 2012 release) and typed autosomal variants and ~26,000 transcripts (RPKM data from RNA-Seq with quantile normalization). To make this dataset similar to UKBEC, we excluded individuals from GTEx that were identified as having neurological diseases based on description of their comorbidities/cause of death or variables (variables excluded were: MHALS (ALS), MHALZDMT (Alzheimer's or dementia), MHALZHMR (Alzheimer's), MHPKNSN

(Parkinson's disease), MHDMENTIA (dementia with unknown cause), MHDPRSSN (major depression), MHSCHZ (schizophrenia), MHENCEPHA (active encephalitis), MHJAKOB (Cruetzfeldt Jakob relatives), MHMENINA (active meningitis), MHMS (multiple sclerosis), MHSZRSU (unexplained seizures)).

eQTL analysis

The R package Matrix eQTL³³ was used to identify eQTLs. *Cis*-eQTLs were defined as SNPs within 1Mb of the transcription start site; *trans*-eQTLs were those SNPs outside this region. Study genotype data was limited to those that were present in the set of analysis-SNPs described above. Due to the relatively low sample numbers, there is a possibility that an eQTL signal may be deemed significant if it is showing a very strong effect in only a few individuals. To minimize this event, analysis-SNPs that had a minor allele frequency < 5% within the analysis population were excluded. For UKBEC, gene expression input was based on residuals from linear regression of gene expression and genotype with the following covariates: sex, cause of death, age at death, post-mortem interval and RIN. Hence, no covariate data were separately considered in Matrix eQTL. Covariate data consisting of three genotyping principal components (PCs), array platform, gender and PEER factors were utilised for GTEx analysis. eQTL analysis, for both *cis* and *trans*, was performed independently for each region per study. eQTL (beta) effect sizes are given as standardized expression units (EU) per allele.

Meta-analyses across brain regions and studies

To determine which eQTLs were significant in multiple regions/studies, we performed a meta-analysis of test statistics (eQTL effect size estimates and standard errors) calculated by Matrix eQTL using the `metagen` function in the R package `meta` with a random-effects

model. False discovery rate (FDR) was calculated using R `p.adjust` with the Benjamini-Hochberg method. Unless otherwise stated we used a genome-wide significant p-value of $p < 5 \times 10^{-8}$ as our cutoff for calling a significant eQTL. For completeness, in supplementary tables, we additionally report the number of eQTLs passing a FDR threshold of $FDR < 0.05$. Four groups of meta-analyses were performed (Figure 1): A) of all ten regions in UKBEC for *cis* and *trans*, separately (p-valued denoted as p_{ukbec}); B) of all ten regions in GTEx for *cis* and *trans*, separately (p_{gtex}); C) of four overlapping regions (cerebellum, frontal cortex, hippocampus and putamen) between UKBEC and GTEx for *cis* (p_{over}); and D) by pairs of overlapping regions (separate meta-analysis for each cerebellum, frontal cortex, hippocampus and putamen) for *trans* (p_{pair}). To increase power for analyses C and D, only eQTLs that had the same effect direction in each study per region were analysed. The first ten principal components of the UKBEC-GTEx meta-analysis for each of the overlapping regions are shown in Supplementary Figures 1-4.

Permutation analysis to test meta-analysis results

Permutation analysis was used to test the robustness of eQTLs identified in the meta-analysis of overlapping regions in UKBEC and GTEx datasets (analyses C and D above, Figure 1). 10 000 iterations were performed for each region-study combination by shuffling sample IDs in the gene expression file and using Matrix eQTL as described above. Meta-analysis of each permuted iteration per combination yielded a meta p-value. Permuted p-value was calculated as the number of meta p-values equal to or less than the nominal p-value divided by the number of iterations (10 000). Those eQTLs with permuted p-values less than 0.05 were considered significant.

Results

Identification of disease-associated eQTLs in UKBEC and GTEx

We sought to determine if SNPs associated with schizophrenia and affective disorders also served as eQTLs in brain tissue. Study-SNPs identified in six GWAS were included in analysis: PGC-SCZ2⁴ for schizophrenia, PGC-BIP⁵ and PGC-MooDs⁶ for bipolar disorder, PGC-MDD⁷ and CONVERGE⁸ for major depressive disorder, and PGC-Cross Disorder⁹ which included schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder and attention-deficit hyperactivity disorder (Figure 1). In order to maximise detection of eQTLs³¹, we included SNPs that are in moderate to high LD ($R^2 \geq 0.5$) with study-SNPs using rAggr (Supplementary Table 1). Combining study-SNPs and those in LD with them yielded 72 208 analysis-SNPs across the six GWAS.

Genotype and gene expression data were obtained from UKBEC²⁹ and GTEx³⁰, which contained information of post-mortem human tissue from samples independent from each other and from the GWAS used to identify analysis-SNPs. Of the 72 208 analysis-SNPs, 58 325 and 58 133 were present in both UKBEC and GTEx, respectively. Expression data originated from ten UKBEC regions (cerebellum, frontal cortex, hippocampus, putamen, temporal cortex, occipital cortex, substantia nigra, white matter, thalamus, medulla) and ten GTEx regions (anterior cingulate cortex (BA24), caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens, putamen). Samples in UKBEC were neurologically-disease free. Those samples with neurological disease in GTEx data were removed from analysis (see Methods), allowing for identification of eQTLs in disease-free tissue.

To identify *cis*-eQTLs in these tissues, we utilised an additive linear model with Matrix eQTL³³. Using this methodology, we initially analyzed *cis*-eQTLs in each of the ten regions

within UKBEC and GTEx datasets, separately (Figure 1, analyses A,B, Supplementary Figures 5,6). The number of detected *cis*-eQTLs using a stringent threshold of $p < 5 \times 10^{-8}$ varied considerably between studies and regions (Supplementary Table 2). For example, in the UKBEC dataset, the cerebellum harboured 1 234 unique *cis*-eQTLs, while none were detected in substantia nigra or medulla. Within the GTEx dataset, 2 879 and 696 were detected in the cerebellum and anterior cingulate cortex, respectively. We meta-analysed *cis*-eQTLs across all ten regions in UKBEC and GTEx separately (Supplementary Figure 7, Supplementary Tables 3 and 4). When collectively considered, 6 188 and 16 720 unique *cis*-eQTLs were significant ($p_{\text{ukbec/gtex}} < 5 \times 10^{-8}$) in the UKBEC and GTEx data, respectively. These *cis*-eQTLs were associated with expression variation of 146 and 506 genes, respectively. 3 169 *cis*-eQTLs, associated with 46 genes, were shared between these datasets at a threshold of $p_{\text{ukbec/gtex}} < 5 \times 10^{-8}$ (Supplementary Figure 8).

The number of *trans*-eQTLs that were found to be significantly associated in any of the individual ten regions per study was greatly reduced compared to *cis*-eQTLs. Zero *trans*-eQTLs were detected in meta-analysis of all ten UKBEC regions (Figure 1, analysis A). In the GTEx dataset (Figure 1, analysis B), 103 *trans*-eQTLs were detected and associated with expression of two non-coding RNAs of unknown function located on chromosome 8: *RNF5P1* (most significant SNP is rs3130349 on chromosome 6p21.32, $p_{\text{gtex}} = 8.3 \times 10^{-222}$) and *FAM66A* (most significant SNP is rs7832708 on chromosome 8p23.1, $p_{\text{gtex}} = 8.8 \times 10^{-62}$).

Meta-analysis of eQTLs in overlapping brain regions across studies

To generate a high-confidence list of eQTLs shared between datasets as well as to discover additional ones, we next sought to determine if eQTLs identified in either UKBEC or GTEx data were significant when assessed across both studies. Four brain regions overlapped

between the two datasets: cerebellum, frontal cortex, hippocampus and putamen. Test-statistics restricted to these regions were meta-analysed (Figure 1, analyses C and D). This reduced the aforementioned 3 169 overlapping *cis*-eQTLs to 1 288 *cis*-eQTLs that were significant with $p_{\text{over}} < 5 \times 10^{-8}$ (Supplementary Table 5). To further validate *cis*-eQTL associations, we performed permutation analysis of 10 000 iterations for each expression dataset and brain region (resulting in eight expression-region groups meta-analysed 10 000 times each). Each of the 1 288 *cis*-eQTLs with $p_{\text{over}} < 5 \times 10^{-8}$ had a permuted p-value $< 1 \times 10^{-4}$.

The 1 288 *cis*-eQTLs were associated with expression of 15 genes (Figure 2A, Table 1). Of these, 784 *cis*-eQTLs associated with 10 genes also had $p_{\text{ukbec}} < 5 \times 10^{-8}$ in the UKBEC meta-analysis of the four overlapping brain regions and 714 *cis*-eQTLs associated with 6 genes also had $p_{\text{gtex}} < 5 \times 10^{-8}$ in the GTEx meta-analysis of the four overlapping brain regions. Demonstrating reproducibility between the two studies, 568 of the *cis*-eQTLs within these two groups had both $p_{\text{ukbec}} < 5 \times 10^{-8}$ and $p_{\text{gtex}} < 5 \times 10^{-8}$ (Figure 2B). Moreover, an additional 358 *cis*-eQTLs were detected when both UKBEC and GTEx datasets were leveraged in combination, thus demonstrating the benefit of a combined analysis. Mean and median of the (absolute value of) effect sizes (irrespective of p_{over}) were 0.33 and 0.34 EU per allele, respectively (Supplementary Figure 9). The largest effect size was observed with rs9461434 (-0.70 ± 0.12 EU per allele for AA relative to GG), associated with the expression of the pseudogene *ZNF603P* ($p_{\text{over}} = 2.1 \times 10^{-8}$, FDR = 1.4×10^{-6} , Supplementary Figure 10).

Of the 1 288 *cis*-eQTLs with $p_{\text{over}} < 5 \times 10^{-8}$, 40.1% were located in introns and 22.8% within intergenic regions (Figure 2C), as classified by Ensembl Variant Effect Predictor. The most significant *cis*-eQTL was rs938682 on chromosome 15q25.1, correlated with expression of

CHRNA5 ($p_{\text{over}} = 1.7 \times 10^{-23}$, $\text{FDR} = 5.8 \times 10^{-19}$), which encodes the $\alpha 5$ subunit of the nicotinic cholinergic receptor. This eQTL has an effect size of 0.26 ± 0.03 EU per allele for AA (allele frequency of A = 0.79 in 1000 Genomes CEU population) relative to GG (Figure 3A,B). Interestingly, the eQTL is located within an intron of *CHRNA3*, whose protein interacts with *CHRNA5* to form a nicotinic acetylcholine receptor³⁴. There was no correlation between *CHRNA3* and *CHRNA5* expression with a Pearson $R^2 = 0.02$. Moreover, no significant *cis*-eQTLs associated with *CHRNA3* expression were observed.

Interestingly, 26 *cis*-eQTLs (Supplementary Table 6) were classified as leading to missense mutations. Five eQTLs resulted in missense mutations in genes whose expression with which they were associated. The most significant of these was rs16969968 ($p_{\text{over}} = 8.4 \times 10^{-13}$), which is located within exon 5 of *CHRNA5*, a highly conserved region (Figure 3C). This finding lends further support to the involvement of *CHRNA5* in these diseases. The minor allele A encodes for an amino acid change to asparagine from aspartic acid (major allele G) at position 398, which may affect receptor function (see Discussion). The effect size of GG, relative to AA, is 0.23 ± 0.03 EU per allele (Supplementary Figure 11), thus implying that both expression changes and receptor activity could be contributing to the effect of this variant.

As noted above, only two *trans*-eQTLs were detected in GTEx. As expected, neither remained significant when the four overlapping regions were meta-analysed together. However, meta-analysis of each region across both studies identified varying amount of *trans*-eQTLs with $p_{\text{pair}} < 5 \times 10^{-8}$ (Figure 1, analysis D; Table 2; Supplementary Tables 7-10). Twelve *trans*-eQTLs associated with five genes within the cerebellum; 255 *trans*-eQTLs associated with twelve genes within the frontal cortex; eleven *trans*-eQTLs associated with

two genes within the hippocampus; and 19 *trans*-eQTLs associated with five genes within the putamen (Figure 4³⁵). These *trans*-eQTLs remained significant after permutation testing, with each having a permuted p-value $< 1 \times 10^{-4}$. In the frontal cortex, nearly 70% of *trans*-eQTLs were located within the *MHC* locus on chromosome 6p21 and were associated with expression of *DDX17* on chromosome 22q13.1, which encodes for a RNA-binding protein. While the specific role of this protein, also known as p72, in brain is still uncertain, it is involved in numerous aspects of RNA metabolism, including promoting microRNA biogenesis and reducing viral RNA stability^{36, 37}. *DDX17* is also a binding-partner to the transcription factor *SOX2*³⁸, thereby potentially influencing master regulation of gene expression.

Overlap with whole blood tissue eQTLs

Previous eQTL analyses for schizophrenia and affective disorders have relied on transcriptome data collected from more readily available biological specimens such as whole blood³⁹⁻⁴¹ or single brain regions⁴¹. An important question is whether disease-related eQTLs can be detected in samples that are more accessible and available from living patients. If so, this may facilitate larger study designs in the future where high quality biological material is more readily selected. To test whether our significant *cis*-eQTLs detected brain regions can be identified in more clinically-accessible tissue, we assessed the associations of 1 288 *cis*-eQTLs with $p_{\text{over}} < 5 \times 10^{-8}$ (Figure 1, analysis C) in whole blood tissue within the GTEx database ($n = 282$ after removal of individuals with neurological disease (see Methods)). Of these, 472 had a $p_{\text{blood}} < 0.05$. However, only 143 (11%) reached our stringent threshold and had a $p_{\text{blood-GTEx}} < 5 \times 10^{-8}$ (Supplementary Figure 12A). Looking in more detail, of the fifteen genes associated with brain *cis*-eQTLs from our cross-study meta-analysis, only twelve had detectable expression in the GTEx whole blood samples. Moreover, only *BTN3A2* was

associated with blood *cis*-eQTLs having $p_{\text{blood-GTEEx}} < 5 \times 10^{-8}$. The most significant brain *cis*-eQTL in whole blood was rs71557332 (chromosome 6p22.2), which correlated with *BTN3A2* expression ($p_{\text{blood-GTEEx}} = 1.8 \times 10^{-52}$ and $\text{FDR} = 3.6 \times 10^{-49}$). Of note, the direction of the effect was the same as that seen in the brain eQTL; the effect size was 0.98 ± 0.05 EU per allele in blood compared to an effect size of 0.55 ± 0.10 EU per allele in brain (Supplementary Figure 12B,D). Contrastingly, the top brain *cis*-eQTL, rs938682, associated with *CHRNA5* expression, only had a $p_{\text{blood-GTEEx}} = 0.03$ and $\text{FDR} = 0.29$ (effect direction was the same).

Similar results were observed when we tested for overlap of *cis*-eQTLs detected in the largest meta-analysis of peripheral whole blood to date ($n = 5\,311$, no overlap with GTEEx whole blood data)⁴². In this analysis, only 69 out of brain *cis*-eQTLs, that were associated with three genes (*BTN3A2*, *CHRNA5* and *GSTO2*), were detected in whole blood (Supplementary Figure 12C). Of these, only *cis*-eQTLs associated with *BTN3A2* and *GSTO2* reached our significance threshold; eQTLs correlated with *CHRNA5* expression had a minimum $p_{\text{blood-Westra}} > 7 \times 10^{-4}$. Thus, whilst *cis*-eQTLs can be detected in whole blood in principle, approximately 90% did not reach genome-wide significance levels. This strongly suggests that eQTL analyses should either be performed in disease-relevant tissue wherever possible or that more extensive studies targeting minimally invasive tissues are necessary to identify suitable brain correlates.

Discussion

In this study, we report the presence of eQTLs in brain tissue for SNPs that are associated with a risk of developing schizophrenia and/or affective disorders. Even with a stringent p-value cut-off of 5×10^{-8} , we identified 1 288 *cis*-eQTLs and 297 *trans*-eQTLs that were correlated with expression of fifteen and twenty-four genes, respectively. These associations

held across four brain regions from two independent studies (UKBEC and GTEx) of neurologically-normal individuals as well as through permutation analyses of 10 000 iterations, supporting the robustness of these findings.

Meta-analyses of eQTLs across different brain regions revealed diverging patterns between *cis* and *trans*: substantially more *cis*-eQTLs were significant when considering the four overlapping brain than *trans*-eQTLs. The pleiotropy of *cis*-eQTLs across different brain regions described here is consistent with their replicative behaviour previously reported. Considering all *cis*-eQTLs in UKBEC, 25-50% were also found in other eQTL datasets, even in non-synonymous tissues like human monocytes^{29, 43}. In our study, ~35% of *cis*-eQTLs detected in brain were also present in GTEx whole blood samples, albeit at lower p-values. In contrast to *cis*, we could only detect *trans*-eQTLs for the same tissue regions between the two datasets. This is consistent with notion that *trans*-eQTLs may act in a tissue-dependent manner⁴⁴⁻⁴⁶. Given the differences in locations and associations of *trans*-eQTLs amongst the various brain regions, this supports the hypothesis that the biology governing these *trans*-associations may be more influenced by the intracellular environment than the *cis*-eQTLs¹⁷.

Of the 1 288 *cis*-eQTLs detected, nearly two-thirds were located within introns or intergenic regions, consistent with previous studies^{26, 41}. Meanwhile, nearly 2% were classified as causing missense mutations in the proteins encoded from genes harbouring the eQTL. The most significant eQTL that results in a missense mutation within the gene whose expression it is associated with, rs16969968, leads to an amino acid change of aspartic acid (for major allele *G*) to asparagine (for minor allele *A*) at position 398 in *CHRNA5*, a subunit in nicotinic acetylcholine receptors. This causes a change from a negatively charged side chain to one that is polar but uncharged. Interestingly, functional *in vitro* studies have demonstrated

that receptors containing this missense mutation are less responsive to a nicotinic agonist than ones with the more common variant^{34, 47}. This leads to reduced cell-depolarization and cholinergic signalling and is consistent with the reported hypofunction of cholinergic signalling in schizophrenia⁴⁸. Moreover, pharmaceutical modulation of cholinergic signalling may improve outcomes for schizophrenia⁴⁹.

There is already modest evidence that this and other eQTLs affecting *CHRNA5* expression are associated with schizophrenia and affective disorders^{4, 50, 51}. However, rs16969968 has also been shown to be associated with increased tobacco use^{34, 52} and incidence of lung cancer^{53, 54}. Therefore, given the disproportionate percentage of individuals with mental illness that smoke⁵⁵, further studies are needed to ensure that these eQTLs are not associated with a confounding behaviour in such individuals. Nonetheless, these *cis*-eQTLs for *CHRNA5* demonstrate how genetic analyses can identify variants that may increase disease-risk while concurrently being potential therapeutic targets.

Our analysis has also replicated several other *cis*-eQTLs found in other studies. Previously, 27 brain eQTLs were identified in a meta-analysis of GWAS SNPs associated with five neuropsychiatric disorders (including the three studied here⁹) using cortical expression data from five separate studies⁵⁶. We identified one of these overlapping with our study: rs4523957 associated with expression of *SRR* on chromosome 17p13.3. Though this was not the most significant eQTL for this gene, the association was still below the genome-wide threshold at $p_{\text{over}} = 2.2 \times 10^{-9}$, $\text{FDR} = 3.6 \times 10^{-7}$. *SRR* encodes for an enzyme that converts L- to D-serine, which has been found to be lower in the CSF of patients with schizophrenia⁵⁷, supporting the role of glutamatergic neurotransmission in the biology of schizophrenia and affective disorders^{58, 59}. Indeed, it is also a candidate drug target for schizophrenia, again

highlighting the potential for eQTL studies to identify pathobiology that might be targeted pharmacologically. *ZSCAN31*, also known as *ZNF323*, is another *cis*-eQTL associated gene that has been previously identified as significantly associated with schizophrenia, bipolar disorder and psychosis in both a GWAS⁶⁰ and an eQTL study using 193 human pre-frontal cortex samples⁶¹. In our study, multiple *cis*-eQTLs associated with this gene were significant ($p_{\text{over}} = 2.9 \times 10^{-8}$ for the top eQTL). Likewise, two genes that were prioritized as putatively causal from integrative analysis of PGC-SCZ2 GWAS with both whole blood and UKBEC data⁶², *SNX19* and *NMRALI*, were associated with *cis*-eQTLs with some evidence of significance, $p_{\text{over}} = 8.3 \times 10^{-5}$ and 2.2×10^{-6} , respectively.

Despite some overlap with previous studies, several reported associations were not significant here. *CACNA1C* and *ZNF804A* are two genes that have been implicated in schizophrenia and affective disorders through multiple studies, including GWAS and case/control brain expression analyses⁶³⁻⁶⁷. While both UKBEC and GTEx data contain expression information regarding these genes and the set of analysis-SNPs contain variants that are in *cis*, we did not find any significant eQTLs in our meta-analysis of UKBEC and GTEx across the four overlapping brain regions. For *CACNA1C* (minimum $p_{\text{over}} = 0.45$), the effect size estimates of the *cis*-eQTLs had different signs across the various regions (Supplementary Table 11). For *ZNF804A*, no detected eQTLs overlapped between all four regions (Supplementary Table 12). Further, of those that were in moderate-high LD with each other, the effect size estimates also demonstrated different signs. This suggests that some disease-implicated *cis*-eQTLs may have varying effects in different brain regions. These are not captured in this study but warrants further investigation.

The greatest number of *trans*-eQTLs detected in this study was predominantly located within the *MHC* locus on chromosome 6p21. eQTL analysis of peripheral blood and immune cells suggest that this region is enriched for *trans*-eQTLs^{68, 69}, especially when restricting variants to those that are trait/disease-associated⁷⁰, as was done in this study. However, given the poor replicability of *trans*-eQTLs thus far, we are cautious about over-interpreting these results; substantially larger sample sizes are needed to detect and validate such eQTLs. Nonetheless, the identification of tissue-specific *trans*-eQTLs that withstand permutation analysis gives confidence that our findings may be biologically significant.

Progress in understanding the biology governing schizophrenia and affective disorders has been hampered by difficulty in accessing disease-relevant tissue. Therefore, peripheral blood, as it is more accessible, has previously been used as a proxy for studying eQTLs of complex diseases^{16, 42, 62}. However, it is unclear as to how robust findings from blood samples are with respect to studying diseases outside of the hematopoietic system. By comparing *cis*-eQTL results from brain and blood, we find that a minority of the brain-eQTLs were detected in blood and their significance (as marked by p-value) was greatly diminished. Therefore, studying eQTLs in disease-related tissues, and in this case brain tissue, may prioritize eQTLs for further validation and mechanistic studies. Indeed this has previously been demonstrated in prostate tissue for eQTLs associated with prostate cancer⁷¹.

In summary, we have identified robust *cis*- and *trans*-eQTLs associated with schizophrenia and affective disorders in human brain tissue. Of the eQTL-associated genes, many have been implicated previously, such as *CHRNA5*, while others are novel associations (i.e. *ZNF603P*) that now merit further analyses. We also demonstrate that eQTL analysis in disease-related tissues allows for prioritization of associations for follow-up studies in diseased-tissue. These

results provide insight into putative mechanisms related to development of schizophrenia and affective disorders, thereby identifying potentially new therapeutic targets.

Files

Supplementary_Figures.pdf

This file contains Supplementary Figures 1-12.

Supplementary_Tables.xlsx

This file contains Supplementary Tables 1-12, with each table on a separate sheet.

UKBEC_Affiliations.pdf

This file contains the UK Brain Expression Consortium authors and respective affiliations.

Acknowledgements

We thank the individuals and their support groups who have contributed to these studies and Sean Byars for help and advice with data analysis.

Funding Sources

APN is supported the Australian Postgraduate Award and the International Postgraduate Research Scholarship from The University of Melbourne. MI is supported by an NHMRC and Australian Heart Foundation Career Development Fellowship (no. 1061435). CRS is supported by a fellowship from the Edmond J. Safra Foundation.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

OGB, MI and CRS designed the study. OGB, APN, and CRS carried out analyses. OGB wrote the manuscript with assistance from APN, MI and CRS.

References

1. Mental Disorders. <http://www.who.int/mediacentre/factsheets/fs396/en/>, 2014, Accessed Date Accessed 2014 Accessed.
2. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C *et al*. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380**(9859): 2197-2223.
3. Uher R. Gene-environment interactions in severe mental illness. *Front Psychiatry* 2014; **5**: 48.
4. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**(7510): 421-427.
5. 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**(10): 977-983.
6. Muhleisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F, Treutlein J *et al*. Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat Commun* 2014; **5**: 3339.
7. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM *et al*. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 2013; **18**(4): 497-511.
8. CONVERGE Consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* 2015; **523**(7562): 588-591.
9. Cross-Disorder Group of the Psychiatric Genomics C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**(9875): 1371-1379.
10. Cross-Disorder Group of the Psychiatric Genomics C, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM *et al*. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**(9): 984-994.
11. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR *et al*. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**(11): 1236-1241.
12. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009; **106**(23): 9362-9367.

13. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H *et al*. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014; **42**(Database issue): D1001-1006.
14. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V *et al*. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 2003; **422**(6929): 297-302.
15. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R *et al*. Genome-wide associations of gene expression variation in humans. *PLoS Genet* 2005; **1**(6): e78.
16. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J *et al*. Genetics of gene expression and its effect on disease. *Nature* 2008; **452**(7186): 423-428.
17. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 2015; **16**(4): 197-212.
18. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C *et al*. Population genomics of human gene expression. *Nat Genet* 2007; **39**(10): 1217-1224.
19. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS *et al*. Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004; **430**(7001): 743-747.
20. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010; **6**(4): e1000888.
21. Flutre T, Wen X, Pritchard J, Stephens M. A statistical framework for joint eQTL analysis in multiple tissues. *PLoS Genet* 2013; **9**(5): e1003486.
22. Peters JE, Lyons PA, Lee JC, Richard AC, Fortune MD, Newcombe PJ *et al*. Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. *PLoS Genet* 2016; **12**(3): e1005908.
23. Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ *et al*. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet* 2010; **6**(5): e1000932.
24. Richards AL, Jones L, Moskvina V, Kirov G, Gejman PV, Levinson DF *et al*. Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol Psychiatry* 2012; **17**(2): 193-201.
25. Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet* 2011; **7**(2): e1001317.

26. Fu J, Wolfs MG, Deelen P, Westra HJ, Fehrmann RS, Te Meerman GJ *et al.* Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet* 2012; **8**(1): e1002431.
27. Bryois J, Buil A, Evans DM, Kemp JP, Montgomery SB, Conrad DF *et al.* Cis and trans effects of human genomic variants on gene expression. *PLoS Genet* 2014; **10**(7): e1004461.
28. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 2010; **6**(5): e1000952.
29. 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**(10): 1418-1428.
30. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**(6235): 648-660.
31. Karch CM, Ezerskiy LA, Bertelsen S, Alzheimer's Disease Genetics C, Goate AM. Alzheimer's Disease Risk Polymorphisms Regulate Gene Expression in the ZCWPW1 and the CELF1 Loci. *PLoS One* 2016; **11**(2): e0148717.
32. Ramasamy A, Trabzuni D, Gibbs JR, Dillman A, Hernandez DG, Arepalli S *et al.* Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. *Nucleic Acids Res* 2013; **41**(7): e88.
33. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012; **28**(10): 1353-1358.
34. Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Grucza RA, Xuei X *et al.* Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry* 2008; **165**(9): 1163-1171.
35. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* 2009; **19**(9): 1639-1645.
36. Moy RH, Cole BS, Yasunaga A, Gold B, Shankarling G, Varble A *et al.* Stem-loop recognition by DDX17 facilitates miRNA processing and antiviral defense. *Cell* 2014; **158**(4): 764-777.
37. Remenyi J, Bajan S, Fuller-Pace FV, Arthur JS, Hutvagner G. The loop structure and the RNA helicase p72/DDX17 influence the processing efficiency of the mice miR-132. *Sci Rep* 2016; **6**: 22848.

38. Alqahtani H, Gopal K, Gupta N, Jung K, Alshareef A, Ye X *et al*. DDX17 (P72), a Sox2 binding partner, promotes stem-like features conferred by Sox2 in a small cell population in estrogen receptor-positive breast cancer. *Cell Signal* 2016; **28**(2): 42-50.
39. de Jong S, van Eijk KR, Zeegers DW, Strengman E, Janson E, Veldink JH *et al*. Expression QTL analysis of top loci from GWAS meta-analysis highlights additional schizophrenia candidate genes. *Eur J Hum Genet* 2012; **20**(9): 1004-1008.
40. Zhang C, Wang Z, Hong W, Wu Z, Peng D, Fang Y. ZNF804A Genetic Variation Confers Risk to Bipolar Disorder. *Mol Neurobiol* 2015.
41. Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K *et al*. Heritability and genomics of gene expression in peripheral blood. *Nat Genet* 2014; **46**(5): 430-437.
42. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J *et al*. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**(10): 1238-1243.
43. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R *et al*. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010; **5**(5): e10693.
44. Petretto E, Mangion J, Dickens NJ, Cook SA, Kumaran MK, Lu H *et al*. Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet* 2006; **2**(10): e172.
45. van Nas A, Ingram-Drake L, Sinsheimer JS, Wang SS, Schadt EE, Drake T *et al*. Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. *Genetics* 2010; **185**(3): 1059-1068.
46. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S *et al*. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* 2012; **44**(10): 1084-1089.
47. Sciacaluga M, Moriconi C, Martinello K, Catalano M, Bermudez I, Stitzel JA *et al*. Crucial role of nicotinic alpha5 subunit variants for Ca²⁺ fluxes in ventral midbrain neurons. *FASEB J* 2015; **29**(8): 3389-3398.
48. Balu DT. The NMDA Receptor and Schizophrenia: From Pathophysiology to Treatment. *Adv Pharmacol* 2016; **76**: 351-382.
49. Foster DJ, Jones CK, Conn PJ. Emerging approaches for treatment of schizophrenia: modulation of cholinergic signaling. *Discov Med* 2012; **14**(79): 413-420.

50. Hong LE, Yang X, Wonodi I, Hodgkinson CA, Goldman D, Stine OC *et al.* A CHRNA5 allele related to nicotine addiction and schizophrenia. *Genes Brain Behav* 2011; **10**(5): 530-535.
51. Jackson KJ, Fanous AH, Chen J, Kendler KS, Chen X. Variants in the 15q25 gene cluster are associated with risk for schizophrenia and bipolar disorder. *Psychiatr Genet* 2013; **23**(1): 20-28.
52. Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA *et al.* Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet* 2007; **16**(1): 36-49.
53. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008; **452**(7187): 633-637.
54. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008; **452**(7187): 638-642.
55. McClave AK, McKnight-Eily LR, Davis SP, Dube SR. Smoking characteristics of adults with selected lifetime mental illnesses: results from the 2007 National Health Interview Survey. *Am J Public Health* 2010; **100**(12): 2464-2472.
56. Kim Y, Xia K, Tao R, Giusti-Rodriguez P, Vladimirov V, van den Oord E *et al.* A meta-analysis of gene expression quantitative trait loci in brain. *Transl Psychiatry* 2014; **4**: e459.
57. Hashimoto K, Fukushima T, Shimizu E, Komatsu N, Watanabe H, Shinoda N *et al.* Decreased serum levels of D-serine in patients with schizophrenia: evidence in support of the N-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia. *Arch Gen Psychiatry* 2003; **60**(6): 572-576.
58. Hu W, MacDonald ML, Elswick DE, Sweet RA. The glutamate hypothesis of schizophrenia: evidence from human brain tissue studies. *Ann N Y Acad Sci* 2015; **1338**: 38-57.
59. Jun C, Choi Y, Lim SM, Bae S, Hong YS, Kim JE *et al.* Disturbance of the glutamatergic system in mood disorders. *Exp Neurol* 2014; **23**(1): 28-35.
60. Saito T, Kondo K, Iwayama Y, Shimasaki A, Aleksic B, Yamada K *et al.* Replication and cross-phenotype study based upon schizophrenia GWASs data in the Japanese population: support for association of MHC region with psychosis. *Am J Med Genet B Neuropsychiatr Genet* 2014; **165B**(5): 421-427.
61. Luo XJ, Mattheisen M, Li M, Huang L, Rietschel M, Borglum AD *et al.* Systematic Integration of Brain eQTL and GWAS Identifies ZNF323 as a Novel Schizophrenia

- Risk Gene and Suggests Recent Positive Selection Based on Compensatory Advantage on Pulmonary Function. *Schizophr Bull* 2015.
62. 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**(5): 481-487.
 63. Green EK, Grozeva D, Jones I, Jones L, Kirov G, Caesar S *et al*. The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol Psychiatry* 2010; **15**(10): 1016-1022.
 64. Nyegaard M, Demontis D, Foldager L, Hedemand A, Flint TJ, Sorensen KM *et al*. CACNA1C (rs1006737) is associated with schizophrenia. *Mol Psychiatry* 2010; **15**(2): 119-121.
 65. Roussos P, Mitchell AC, Voloudakis G, Fullard JF, Pothula VM, Tsang J *et al*. A role for noncoding variation in schizophrenia. *Cell Rep* 2014; **9**(4): 1417-1429.
 66. Tao R, Cousijn H, Jaffe AE, Burnet PW, Edwards F, Eastwood SL *et al*. Expression of ZNF804A in human brain and alterations in schizophrenia, bipolar disorder, and major depressive disorder: a novel transcript fetally regulated by the psychosis risk variant rs1344706. *JAMA Psychiatry* 2014; **71**(10): 1112-1120.
 67. Nie F, Wang X, Zhao P, Yang H, Zhu W, Zhao Y *et al*. Genetic analysis of SNPs in CACNA1C and ANK3 gene with schizophrenia: A comprehensive meta-analysis. *Am J Med Genet B Neuropsychiatr Genet* 2015; **168**(8): 637-648.
 68. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A *et al*. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* 2012; **44**(5): 502-510.
 69. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E *et al*. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 2014; **343**(6175): 1246949.
 70. Fehrmann RS, Jansen RC, Veldink JH, Westra HJ, Arends D, Bonder MJ *et al*. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet* 2011; **7**(8): e1002197.
 71. Thibodeau SN, French AJ, McDonnell SK, Cheville J, Middha S, Tillmans L *et al*. Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. *Nat Commun* 2015; **6**: 8653.

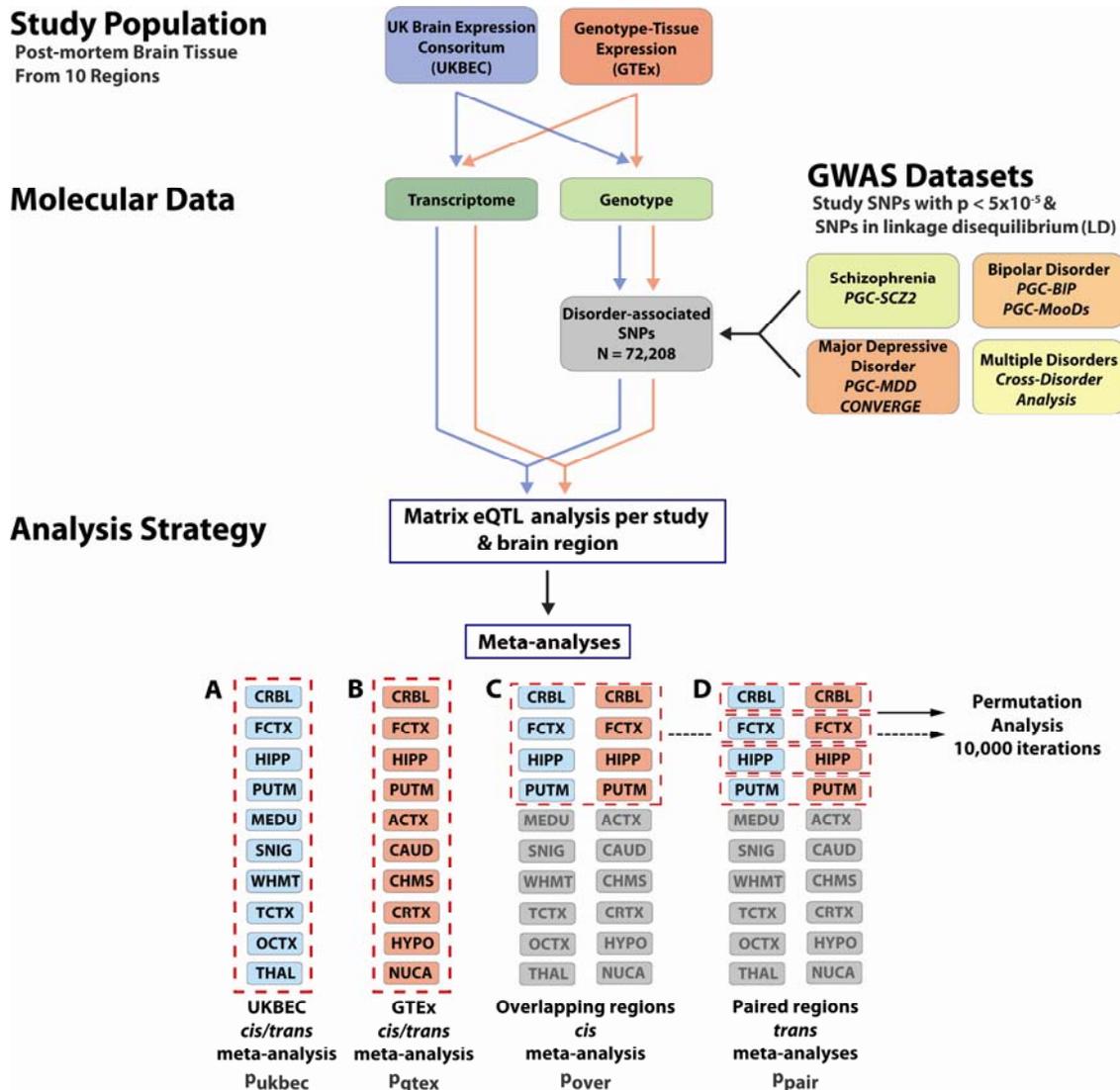


Figure 1 | Study design for identification of brain eQTLs. Two studies with genotype and transcriptomic data (UKBEC and GTEx) were analysed for eQTLs based on SNPs that were found to be associated with schizophrenia, bipolar disorder and major depressive disorder through six GWAS studies. UKBEC consisted of samples from neurologically-normal individuals; samples from individuals with neurological disease (see Methods) were removed from GTEx prior to analyses. For each study, each of the ten brain regions was analysed separately by Matrix eQTL for identification of *cis*- and *trans*-eQTLs. Matrix eQTL results were meta-analysed in the following manner: *A, B*, within each study, all ten regions were meta-analysed for both *cis*- and *trans*-eQTLs (p-values are denoted as p_{ukbec} and p_{gtex},

respectively); *C*, overlapping regions between the two studies were meta-analysed for *cis*-eQTLs (p_{over}); and *D*, each pair of overlapping regions was meta-analysed separately for *trans*-eQTLs (p_{pair}). *CRBL*, cerebellum; *FCTX*, frontal cortex; *HIPP*, hippocampus; *PUTM*, putamen; *MEDU*, medulla; *SNIG*, substantia nigra; *WHMT*, white matter; *TCTX*, temporal cortex; *OCTX*, occipital cortex; *THAL*, thalamus; *ACTX*, anterior cingulate cortex; *CAUD*, caudate; *CHMS*, cerebellar hemisphere; *CRTX*, cortex; *HYP0*, hypothalamus; *NUCA*, nucleus accumbens.

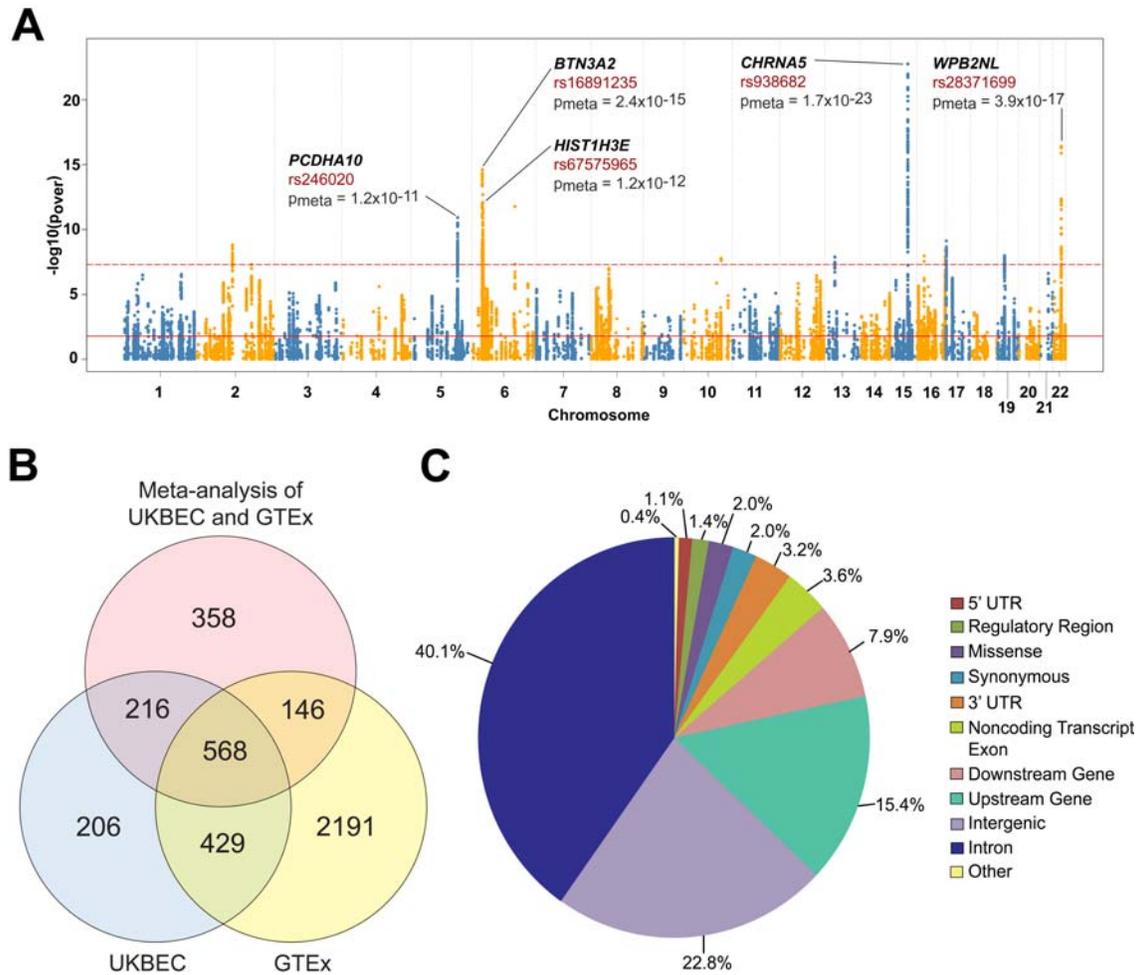


Figure 2 | Manhattan plot and effect predictions of meta-analysed *cis*-eQTLs. *A*, $-\log_{10}(p_{\text{Over}})$ are plotted for *cis*-eQTLs identified in a meta-analysis of four overlapping brain regions between UKBEC and GTEEx (Figure 1, analysis C). Top five protein-coding genes associated with these eQTLs are provided, along with the associated *cis*-eQTL and p_{Over} . Solid red line denotes threshold at which $\text{FDR} = 0.05$. Dashed red line denotes threshold at which $p\text{-value} = 5 \times 10^{-8}$. *B*, Overlap of *cis*-eQTLs between meta-analysis of four brain regions common between UKBEC and GTEEx (Figure 1, analysis C) with $p_{\text{Over}} < 5 \times 10^{-8}$, meta-analysis of those regions in UKBEC alone with $p_{\text{ukbec}} < 5 \times 10^{-8}$ and meta-analysis of those regions in GTEEx alone with $p_{\text{gtex}} < 5 \times 10^{-8}$. *C*, SNP effect prediction (from Ensembl

Variant Effect Predictor) for 1 288 *cis*-eQTLs with $p_{\text{over}} < 5 \times 10^{-8}$. Overlapping brain regions are cerebellum, frontal cortex, hippocampus and putamen.

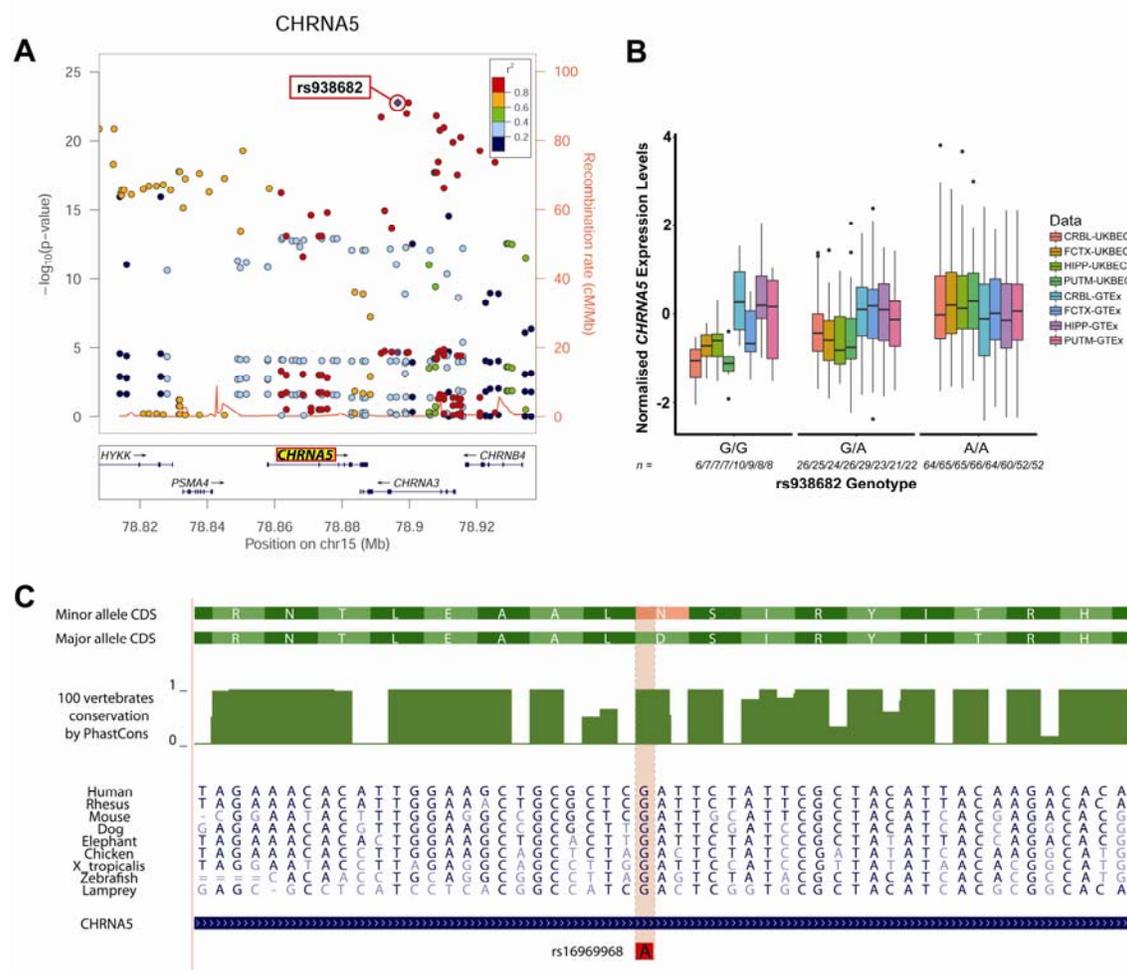


Figure 3 | Mapping and gene expression effects of *CHRNA5* associated eQTLs. *A*, LocusZoom plot for the most significant *cis*-eQTL (rs938682) identified in meta-analysis of UKBEC and GTEx (Figure 1, analysis C, Table 1). This SNP is *cis* to *CHRNA5*, denoted in yellow highlight. Other SNPs that are within the study and *cis* to *CHRNA5* are plotted for their p_{over} (left y-axis) and LD-value (r^2 denoted by color of circle, calculated as relative to rs938682). Recombination rate of genomic region is plotted in red. *B*, Boxplot of gene expression levels (normalised separately for UKBEC and GTEx per region) by genotype of the rs938682 eQTL. Vertical lines for each plot captures data between $-1.5 \times$ interquartile range and $1.5 \times$ interquartile range, with outliers depicted as black circles. The number of individuals with a particular genotype per study-region is denoted in italics. *C*, Amino acid

conservation plot for CHRNA5. Missense mutation resulting from the eQTL rs16969968 is highlighted.

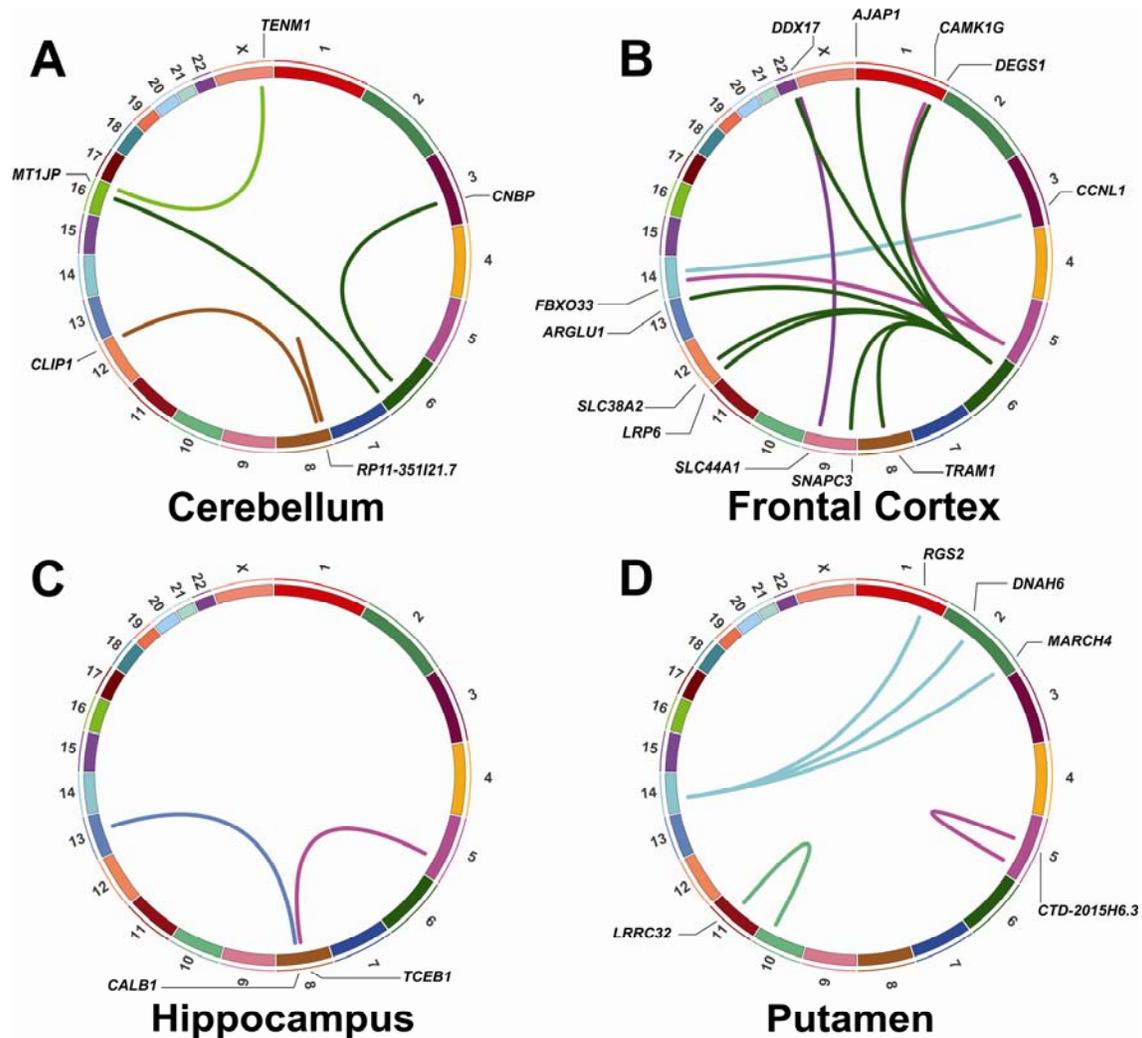


Figure 4 | Circos³⁵ plots of *trans*-eQTLs from meta-analysis by brain region. Locations of *trans*-eQTLs (with $p_{\text{pair}} < 5 \times 10^{-8}$ per overlapping brain region, A-D) and their respective associated genes. Chromosomal location of the eQTL is denoted by the color of the connecting line. For example, within the putamen (D), *trans*-eQTLs are located on chromosome 14 (light blue) and are associated with expression of three genes on chromosomes 1 and 2.

Table 1: Most significant *cis*-eQTL for each associated gene.

Gene	Gene Function	eQTL	Alleles [†]	eQTL Location	$\beta_{\text{over}} \pm \text{se}$	p_{over}	FDR
<i>CHRNA5</i>	Ligand-gated ion channel	rs938682	A/G	Chr 15, <i>CHRNA3</i> intron	0.26 ± 0.03	1.7E-23	5.8E-19
<i>WBP2NL</i>	Sperm-specific WW domain-binding protein	rs28371699	A/C	Chr 22, <i>CYP2D6</i> intron	-0.13 ± 0.02	3.9E-17	6.6E-14
<i>BTN3A2</i>	Immunoglobulin superfamily	rs16891235	T/C	Chr 6, <i>HIST1H1A</i> exon (missense variant)	0.15 ± 0.02	2.4E-15	2.6E-12
<i>ZNF603P</i>	Non-coding RNA	rs200976	A/G	Chr 6, intergenic variant	-0.48 ± 0.06	2.0E-13	1.4E-10
<i>HIST1H3E</i>	Nucleosome component	rs67575965	A/G	Chr 6, intergenic variant	-0.30 ± 0.04	1.2E-12	5.9E-10
<i>C6orf3</i>	Non-coding RNA	rs6912446	T/A	Chr 6, <i>TRAF3IP2</i> intron	0.17 ± 0.02	1.7E-12	7.9E-10
<i>PCDHA10</i>	Calcium-dependent cell-adhesion	rs246020	A/G	Chr 5, intron in multiple genes including <i>PCDHA10</i>	-0.35 ± 0.05	1.2E-11	4.2E-09
<i>SRR</i>	Catalyzes the synthesis of D-serine	rs395566	C/T	Chr 17, <i>SMG6</i> intron	-0.16 ± 0.03	7.6E-10	1.5E-07
<i>SEPT10</i>	Filament-forming cytoskeletal GTPase	rs6752531	G/A	Chr 2, <i>SEPT10</i> intron	0.12 ± 0.02	1.5E-09	2.7E-07
<i>GAS8</i>	Cytoskeletal linker	rs11645893	G/C	Chr 16, <i>FANCA</i> intron	0.08 ± 0.01	4.1E-09	5.5E-07
<i>MAU2</i>	Association of cohesion complex with chromatin during interphase	rs8101499	G/A	Chr 19, downstream of <i>MAU2</i>	0.06 ± 0.01	1.0E-08	9.3E-07
<i>RRN3</i>	Transcription initiation	rs11075255	G/A	Chr 16, intron in <i>RRN3</i> and <i>PDXDC1</i>	-0.12 ± 0.02	1.1E-08	9.4E-07
<i>SMIM2-AS1</i>	Non-coding RNA	rs12869383	A/G	Chr 13, intergenic	-0.52 ± 0.09	1.3E-08	9.8E-07
<i>GSTO2</i>	Glutathione-dependent thiol transferase activity	rs7079338	C/T	Chr 10, <i>GSTO2</i> intron	0.32 ± 0.06	1.6E-08	1.2E-06
<i>ZSCAN31</i>	Transcription factor	rs2142730	T/C	Chr 6, <i>PGBD1</i> intron	-0.34 ± 0.06	2.9E-08	1.7E-06

[†]First allele listed is the reference allele for effect size. For example, compared to G, A has a β of 0.26 ± 0.03 EU per allele with respect to expression levels of *CHRNA5*.

Table 2: Most significant *trans*-eQTL for each associated gene by brain region.

	eQTL	eQTL Location	Gene	Gene Chromosome	$\beta_{\text{pair}} \pm \text{se}$	p_{pair}	FDR*
Cerebellum	rs1883137	6q21	<i>CNBP</i>	3	0.22 ± 0.04	6.7E-09	2.9E-06
	rs1995942	6q27	<i>MT1JP</i>	16	0.35 ± 0.06	9.3E-09	2.9E-06
	rs6601450	8p23.1	<i>RP11-351I21.7</i>	8	-0.42 ± 0.07	2.2E-08	4.3E-06
	rs7839435	8p21.1	<i>CLIP1</i>	12	0.21 ± 0.04	3.8E-08	4.3E-06
	rs4843852	16q24.1	<i>TENM1</i>	X	0.25 ± 0.05	4.9E-08	4.9E-06
Frontal Cortex	rs28895257	6p21.32	<i>DDX17</i>	22	0.21 ± 0.03	3.1E-14	2.6E-11
	rs72795371	5q33.1	<i>CAMK1G</i>	1	0.40 ± 0.07	1.9E-09	1.7E-08
	rs6876723	5q33.1	<i>FBXO33</i>	14	0.26 ± 0.05	7.1E-09	6.0E-08
	rs522308	6p21.32	<i>SLC38A2</i>	12	0.24 ± 0.04	7.8E-09	6.6E-08
	rs9268659	6p21.32	<i>SNAPC3</i>	9	-0.24 ± 0.04	9.4E-09	7.7E-08
	rs5018660	6p21.32	<i>LRP6</i>	12	0.20 ± 0.04	9.5E-09	7.8E-08
	rs9271620	6p21.32	<i>TRAM1</i>	8	0.18 ± 0.03	1.5E-08	1.2E-07
	rs9268458	6p21.32	<i>AJAP1</i>	1	0.20 ± 0.03	2.2E-08	1.6E-07
	rs9271763	6p21.32	<i>ARGLU1</i>	13	0.21 ± 0.04	2.2E-08	1.6E-07
	rs9271480	6p21.32	<i>DEGS1</i>	1	0.26 ± 0.05	2.7E-08	1.8E-07
	rs7142068	14q23.3	<i>CCNL1</i>	3	0.37 ± 0.07	4.1E-08	2.5E-07
	rs145069606	22q13.33	<i>SLC44A1</i>	9	-0.48 ± 0.09	4.8E-08	2.5E-07
Hip	rs7448129	5q23.3	<i>TCEB1</i>	8	0.43 ± 0.08	1.7E-08	2.7E-06
	rs1831189	13q21.33	<i>CALB1</i>	8	-0.35 ± 0.06	3.4E-08	3.0E-06
Putamen	rs11167589	5q33.1	<i>CTD-2015H6.3</i>	5	0.42 ± 0.07	8.3E-10	6.3E-07
	rs57275892	14q13.2	<i>MARCH4</i>	2	-0.28 ± 0.05	1.8E-08	1.4E-06
	rs11817486	10q24.31	<i>LRRC32</i>	11	-0.36 ± 0.06	2.8E-08	1.8E-06
	rs12887279	14q13.2	<i>RGS2</i>	1	-0.29 ± 0.05	3.9E-08	1.9E-06
	rs57725522	14q13.2	<i>DNAH6</i>	2	-0.39 ± 0.07	4.2E-08	1.9E-06

* FDR values were calculated per brain region. *Hip*, Hippocampus.

