The genomic basis of mood instability: identification of 46 loci in 363,705 UK Biobank participants, genetic correlation with psychiatric disorders, and association with gene expression and function.

Joey Ward MSc[1], Elizabeth M. Tunbridge D.Phil[2], Cynthia Sandor PhD[3], Laura M. Lyall PhD [1], Amy Ferguson BSc[1], Rona J. Strawbridge PhD[1,4], Donald M. Lyall PhD[1], Breda Cullen PhD[1], Nicholas Graham MBChB[1], Keira J.A. Johnston MSc[1], Caleb Webber PhD[3,5],Valentina Escott-Price PhD[6], Michael ODonovan PhD [6], Jill P. Pell[1] PhD, Mark E.S. Bailey PhD [7], Paul J. Harrison[2] DM (Oxon), Daniel J. Smith MD[1]

[1]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK, [2]Department of Psychiatry, University of Oxford, UK, [3]UK Dementia Research Institute, Cardiff University, Cardiff, UK, [4]Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden, [5]Department of Physiology, Anatomy and Genetics, Oxford, UK, [6]University of Cardiff, Cardiff, UK, [7]School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.

Corresponding Author: Professor Daniel Smith

Address: Mental Health and Wellbeing, Gartnavel Royal Hospital, 1055 Great Western Road, Glasgow G12 0XH,

Telephone: 0141 211 3930

Email: Daniel.Smith@glasgow.ac.uk

Abstract

Genome-wide association studies (GWAS) of psychiatric phenotypes have tended to focus on categorical diagnoses, but to understand the biology of mental illness it may be more useful to study traits which cut across traditional boundaries. Here we report the results of a GWAS of mood instability (MI) as a trait in a large population cohort (UK Biobank, n=363,705). We also assess the clinical and biological relevance of the findings, including whether genetic associations show enrichment for nervous system pathways. Forty six unique loci associated with MI were identified with a heritability estimate of 9%. Linkage Disequilibrium Score Regression (LDSR) analyses identified genetic correlations with Major Depressive Disorder (MDD), Bipolar Disorder (BD), Schizophrenia (SZ), anxiety and Post Traumatic Stress Disorder (PTSD). Gene-level and gene set analyses identified total 244 significant genes and 6 enriched gene sets. Tissue expression analysis from the SNP level data found enrichment in multiple brain regions, and eQTL analyses highlighted an inversion on chromosome 17 plus two brain-specific eQTLs. Additionally, we used a Phenotype Linkage Network (PLN) analysis and community analysis to assess for enrichment of nervous system gene sets using mouse orthologue databases. The PLN analysis found enrichment in nervous system PLNs for a community containing serotonin and melatonin receptors. In summary, this work has identified novel loci, tissues, and gene sets contributing to MI as a normal trait and will inform future work on the biology of mood and psychotic disorders, and to point the way towards potential for new stratified medicine approaches and the identification of novel trans-diagnostic drug targets.

Introduction

Mood instability (MI) is a subjective emotional state defined as rapid oscillations of intense affect, with difficulty regulating these oscillations and their behavioural consequences[1]. As a psychopathological phenotype, MI may be useful for psychiatric research within a Research Domain Classification (RDoC) framework[2] because it is a symptom that occurs in several psychiatric disorders, particularly major depressive disorder (MDD) and bipolar disorder (BD). It is also present within general population samples, and is known to be associated with a range of adverse health outcomes[3].

We recently identified four loci associated with MI within a subsample of the UK Biobank cohort (N=113,968) and found genetic correlation with both MDD and schizophrenia[4]. Here, we report a significantly larger genome-wide association study (GWAS) of MI in the full UK Biobank dataset (N=363,705), using a BOLT-LMM approach to maximize statistical power. We also revisit the assessment of genetic correlations with psychiatric disorders, including the use of more recent GWAS outputs for MDD, schizophrenia and BD. Furthermore, we contextualize our findings in terms of affected tissues, eQTL analysis and Phenotype Linkage Network (PLN) analysis. PLN is a new methodology that harnesses the fact that variation in many complex traits results from perturbations of multiple molecular components within a smaller number of cellular pathways. These pathways can then be identified with using gene network approaches.

Methods

UK Biobank sample

UK Biobank is a large cohort of over 500,000 United Kingdom residents, aged between 39 and 69 years[5]. UK Biobank was created to study the genetic, environmental and lifestyle factors that cause or prevent a range of morbidities in middle and older age. Baseline assessments occurred over a 4-year period, from 2006 to 2010, across 22 UK centres. These assessments covered a wide range of social, cognitive, lifestyle and physical health measures. Informed consent was obtained from all participants, and this study was conducted under generic approval from the NHS National Research Ethics Service (approval letter dated 13 May 2016, Ref 16/NW/0274) and under UK Biobank approvals for application #6553 'Genome-wide association studies of mental health' (PI Daniel Smith).

Genotyping, imputation and quality control

In March 2018 UK Biobank released genetic data for 487,409 individuals, genotyped using the Affymetrix UK BiLEVE Axiom or the Affymetrix UK Biobank Axiom arrays (Santa Clara, CA, USA) which contain over 95% common SNP content[6]. Pre-imputation quality control, imputation and post-imputation cleaning were conducted centrally by UK Biobank (described in the UK Biobank release documentation).

Phenotyping

UK Biobank participants were asked as part of their baseline assessment: “Does your mood often go up and down?” Those who responded ‘yes’ to this question were defined as MI cases and those who responded ‘no’ were defined as controls. To minimise any impact of psychiatric disorders on observed genetic associations with MI, individuals reporting depression, bipolar disorder, schizophrenia, ‘nervous breakdown’, self-harm or suicide attempt (all from UK Biobank data field 20002), and those who reported taking psychotropic medications (data field 20003) were excluded

from the analysis. Participants were also excluded if: their self-reported sex did not match their genetically determined sex; Biobank had determined them to have sex chromosome aneuploidy; they were considered by UK Biobank to be heterozygous outliers; they were missing over 10% of their genetic data; or they were not in subset classified the British participants of European ancestry.

Genetic association and heritability

Genetic association analysis was performed using BOLT-LMM[7, 8] to control as robustly as possible for population structure while maximising power by avoiding the need to exclude related individuals. This also removes the need to adjust the model for principal genetic components (PGCs). Additionally, BOLT-LMM builds an infinitesimal model including all directly genotyped SNPs simultaneously, thereby further increasing power compared to logistic regression approaches that test each SNP in turn. This 'genotyped SNPs only' model has the imputed SNPs tested against it allowing for the imputation score cut criteria to be substantially reduced and increases the number of SNPs available to test for association with the outcome. Models were adjusted for age, sex and genotyping array. SNPs were filtered to remove those with MAF < 0.01, Hardy-Weinberg Equilibrium $P < 1*10^{-6}$, or imputation quality score < 0.3. BOLT-LMM was also used to provide a heritability estimate and λ_{GC} estimate. The summary statistics were processed by FUMA[9] (http://fuma.ctglab.nl/) for visualisation, MAGMA Gene Analysis, Gene-set Analysis and Tissue Expression Analysis[10]. The Gene-level Analysis operates by grouping p values for individual SNPs into a gene test statistic using the mean chi-sq statistic for the gene whilst accounting for LD via the use of a European ancestry reference panel. The Gene-set Analysis groups genes according to MsigDB v6.1[11], a collection of both curated gene sets and GO terms, and tests each set in turn against all the other sets. The Tissue Expression Analysis performs a one-sided test based on the correlation between tissue-specific gene expression profiles and trait-gene associations.

As FUMA only makes use of the 1000Genomes reference panel, regional plots were made via LocusZoom v1.4[12] as SNPs from the HRC reference panel were also imputed in the UK biobank

genetic data release. We defined a locus as the region of containing a lead SNP and all other SNPs (r^2 >0.1) within a 5MB radius of the lead SNP. The LD was calculated using 10,000 unrelated Biobank participants who had passed the same genetic QC as those used for the GWAS.

Genetic correlations

Linkage Disequilibrium Score Regression (LDSR)[13] was used to calculate genetic correlations with psychiatric disorders. The intercept was left unconstrained to allow for sample overlap. For the MDD[14], BD[15], schizophrenia[15] and PTSD[16] phenotypes, we used the most up-to-date GWAS summary statistics provided by the Psychiatric Genomics Consortium. Anxiety disorder summary statistics came from the Anxiety NeuroGenetics STudy (ANGST) Consortium[17].

Tissue-specific expression and eQTL analysis

The lead SNP for each locus (unless otherwise noted) was assessed for *cis* effects on gene expression (eQTLs) in publicly available human dorsolateral prefrontal cortex RNASeq datasets using the Lieber Institute for Brain Development (LIBD) eQTL browser (See URLs). Each locus was initially examined in the LIIBD BrainSeq dataset (n=738; See URLs); SNPs showing significant eQTLs were then assessed for replication in the Common Mind Consortium (CMC) dataset (n=547; See URLs). Only eQTLs that reached a threshold of p = 0.05 (FDR corrected) in both the LIBD and CMC datasets, and showed the same direction of effect in both, are reported. Tissue-specific expression patterns were assessed for implicated genes using the GTEx portal[18]. All p values quoted in the text are FDR corrected.

Principal component generation

Principal components were created using plink 2[19] using pca approx. (with default settings).

Pathway analysis

PLN analysis builds on the fact that variation in complex traits results from perturbations of multiple molecular components within a smaller number of cellular pathways that can be identified using

gene network approaches. No single dataset or data type can provide a complete picture of the functional association between genes but a recent method combines information from multiple data types by weighing functional similarities between genes according to their likelihood of influencing the same mammalian phenotype(s). This approach has a greater specificity and sensitivity than analyses using a single data type and other comparable integrative methods[20]. The PLN approach exploits phenotypic information from over seven thousand genes whose function has been experimentally perturbed in the mouse and evaluates the ability of different data types such as protein-protein interactions (PPI), co-expression (RNA or protein), and semantic similarity score based on literature or Gene Ontology (GO) annotations or pathway annotations (KEGG), to predict whether knockout of the orthologues of a given pair of human genes will yield similar phenotypes. By weighting those data types accordingly, they are integrated to generate a single combined measure of functional similarity between gene pairs. The resulting network of pairwise gene functional similarities is termed a phenotypic-linkage network (PLN)[20].To increase the sensitivity and specificity to detect functional associations relevant for a specific disease/trait, it is possible to select only those mouse phenotypes that are relevant for a specific disorder in the data type weighting evaluation step[21]. Following this approach, we re-weighted our generic PLN to be more sensitive to functional genomics data most informative to MI by considering only phenotypes within the over-arching mouse phenotype ontology (MPO) category *Nervous System* (MP:0003631). The PLN and nervous-PLN were built using the same 16 functional genomics datasets described by Honti et al[20], with 64,640,972 and 49,656,123 weighted links respectively.

Following the approach described by Sandor et al., we identified 'communities' of densely interconnected groups of genes (including at least 20 genes) within each PLN and tested whether any communities were enriched in genes harboured by GWA/subGWA intervals. This test examines how many of these intervals harboured at least one gene belonging to a given Community as compared to randomly shifted intervals equal in gene number. This approach makes no prior hypothesis about the number or nature of genes within each GWA interval.

Definition of GWA and subGWA intervals

The GWAS and subGWA intervals were defined by considering SNPs attaining an association p-value of $5x10^{-8}$ and $1x10^{-6}$ identifying 6375 and 9358 SNPs, respectively. We then identified the haplotypic block within which each SNPs using genotypes in the 1000 Genome Project and the python pipeline developed by Brent Pedersen (https://gist.github.com/brentp/5050522). We defined GWA/sub GWA intervals by identifying the most distant block on a chromosome within a region of 500Kb of the lead SNP. We then added an additional 300 kb on either side of the interval to include genes that may be regulated by regulatory variants with effects captured by the lead SNPs. For subGWA regions, we excluded those subGWA intervals harbouring genes present in GWA intervals.

Results

Demographics

In the GWAS sample of 363,705 individuals, 43.2% reported MI (n=157,039) and the rest did not (n=206,666). There was a higher proportion of females in MI cases than in controls (55.4% versus 51.2% respectively), and the mean age of cases was lower than controls (55.8 years versus 57.7 years).

GWAS findings

We detected 46 loci across the genome with $p < 5*10^{-8}$ (Figure 1 and Table S1) and an estimated SNP heritability (H_2) of 0.09 (S.E.=0.02). The distribution of test statistics was consistent with a polygenic contribution to risk (λ_{GC}= 1.21; λ_{1000} = 1.001; LDSR intercept = 1.041; SE = 0.006).

Gene-level and Gene Set Analysis

244 significant genes were detected by MAGMA (Supplementary Table S2) and FUMA gene analysis. The Gene Set Analysis returned 6 enriched gene sets that met the threshold for significance after Bonferroni correction (Supplementary Table S3). Of these, 4 sets were related to brain development and differentiation of neurons, glial cells and astrocytes or neurogenesis. Other enriched sets included the Nikolsky breast cancer 16q24 amplicon genes and the prepulse inhibition gene sets.

Tissue expression analysis

MAGMA tissue expression analysis identified 11 tissue categories, all of which were in the brain (Figure S1). Indeed, all sampled brain areas except substantia nigra showed enrichment (i.e. frontal and anterior cingulate cortex, basal ganglia, hippocampus, amygdala, hypothalamus and cerebellum).

Genetic correlations

Genetic correlations were calculated between MI and the five psychiatric phenotypes of interest (Table 1). All genetic correlations remained significant after False Discovery Rate (FDR) correction (Q <0.05). The largest correlations were with MDD (r_g=0.66, p=1.28*$10^{-35}$and anxiety (r_g=0.64, p=3.23*10^{-6}). PTSD had a moderate correlation with MI (rg=0.32, p=1.01*10^{-6}) and both schizophrenia and bipolar disorder had weak but significant correlations (SCZ r_g=0.14, p=1.01*10^{-6}, BD r_g=0.09, p=0.003).

eQTL analysis

Nine of the GWAS loci showed significant eQTLs (Supplementary Table S4). The strongest evidence of association with expression levels was for rs669915, a eQTL located within a region of strong linkage disequilibrium (LD) in chromosome 17q21 resulting from the existence of a 900kb inversion polymorphism that is common in European populations[22]. The extended region of LD across this portion of the chromosome makes it challenging to identify causal SNPs or the genes they regulate. The rs669915 eQTL was most strongly associated with expression of *LRRC37A4P* (LIBD dataset minimum p=1.96 x 10^{-99}; CMC dataset p=3.99 x 10^{-65}), an expressed pseudogene, but there are many alternative candidates for genes regulated by this SNP, including MAPT and CRHR1, for which it was also an eQTL. (Supplementary Table S4).

The chromosome 17q21 inversion polymorphism has itself been reported to affect the expression of genes in this region[23]. We therefore investigated whether rs669915 might 'tag' the expression effects mediated by the inversion polymorphism in our sample. Using the method of de Jong and colleagues, we constructed genetic principal components (GPCs) from SNPs within the region between base positions 40,850,001 and 41,850,000 on chromosome 17. A plot of the first two PCs is shown in (Figure S2) and reveals three distinct clusters of individuals, each representing one of the

three inversion polymorphism genotypes, H1/H1 (right-most cluster; n=162,113), H1/H2 (middle cluster; n= 158,506) and H2/H2 (left-most cluster; n= 38,597). The H1 inversion allele had a population frequency of 0.32, far higher than the frequency reported by de Jong. In linear regression analyses, there was no association between MI phenotype and inversion genotype using a model of additive allelic effects (no. of H2 alleles) and adjusting for age, sex and genotyping array (p=0.835).

Many of the eQTL-positive loci showed associations with specific mRNA isoforms. Thus, rs763646 predicted expression of a specific junction in *EXOC4* (LIBD dataset p=0.001; CMC dataset p=0.03). Similarly, rs1962104 predicted expression of a splice junction and specific exons of *PTK2* (LIBD dataset minimum p=6.02 x 10^{-7}; CMC dataset p=0.0007). Rs11039182 predicted expression of a novel junction within *SLC39A13* (LIBD dataset p=8.20 x 10^{-11}; CMC dataset p=4.62 x 10^{-12}). *SLC39A13* shows moderate expression across human tissues. Rs2898260 predicted expression of a specific start site in *XKR6* (LIBD dataset p=4.25 x 10^{-5}; CMC dataset p=0.00002), and specific junctions in *TDH* (LIBD dataset p=0.0001; CMC dataset p=0.0009) and *SOX7* (LIBD dataset p = 0.0008; CMC dataset p=0.0004), all of which are expressed at low levels in human tissues.

Two loci predicted expression of genes whose function is largely unknown but which show greater abundance in brain compared with other human tissues. Specifically, rs2729940 predicts expression of *RP11-481A20.10* (LIBD dataset minimum p = 0.01; CMC dataset p=0.0008) and *RP11-481A20.11* (LIBD dataset minimum p=0.0003; CMC dataset p=0.00003), rs7818437 predicts expression of *RP11-981G7.1* (LIBD dataset minimum p=2.38 x 10^{-7}; CMC dataset p=0.004), which is uniquely expressed in the brain.

Other genes implicated by our eQTL analysis include *SPINK9*, associated with rs6889822 (LIBD dataset minimum p=0.0002; CMC dataset p=8.99 x 10^{-7}), which is expressed at low levels, but is more abundant in the brain than other tissues. Expression of *FAM86B3P* (LIBD dataset minimum p=6.79 x 10^{-11}; CMC dataset p=8.02 x 10^{-7}), *FAM85B* (LIBD dataset minimum p=7.65 x 10^{-9}; CMC dataset p=0.001) and *ALG1L13P* (LIBD dataset minimum p=2.28 x 10^{-5}; CMC dataset p=0.004) was

associated with rs4398922. All three genes show low but consistent expression across different human tissues.

Nervous system-PLN analyses

Amongst both GWA and subGWA gene sets, we found a disproportionate aggregation of genes within only one community, Community 26 within the NS-PLN (21 GWAS loci including at least one gene, q=0.011; 25 “subGWAS “loci including at least one, q=0.018) (**Fig 2 A**). Examining the entire NS-PLN Community 26 gene, we found that it was significantly enriched in genes, whose unique 1:1 orthologues in the mouse when disrupted induce abnormities in synaptic transmission (Mouse Phenotype Ontology term MP:0003635; $q=2.77e^{-118}$, 75 genes expected vs 259 gene observed). However, we did not find evidence that the unique mouse orthologues of MI GWA and subGWA genes that belonged to Community 26 were enriched for any particular mouse phenotype. Nonetheless, we found that the 37 and 35 GWA and subGWA genes present in the Community 26 were highly functionally connected with other Community 26 genes annotated with abnormal synaptic transmission phenotype term (Figure 2 B).

Discussion

Main findings

These analyses represent the largest genetic study of MI to date. Forty six unique loci associated with MI were identified, with a heritability estimate of 9%. Our findings confirm the four loci identified in our initial GWAS on the UK Biobank interim data release[4] and are further validated by tissue expression analyses (enrichment for 11 brain regions) and pathway analyses (6 enrichment pathways, 4 of which relate to the development and differentiation of neurons). The large number of individuals in this study also provided substantial power to detect genetic correlations with psychiatric traits via LDSR. All six psychiatric traits assessed had a significant genetic correlation with MI. Some of these correlations were strong (particularly for MDD and anxiety) but others were weaker than expected: the genetic correlation between MI and BD was only 9%, perhaps suggesting that the MI phenotype in this study differs from the affective instability that is a core feature of BD.

Biology of mood instability

Loci associated with MI included genes that are involved across a variety of biochemical pathways, as well as brain development and function. For example, several gene products localised to the synapse. *PLCL1* and *PLCL2* are involved in GABA signalling[24] and melatonin signalling respectively, and *RAPSN* assists in anchoring nicotinic acetylcholine receptors at synaptic sites[25]. *PLCL1* has already been identified in a GWAS of schizophrenia[26] and *PLCL2* has been shown to be upregulated in bipolar disorder[27]. Additionally, we identified *CALB2* which has many biological functions, including a role in modulating neuronal excitability[28]. Both *DCC* (identified in the previous MI GWAS) and *BSN* facilitate the release of neurotransmitters within the active zone of some axons[29]. *BSN* has also been shown to be associated with schizoaffective disorder via GABA signalling[30]. *FARP1* promotes dendritic growth[31] and, although it has so far not been directly linked to mental health outcomes, it has been shown to regulate dendritic complexity[32]; reduced dendritic complexity is recognised as a feature of schizophrenia[33].

We identified several developmental genes, including *NEGR1*[34], *RARB*[35] and *EPHB1*[36], and transcription factors such as *HIVEP2* (loss of function of which causes intellectual disability[37]) and *TCF4* (previously associated with schizophrenia[38]). *NEGR1* was identified by 23andMe within their GWAS of MDD[39] and increased levels of NEGR1 protein in spinal fluid have been identified in both MDD and BD[40]. *RARB* is involved in retinoic acid synthesis pathways that have been associated depressive symptoms in mice[41] and has also been found to have increased expression in patients with schizophrenia[42]. The methylation state of the *EPHB1* gene has been linked to MDD[43] and SNP-based analyses have identified association between *EPHB1*'s and symptoms of schizophrenia[44].

We also found association with several genes involved in mitochondrial energy production, such as *NDUFAF3*, *NDUFS3*, *PTPMT1*, *KBTBD4* and *MTCH2*, suggesting that part of the physiology of MI may relate to energy dysregulation.

In addition to protein coding genes, several loci were identified in regions containing non-coding protein sequences such as *AC019330.1*, *AC133680.1*, *RP11-6N13.1* and *RP11-436d23.1*. Additionally, eQTL analyses identified three more possible non-coding genes (*RP11-481A20.10*, *RP11-481A20.11* and *FAM85B*) suggesting a possible RNA interference or post-transcriptional regulation basis to MI.

Furthermore, the eQTL analyses highlighted the 17q21 inversion. Our principal component analysis of this region did not detect a significant association leading us to conclude that it is the SNPs in the region (not the inversion itself) driving the association. It is possible that lead SNPs may tag, enhancer RNA or eRNA which we were unable to detect here, as LIBD data are based on poly A-selected mRNA. However, our findings are consistent with the association of dopamine neurons with *LRRC37A4P* reported recently[45].

Genes within regions associated with MI were functionally associated with synaptic transmission, a key pathway for psychiatric disorders, albeit this functional association was only detectable after focussing our gene network towards data types most informative for mammalian nervous system

phenotypes. Among the genes lying within associated loci that contribute to this functional association are several interesting candidate genes. *HTR4* is a member of the family of serotonin receptors, G protein coupled receptors that stimulate cAMP production in response to serotonin (5-hydroxytryptamine). *MCHR1*, melanin concentrating hormone receptor 1, is a G protein-coupled which binds melanin-concentrating hormone. *MCHR1* can inhibit cAMP accumulation and stimulate intracellular calcium flux, and may be involved in the neuronal regulation of food consumption[46] and this locus showed association with schizophrenia in a Danish sample[47].

Strengths and Limitations

As noted above, this is the largest GWAS of a MI phenotype to date and has successfully identified new loci, eQTLs, genetic correlations and gene network enrichments. However, there are several limitations, most notably the use of a single question to define MI, and the lack of objective verification of the phenotype. Similarly, exclusions for psychiatric disorder were based on self-report. Nevertheless, this approach has previously identified robust associations with a range of health outcomes and disorders[1, 3].

It is also important to note that direct links between genetic risk loci and network constituents in the PLN analysis will have to await the release of more completely annotated gene databases. The incompleteness of phenotypic annotations is likely to explain why the genes identified in the PLN analysis don't have corresponding organismal or physiological phenotypes, but the fact that there were strong functional associations between the genes in the network we detected and mouse orthologues that have the synaptic transmission phenotype annotation suggests that the MI genes will also reveal this phenotype when more completely annotated databases become available.

We also note the large difference in frequencies of the inversion polymorphism on chromosome 17 from that reported by De Jong[23]. This difference could be due to the populations sampled to estimate the frequency or just over prevalence in those who joined UK Biobank. It is however

important to note that the inversion itself, whether acting under any of the models tested would contribute such a small fraction of MI phenotype that even larger sample sizes than used here would be needed to detect an significant correlation.

Conclusion

In summary, with a tripling in sample size from the previous GWAS, we identified substantially more associations with MI in the UK Biobank cohort[4]. This has allowed us to more confidently place these findings within a relevant biological context. Future analyses of the precise roles that the associations reported here play in the clinical expression of MI will likely be relevant for a wide range of psychiatric phenotypes and we anticipate that our findings will stimulate further research on the biology and treatment of MI across a range of mood and psychotic disorders.

Conflict of interest

The authors have no conflicts of interest to declare.

URLs

LIBD website - http://eqtl.brainseq.org/

LIBD eQTL Browser phase 1 - http://eqtl.brainseq.org/phase1/eqtl/

CommonMind Consortium public–private partnership http://commonmind.org/WP

Acknowledgements

Data were generated as part of the CommonMind Consortium supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916, P50M096891, P50MH084053S1, R37MH057881 and R37MH057881S1, HHSN271201300031C, AG02219, AG05138 and MH06692. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories and the NIMH Human Brain Collection Core. CMC Leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-Gyu Hahn (University of Pennsylvania), Keisuke Hirai, Hiroyoshi Toyoshiba (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent Essioux (F. Hoffman-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks), Thomas Lehner, Barbara Lipska (NIMH). JW is supported by the JMAS Sim Fellowship for depression research from the Royal College of Physicians of Edinburgh (173558). AF is supported by an MRC Doctoral Training Programme Studentship at the University of Glasgow (MR/K501335/1). RJS is supported by a UKRI Innovation- HDR-UK Fellowship (MR/S003061/1). KJAJ is supported by an MRC Doctoral Training Programme Studentship at the Universities of Glasgow and Edinburgh. DJS acknowledges the support of a Lister Prize Fellowship (173096) and the MRC Mental Health Data Pathfinder Award (MC_PC_17217).

References

1. Marwaha S, He Z, Broome M, Singh SP, Scott J, Eyden J *et al.* How is affective instability defined and measured? A systematic review. *Psychol Med* 2014; **44**(9)**:** 1793-1808.

2. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* 2013; **11:** 126.

3. Broome MR, Saunders KEA, Harrison PJ, Marwaha S. Mood instability: significance, definition and measurement. *The British journal of psychiatry : the journal of mental science* 2015; **207**(4)**:** 283-285.

4. Ward J, Strawbridge RJ, Bailey MES, Graham N, Ferguson A, Lyall DM *et al.* Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia. *Translational Psychiatry* 2017; **7**(11)**:** 1264.

5. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**(3)**:** e1001779.

6. Biobank U. Genotyping of 500,000 UK Biobank participants. Description of sample processing workflow and preparation of DNA for genotyping. 2015;11 September 2015.

7. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015; **47**(3)**:** 284-290.

8. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nature Genetics* 2018.

9. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* 2017; **8**(1)**:** 1826.

10. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology* 2015; **11**(4)**:** e1004219.

11. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 2005; **102**(43)**:** 15545.

12. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England)* 2010; **26**(18)**:** 2336-2337.

13. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**(3)**:** 291-295.

14. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018; **50**(5)**:** 668-681.

15. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* 2018; **173**(7)**:** 1705-1715.e1716.

16. Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE *et al.* Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry* 2018; **23**(3)**:** 666-673.

17. Otowa T, Hek K, Lee M, Byrne EM, Mirza SS, Nivard MG *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular psychiatry* 2016; **21**(10)**:** 1391-1399.

18. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nature genetics* 2013; **45**(6)**:** 580-585.

19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira Manuel A R, Bender D *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics* 2007; **81**(3)**:** 559-575.

20. Honti F, Meader S, Webber C. Unbiased functional clustering of gene variants with a phenotypic-linkage network. *PLoS Comput Biol* 2014; **10**(8)**:** e1003815.

21. Sandor C, Beer NL, Webber C. Diverse type 2 diabetes genetic risk factors functionally converge in a phenotype-focused gene network. *PLoS Comput Biol* 2017; **13**(10)**:** e1005816.

22. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J *et al.* A common inversion under selection in Europeans. *Nature Genetics* 2005; **37:** 129.

23. de Jong S, Chepelev I, Janson E, Strengman E, van den Berg LH, Veldink JH *et al.* Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. *BMC Genomics* 2012; **13:** 458.

24. Kanematsu T, Jang I-S, Yamaguchi T, Nagahama H, Yoshimura K, Hidaka K *et al.* Role of the PLC-related, catalytically inactive protein p130 in GABA(A) receptor function. *The EMBO journal* 2002; **21**(5)**:** 1004-1011.

25. Muller JS, Baumeister SK, Rasic VM, Krause S, Todorovic S, Kugler K *et al.* Impaired receptor clustering in congenital myasthenic syndrome with novel RAPSN mutations. *Neurology* 2006; **67**(7)**:** 1159-1164.

26. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**(7510)**:** 421-427.

27. Nakatani N, Hattori E, Ohnishi T, Dean B, Iwayama Y, Matsumoto I *et al.* Genome-wide expression analysis detects eight genes with robust alterations specific to bipolar I disorder: relevance to neuronal network perturbation. *Human Molecular Genetics* 2006; **15**(12)**:** 1949-1962.

28. Camp AJ, Wijesinghe R. Calretinin: modulator of neuronal excitability. *Int J Biochem Cell Biol* 2009; **41**(11)**:** 2118-2121.

29. Winter C, tom Dieck S, Boeckers TM, Bockmann J, Kampf U, Sanmarti-Vila L *et al.* The presynaptic cytomatrix protein Bassoon: sequence and chromosomal localization of the human BSN gene. *Genomics* 1999; **57**(3)**:** 389-397.

30. Hamshere ML, Green EK, Jones IR, Jones L, Moskvina V, Kirov G *et al.* Genetic utility of broadly defined bipolar schizoaffective disorder as a diagnostic concept. *Br J Psychiatry* 2009; **195**(1)**:** 23-29.

31. Zhuang B, Su YS, Sockanathan S. FARP1 promotes the dendritic growth of spinal motor neuron subtypes through transmembrane Semaphorin6A and PlexinA4 signaling. *Neuron* 2009; **61**(3)**:** 359-372.

32. Cheadle L, Biederer T. Activity-Dependent Regulation of Dendritic Complexity by Semaphorin 3A through Farp1. *The Journal of Neuroscience* 2014; **34**(23)**:** 7999.

33. Broadbelt K, Byne W, Jones LB. Evidence for a decrease in basilar dendrites of pyramidal cells in schizophrenic medial prefrontal cortex. *Schizophrenia Research* 2002; **58**(1)**:** 75-81.

34. Takita J, Chen Y, Okubo J, Sanada M, Adachi M, Ohki K *et al.* Aberrations of NEGR1 on 1p31 and MYEOV on 11q13 in neuroblastoma. *Cancer Sci* 2011; **102**(9)**:** 1645-1650.

35. van der Wees J, Schilthuis JG, Koster CH, Diesveld-Schipper H, Folkers GE, van der Saag PT *et al.* Inhibition of retinoic acid receptor-mediated signalling alters positional identity in the developing hindbrain. *Development* 1998; **125**(3)**:** 545-556.

36. Wilkinson DG. Multiple roles of EPH receptors and ephrins in neural development. *Nat Rev Neurosci* 2001; **2**(3)**:** 155-164.

37. Srivastava S, Engels H, Schanze I, Cremer K, Wieland T, Menzel M *et al.* Loss-of-function variants in HIVEP2 are a cause of intellectual disability. *European Journal of Human Genetics* 2016; **24**(4)**:** 556-561.

38. Williams HJ, Moskvina V, Smith RL, Dwyer S, Russo G, Owen MJ *et al.* Association between TCF4 and schizophrenia does not exert its effect by common nonsynonymous variation or by influencing cis-acting regulation of mRNA expression in adult human brain. *Am J Med Genet B Neuropsychiatr Genet* 2011; **156b**(7)**:** 781-784.

39. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature genetics* 2016; **48**(9)**:** 1031-1036.

40. Maccarrone G, Ditzen C, Yassouridis A, Rewerts C, Uhr M, Uhlen M *et al.* Psychiatric patient stratification using biosignatures based on cerebrospinal fluid protein expression clusters. *Journal of Psychiatric Research* 2013; **47**(11)**:** 1572-1580.

41. O'Reilly KC, Shumake J, Gonzalez-Lima F, Lane MA, Bailey SJ. Chronic Administration of 13-Cis-Retinoic Acid Increases Depression-Related Behavior in Mice. *Neuropsychopharmacology* 2006; **31:** 1919.

42. Tsai SY, Catts VS, Fullerton JM, Corley SM, Fillman SG, Weickert CS. Nuclear Receptors and Neuroinflammation in Schizophrenia. *Molecular Neuropsychiatry* 2017; **3**(4)**:** 181-191.

43. Davies MN, Krause L, Bell JT, Gao F, Ward KJ, Wu H *et al.* Hypermethylation in the ZBTB20 gene is associated with major depressive disorder. *Genome Biology* 2014; **15**(4)**:** R56-R56.

44. Su L, Ling W, Jiang J, Hu J, Fan J, Guo X *et al.* Association of EPHB1 rs11918092 and EFNB2 rs9520087 with psychopathological symptoms of schizophrenia in Chinese Zhuang and Han populations. *Asia Pac Psychiatry* 2016; **8**(4)**:** 306-308.

45. Dong X, Liao Z, Gritsch D, Hadzhiev Y, Bai Y, Locascio JJ *et al.* Enhancers active in dopamine neurons are a primary link between genetic variation and neuropsychiatric disease. *Nature Neuroscience* 2018; **21**(10)**:** 1482-1492.

46. Fontaine-Bisson B, Thorburn J, Gregory A, Zhang H, Sun G. Melanin-concentrating hormone receptor 1 polymorphisms are associated with components of energy balance in the Complex Diseases in the Newfoundland Population: Environment and Genetics (CODING) study. *Am J Clin Nutr* 2014; **99**(2)**:** 384-391.

47. Demontis D, Nyegaard M, Christensen JH, Severinsen J, Hedemand A, Hansen T *et al.* The gene encoding the melanin-concentrating hormone receptor 1 is associated with schizophrenia in a Danish case-control sample. *Psychiatr Genet* 2012; **22**(2)**:** 62-69.

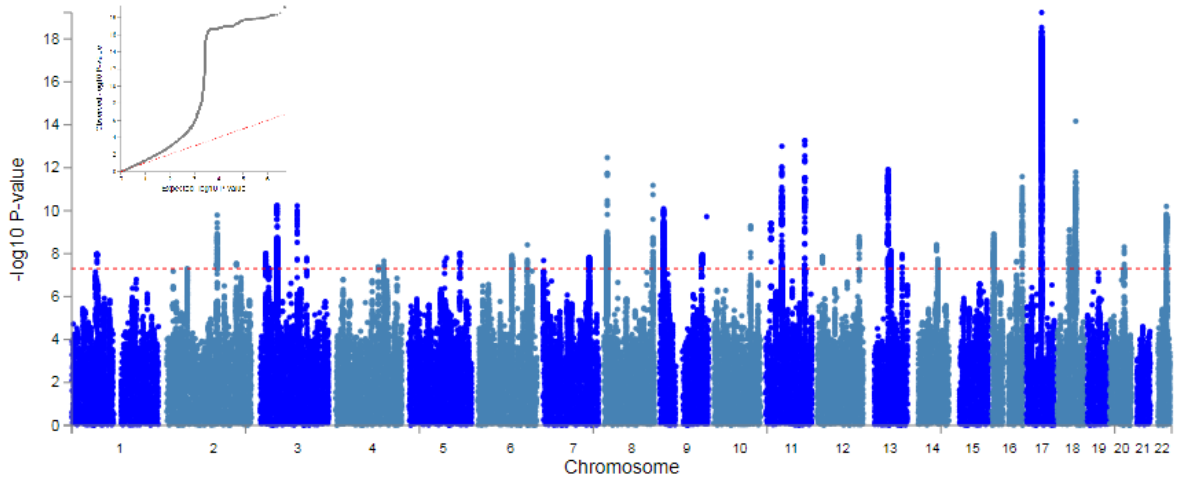

Figure 1 Manhattan and QQ plot of mood instability GWAS

Supplementary Table S1 genomic Loci associated with MI. SNP = lead SNP for the Peak, CHR = chromosome number, BP = Base position of lead SNP, A1 = minor frequency allele, A2 = other allele, Beta = coefficient for the lead SNP, SE = Standard Error for A1 allele, P = p value, start_BP = start of associated region, stop_bp = end of associated region.

Supplementary Table S2 genes identified by MAGMA, geneName = the name of the gene, strand = the direction of the gene on the chromosome, txStart = the base position where the gene starts, txEnd = the base position where the gene ends.

Supplementary Table S3, significant pathways. Gene Set = the gene set that is significant, N genes = the number of genes in the gene set, Beta = the coefficient of the genes in the gene set, Beta STD= the standardised coefficient of the genes in the gene set, SE = The standard error of the genes in the gene set, P = p value, Pbon = the bonferoni corrected p value

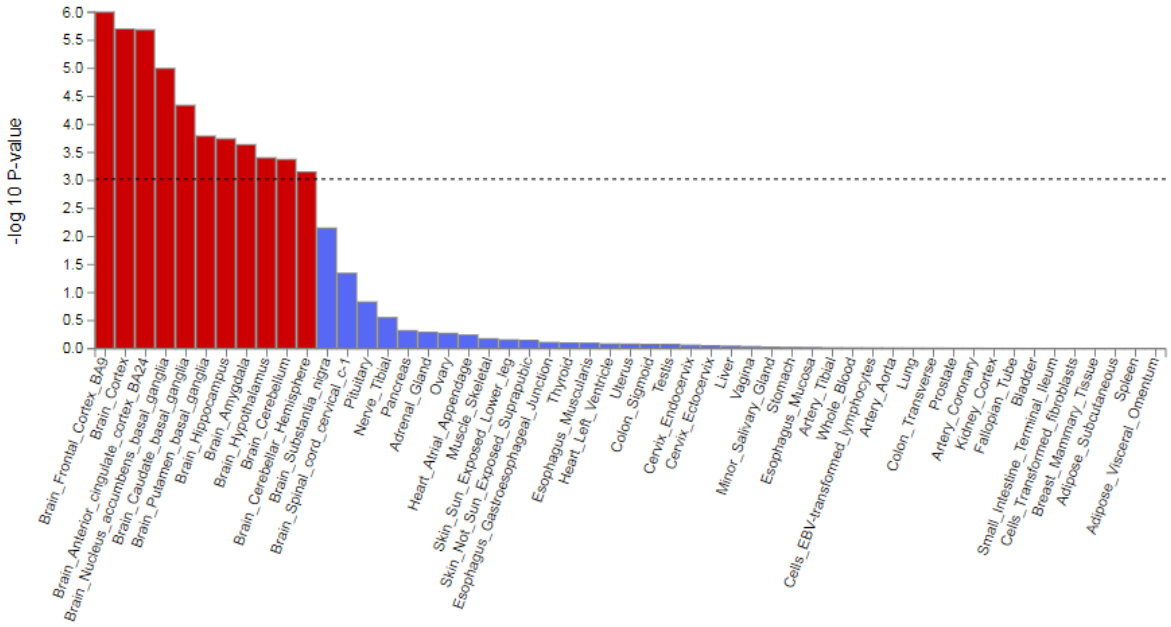

Figure S1 MAGMA tissue expression analysis.

Trait	Rg	se	z	P	Q
MDD	0.74	0.03	26.7	$1.28*10^{-35}$	$6.40*10^{-35}$
Anxiety	0.64	0.14	4.7	$3.23*10^{-6}$	$5.38*10^{-6}$
PTSD	0.32	0.13	2.5	$1.12*10^{-2}$	$1.12*10^{-2}$
Schizophrenia	0.14	0.03	4.4	$1.01*10^{-6}$	$2.53*10^{-6}$
Bipolar Disorder	0.09	0.04	2.5	$2.8*10^{-3}$	$3.5*10^{-3}$

Table 1 Genetic correlations of mood instability with psychiatric phenotypes. Rg genetic correlation with mood instability, SE standard error of the genetic correlation, Z the test statistic, P the p value, Q the False discovery rate corrected p value. MDD major depressive disorder, PTSD post-traumatic stress disorder.

Supplementary Table S4. eQTL analysis.

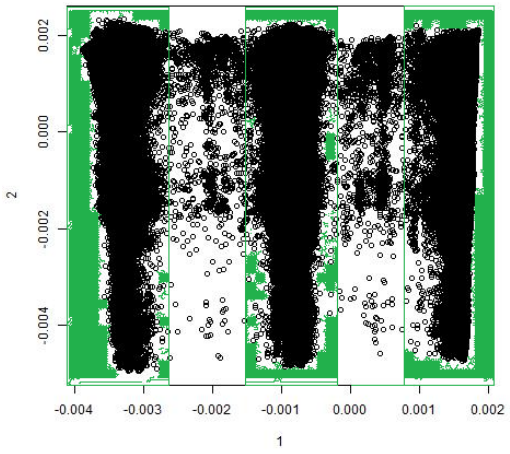

Figure S2, Analysis of chromosome 17q21.31 inversion polymorphism genotype using genetic principal components. PCs 1 and 2 are plotted, calculated using SNP data from individuals used in the GWAS. Individuals were assigned to inversion genotypes if they fell within regions defined by the green boxes, otherwise they were excluded from the association analysis. The three green regions from left to right represent H2/H2 homozygotes, H1/H2 heterozygotes and H1/H1 homozygotes, respectively. Data points lying between the green regions probably represent genotyping errors or rare intra-haplotypic recombinant individuals.

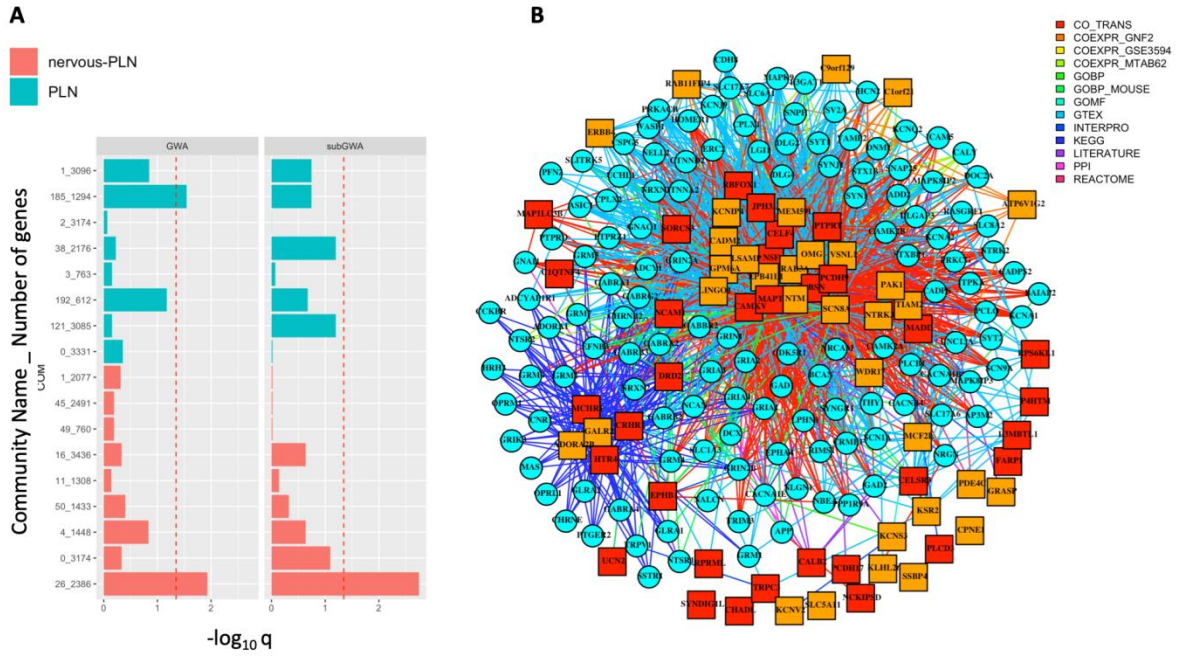

Figure. 2 Different Mood associated genetic risk variants converge in a nervous specific gene network

A Enrichments of gene functional communities from a generic PLN and from a Nervous-System (NS) PLN within Mood-GWA and subGWA loci (see Methods). The Community ID is given first in the descriptor followed by the number of genes within that community. Only communities formed from over 20 genes are shown. (B) Gene subnetwork of Community 26 from NS-PLN showing functional associations between genes residing in Mood-associated GWA (red squares) and subGWA (orange squares) intervals and genes whose unique mouse orthologues are annotated with *abnormal synaptic transmission phenotype* (cyan squares). To increase clarity, only genes with *abnormal synaptic transmission phenotype* annotation with at least three functional links to genes residing in GWA and subGWA regions are shown. The colour of the link connecting two genes indicates the strongest information source supporting the functional association.

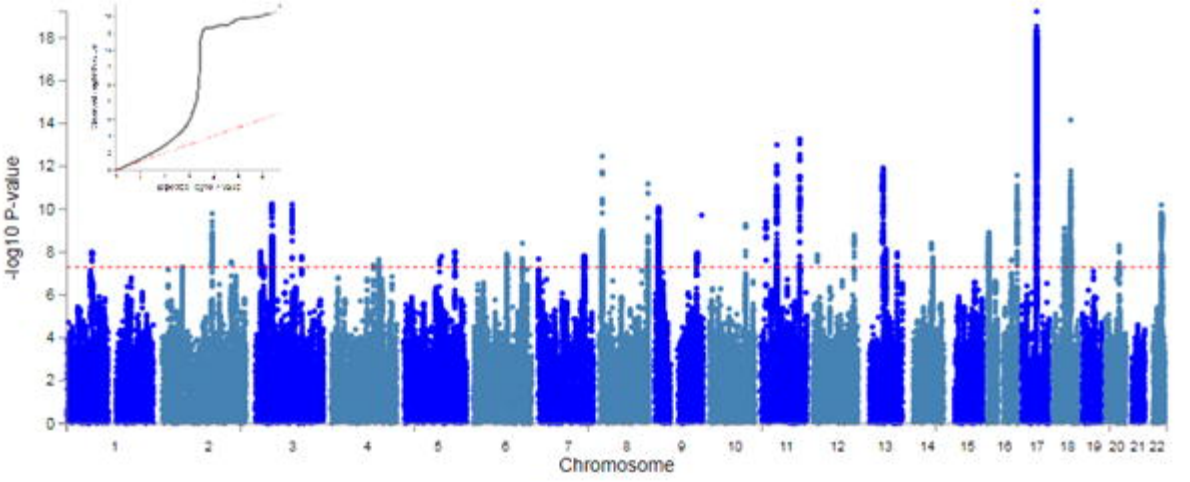
-log10 P-value
Chromosome
0
2
4
6
8
10
12
14
16
18
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Figure 1.

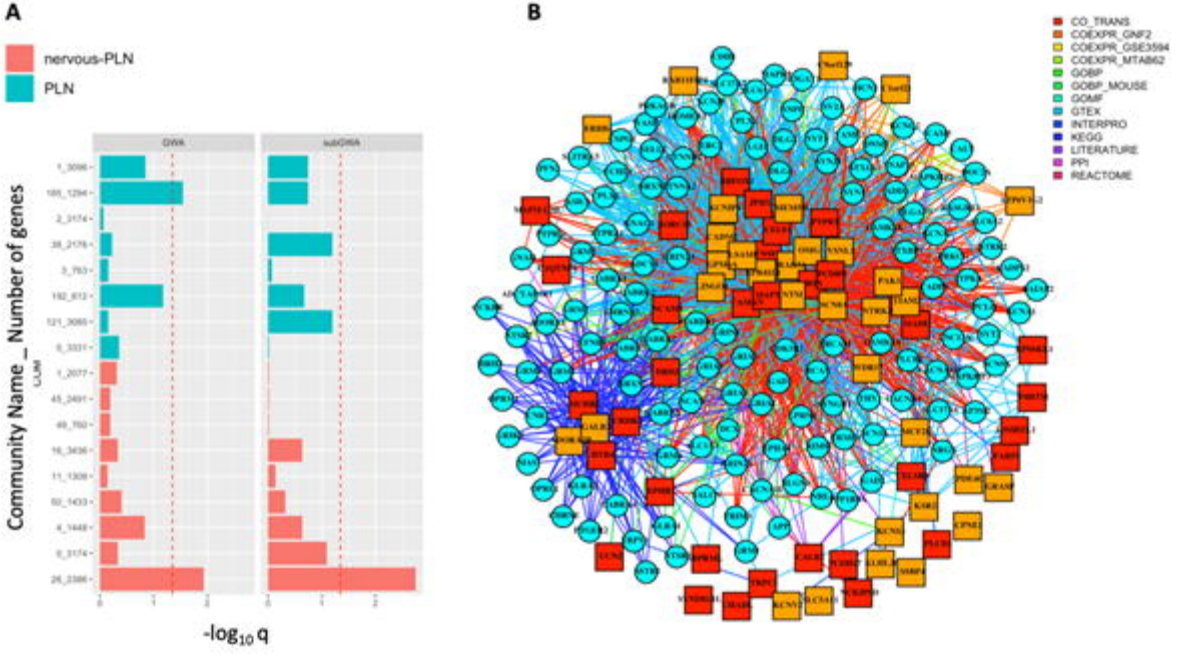
A
nervous-PLN
PLN
GWA
subGWA
Community Name _ Number of genes
-log10 q
B
CO_TRANS
COEXPR_GNF2
COEXPR_GSE3594
COEXPR_MTAB62
GOBP
GOBP_MOUSE
GOMF
GTEX
INTERPRO
KEGG
LITERATURE
PPI
REACTOME

Figure 2.

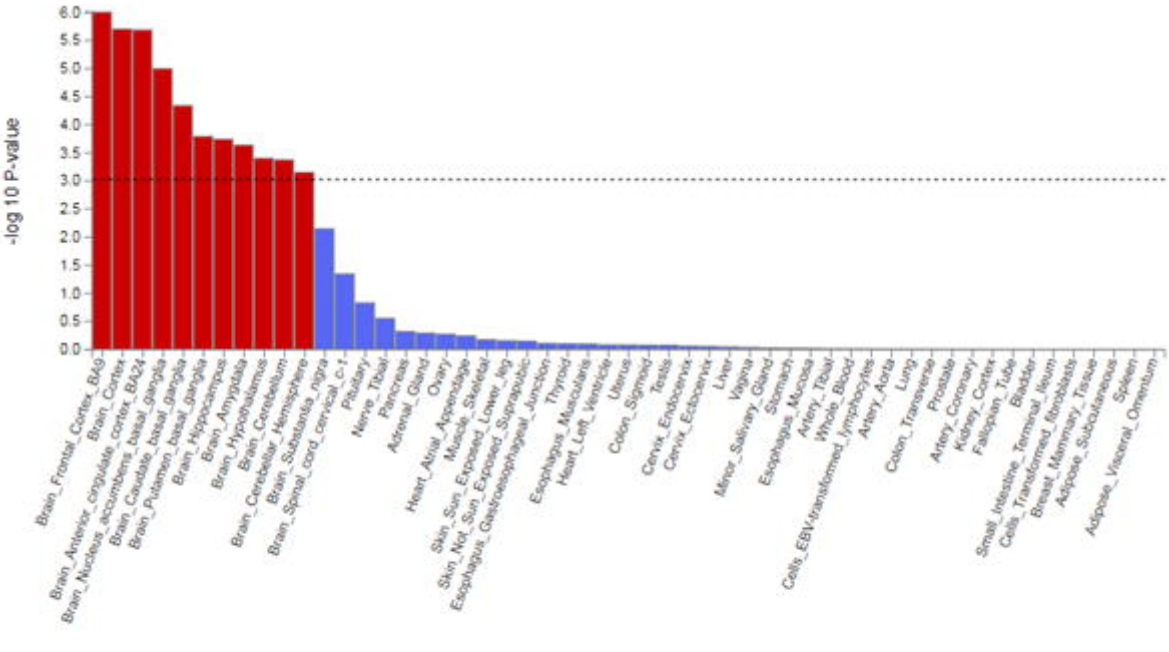
-log 10 P-value
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Brain_Frontal_Cortex_BA9
Brain_Cortex
Brain_Anterior_cingulate_cortex_BA24
Brain_Nucleus_accumbens_basal_ganglia
Brain_Caudate_basal_ganglia
Brain_Putamen_basal_ganglia
Brain_Hippocampus
Brain_Amygdala
Brain_Hypothalamus
Brain_Cerebellum
Brain_Cerebellar_Hemisphere
Brain_Substantia_nigra
Brain_Spinal_cord_cervical_c-1
Pituitary
Nerve_Tibial
Pancreas
Adrenal_Gland
Ovary
Heart_Atrial_Appendage
Muscle_Skeletal
Skin_Sun_Exposed_Lower_leg
Skin_Not_Sun_Exposed_Suprapubic
Esophagus_Gastroesophageal_Junction
Thyroid
Esophagus_Muscularis
Heart_Left_Ventricle
Uterus
Colon_Sigmoid
Testis
Cervix_Endocervix
Cervix_Ectocervix
Liver
Vagina
Minor_Salivary_Gland
Stomach
Esophagus_Mucosa
Artery_Tibial
Whole_Blood
Cells_EBV-transformed_lymphocytes
Artery_Aorta
Lung
Colon_Transverse
Prostate
Artery_Coronary
Kidney_Cortex
Fallopian_Tube
Bladder
Small_Intestine_Terminal_Ileum
Cells_Transformed_fibroblasts
Breast_Mammary_Tissue
Adipose_Subcutaneous
Spleen
Adipose_Visceral_Omentum

Figure S1.

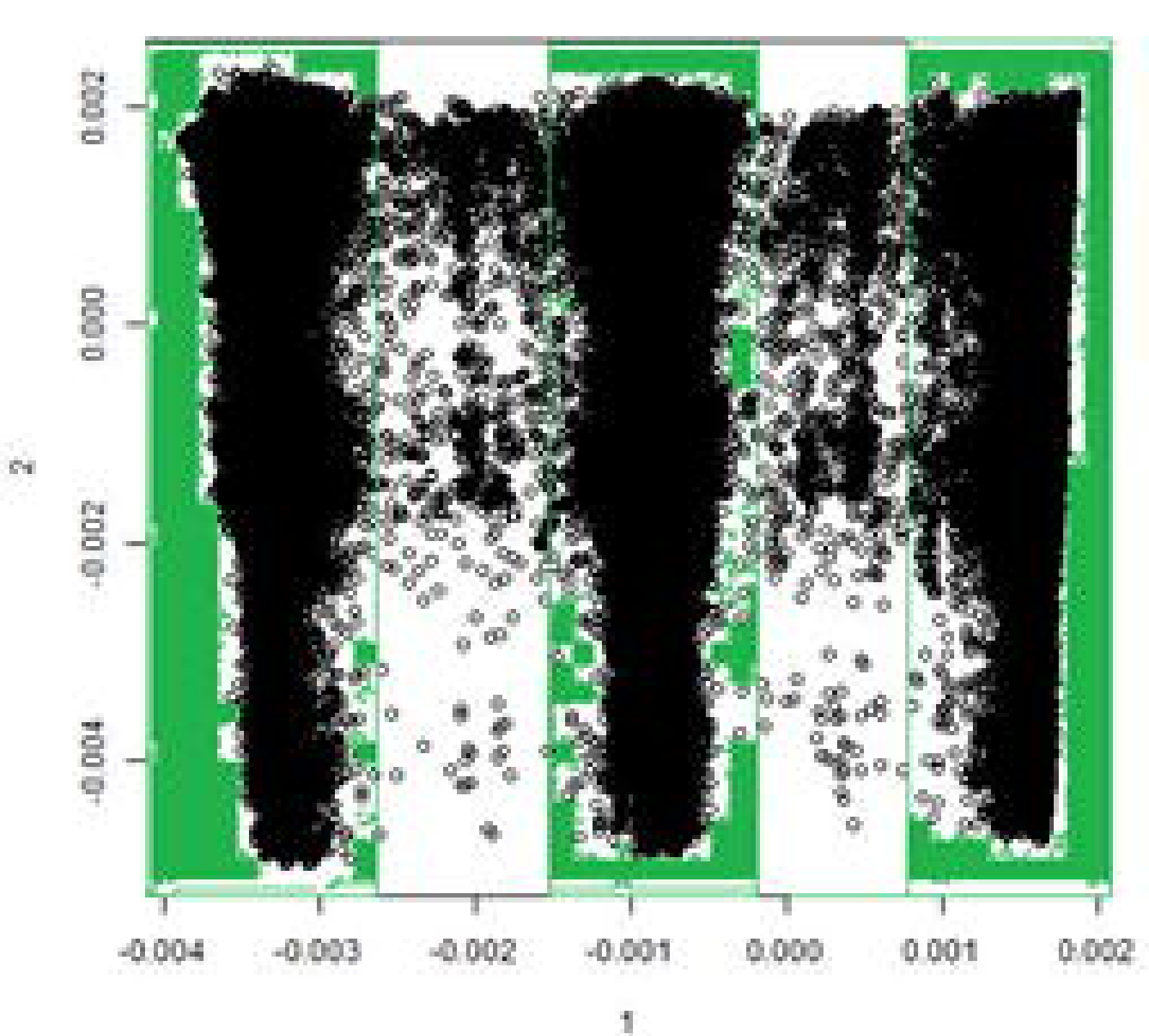

Figure S2