1   **The genomic basis of mood instability: identification of 46 loci in 363,705 UK Biobank participants,**

2   **genetic correlation with psychiatric disorders, and association with gene expression and function.**

3   Joey Ward MSc[1], Elizabeth M. Tunbridge DPhil[2, 3], Cynthia Sandor PhD DVM[4], Laura M. Lyall PhD [1],

4   Amy Ferguson BSc[1], Rona J. Strawbridge PhD[1, 5], Donald M. Lyall PhD[1], Breda Cullen PhD[1], Nicholas

5   Graham MBChB[1], Keira J.A. Johnston MSc[1], Caleb Webber PhD[4, 6], Valentina Escott-Price PhD[7],

6   Michael O'Donovan PhD [7], Jill P. Pell[1] PhD, Mark E.S. Bailey PhD [8], Paul J. Harrison[2, 3] DM (Oxon),

7   Daniel J. Smith MD[1]

8   [1]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK, [2]Department of Psychiatry,

9   University of Oxford, UK, [3]Oxford Health NHS Foundation Trust, [4]UK Dementia Research Institute,

10   Cardiff University, Cardiff, UK, [5]Department of Medicine Solna, Karolinska Institute, Stockholm,

11   Sweden, [6]Department of Physiology, Anatomy and Genetics, Oxford, UK, [7]University of Cardiff,

12   Cardiff, UK, [8]School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of

13   Glasgow, Glasgow, UK.

14

15   Corresponding Author: Professor Daniel Smith

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

18   Telephone: 0141 211 3930

19   Email: Daniel.Smith@glasgow.ac.uk

20

21

22

23

24

25

26

1   **Abstract**

2   Genome-wide association studies (GWAS) of psychiatric phenotypes have tended to focus on

3   categorical diagnoses, but to understand the biology of mental illness it may be more useful to study

4   traits which cut across traditional boundaries. Here we report the results of a GWAS of mood

5   instability as a trait in a large population cohort (UK Biobank, n = 363,705). We also assess the

6   clinical and biological relevance of the findings, including whether genetic associations show

7   enrichment for nervous system pathways. Forty six unique loci associated with mood instability were

8   identified with a SNP heritability estimate of 9%. Linkage Disequilibrium Score Regression (LDSR)

9   analyses identified genetic correlations with Major Depressive Disorder (MDD), Bipolar Disorder

10   (BD), Schizophrenia, anxiety and Post Traumatic Stress Disorder (PTSD). Gene-level and gene set

11   analyses identified 244 significant genes and 6 enriched gene sets. Tissue expression analysis of the

12   SNP-level data found enrichment in multiple brain regions, and eQTL analyses highlighted an

13   inversion on chromosome 17 plus two brain-specific eQTLs. Additionally, we used a Phenotype

14   Linkage Network (PLN) analysis and community analysis to assess for enrichment of nervous system

15   gene sets using mouse orthologue databases. The PLN analysis found enrichment in nervous system

16   PLNs for a community containing serotonin and melatonin receptors. In summary, this work has

17   identified novel loci, tissues and gene sets contributing to mood instability. These findings may be

18   relevant for the identification of novel trans-diagnostic drug targets and could help to inform future

19   precision medicine innovations in mental health.

20

1 **Introduction**

2 Mood instability is a subjective emotional state defined as rapid oscillations of intense affect, with

3 difficulty regulating these oscillations and their behavioural consequences[1]. As a

4 psychopathological phenotype, mood instability may be useful for psychiatric research within a

5 Research Domain Classification (RDoC) framework[2] because it is a symptom that occurs in several

6 psychiatric disorders, particularly major depressive disorder (MDD) and bipolar disorder (BD). It is

7 also present within general population samples, and is known to be associated with a range of

8 adverse health outcomes[3].

9 We recently identified four loci associated with mood instability within a subsample of the UK

10 Biobank cohort (n = 113,968) and found genetic correlation with both MDD and schizophrenia[4].

11 Here, we report a significantly larger genome-wide association study (GWAS) of mood instability

12 in the European ancestry subset of UK Biobank dataset (n = 363,705), using a BOLT-LMM

13 approach to maximize statistical power. We also revisit the assessment of genetic correlations

14 with psychiatric disorders, including the use of more recent GWAS outputs for MDD,

15 schizophrenia and BD. Furthermore, we contextualize our findings in terms of affected tissues,

16 eQTL analysis and Phenotype Linkage Network (PLN) analysis. PLN is a new methodology that

17 harnesses the fact that variation in many complex traits results from perturbations of multiple

18 molecular components within a smaller number of cellular pathways. These pathways can then

19 be identified using gene network approaches.

20

1   **Methods**

2   **UK Biobank sample**

3   UK Biobank is a large cohort of over 500,000 United Kingdom residents, aged between 39 and 69

4   years[5]. UK Biobank was created to study the genetic, environmental and lifestyle factors that

5   cause or prevent a range of morbidities in middle and older age. Baseline assessments occurred

6   over a 4-year period, from 2006 to 2010, across 22 UK centres. These assessments covered a

7   wide range of social, cognitive, lifestyle and physical health measures. Informed consent was

8   obtained from all participants, and this study was conducted under generic approval from the

9   NHS National Research Ethics Service (approval letter dated 13 May 2016, Ref 16/NW/0274) and

10   under UK Biobank approvals for application #6553 'Genome-wide association studies of mental

11   health' (PI Daniel Smith).

12   **Genotyping, imputation and quality control**

13   In March 2018 UK Biobank released genetic data for 487,409 individuals, genotyped using the

14   Affymetrix UK BiLEVE Axiom or the Affymetrix UK Biobank Axiom arrays (Santa Clara, CA, USA) which

15   contain over 95% common SNP content[6]. Pre-imputation quality control, imputation and post-

16   imputation cleaning were conducted centrally by UK Biobank (described in the UK Biobank release

17   documentation).

18   **Phenotyping**

19   UK Biobank participants were asked as part of their baseline assessment: "Does your mood often go

20   up and down?" Those who responded 'yes' to this question were defined as mood instability cases

21   and those who responded 'no' were defined as controls. To minimise any impact of psychiatric

22   disorders on observed genetic associations with mood instability, individuals reporting depression,

23   bipolar disorder, schizophrenia, 'nervous breakdown', self-harm or suicide attempt (all from UK

24   Biobank data field 20002), and those who reported taking psychotropic medications (data field

4

1   20003) were excluded from the analysis. Participants were also excluded if: their self-reported sex

2   did not match their genetically determined sex; UK Biobank had determined them to have sex

3   chromosome aneuploidy; they were considered by UK Biobank to be heterozygous outliers; they

4   were missing over 10% of their genetic data; or they were not in the subset classified as British

5   participants of European ancestry.

6   **Genetic association and heritability**

7   Genetic association analysis was performed using BOLT-LMM[7, 8]. This approach makes use of a

8   genetic relationship matrix (GRM) to control as robustly as possible for population structure without

9   the need to adjust the model for principal components (PCs), while maximising power by avoiding

10   the need to exclude related individuals. Additionally, BOLT-LMM builds an infinitesimal model

11   including all directly genotyped SNPs simultaneously, thereby further increasing power compared to

12   logistic regression approaches that test each SNP in turn. This 'genotyped SNPs only' model has the

13   imputed SNPs tested against it allowing for the imputation score cut-off criterion to be substantially

14   reduced and increases the number of SNPs available to test for association with the outcome. BOLT-

15   LMM treats binary variables as a linear trait but is able to handle binary outcomes well when the

16   sample size is large and when the number of cases and controls are evenly balanced, as is the case

17   here.

18   Models were adjusted for age, sex and genotyping array. SNPs were filtered to remove those with

19   MAF < 0.01, Hardy-Weinberg Equilibrium $p < 1 \times 10^{-6}$, or imputation quality score < 0.3. BOLT-LMM

20   was also used to provide a heritability estimate and λ_{GC} estimate. A secondary analysis was also

21   performed on a subsample of the cohort which excluded those used in the previous GWAS and

22   anyone who was related to another participant.

23   Regional plots were made via LocusZoom v1.4[9] as SNPs from the HRC reference panel were also

24   imputed in the UK biobank genetic data release. We defined a locus as the region of containing a

5

1   lead SNP and all other SNPs ($r^2 > 0.1$) within a 5MB radius of the lead SNP. The LD was calculated

2   using 10,000 unrelated Biobank participants who had passed the same genetic QC as those used for

3   the GWAS.

4   The summary statistics were processed by FUMA[10] (See URLs) for visualisation, MAGMA Gene

5   Analysis, Gene-set Analysis and Tissue Expression Analysis[11]. The Gene-level Analysis operates by

6   grouping p values for individual SNPs into a gene test statistic using the mean χ^2 statistic for the

7   gene whilst accounting for LD via the use of a European ancestry reference panel. The Gene-set

8   Analysis groups genes according to MsigDB v6.1[12], a collection of both curated gene sets and GO

9   terms, and tests each set in turn against all the other sets. The Tissue Expression Analysis performs a

10   one-sided test based on the correlation between tissue-specific gene expression profiles and trait-

11   gene associations.

12   **Genetic correlations**

13   Linkage Disequilibrium Score Regression (LDSR)[13] was used to calculate genetic correlations with

14   psychiatric disorders. The intercept was left unconstrained to allow for sample overlap. For the

15   MDD[14], BD[15], schizophrenia[15] and PTSD[16] phenotypes, we used the most up-to-date GWAS

16   summary statistics provided by the Psychiatric Genomics Consortium. Anxiety disorder summary

17   statistics came from the Anxiety NeuroGenetics STudy (ANGST) Consortium[17].

18   **Tissue-specific expression and eQTL analysis**

19   The lead SNP for each locus (unless otherwise noted) was assessed for *cis* effects on gene expression

20   (eQTLs) in publicly available human dorsolateral prefrontal cortex RNASeq datasets using the Lieber

21   Institute for Brain Development (LIBD) eQTL browser (See URLs). Each locus was initially examined in

22   the LIIBD BrainSeq dataset (n = 738; See URLs); SNPs showing significant eQTLs were then assessed

23   for replication in the Common Mind Consortium (CMC) dataset (n = 547; See URLs). Only eQTLs that

24   reached a False Discovery Rate (FDR) corrected threshold of q ≤ 0.05 in both the LIBD and CMC

1   datasets, and showed the same direction of effect in both, are reported. Tissue-specific expression

2   patterns were assessed for implicated genes using the GTEx portal[18]. All q values quoted in the text

3   are FDR corrected.

4   **Genetic principal component generation**

5   Genetic principal components were created using plink 2[19] using pca approx (with default settings)

6   within the region between base positions 40,850,001 and 41,850,000 on chromosome 17 for the

7   analysis of the inversion.

8   **Pathway analysis**

9   PLN analysis builds on the fact that variation in complex traits results from perturbations of multiple

10   molecular components within a smaller number of cellular pathways that can be identified using

11   gene network approaches. No single dataset or data type can provide a complete picture of the

12   functional association between genes but a recent method combines information from multiple data

13   types by weighing functional similarities between genes according to their likelihood of influencing

14   the same mammalian phenotype(s). This approach has a greater specificity and sensitivity than

15   analyses using a single data type and other comparable integrative methods[20]. The PLN approach

16   exploits phenotypic information from over seven thousand genes whose function has been

17   experimentally perturbed in the mouse and evaluates the ability of different data types such as

18   protein-protein interactions (PPI), co-expression (RNA or protein), and semantic similarity score

19   based on literature or Gene Ontology (GO) annotations or pathway annotations (KEGG), to predict

20   whether knockout of the orthologues of a given pair of human genes will yield similar phenotypes.

21   By weighting those data types accordingly, they are integrated to generate a single combined

22   measure of functional similarity between gene pairs. The resulting network of pairwise gene

23   functional similarities is termed a phenotypic-linkage network (PLN)[20].To increase the sensitivity and

24   specificity to detect functional associations relevant for a specific disease/trait, it is possible to select

1   only those mouse phenotypes that are relevant for a specific disorder in the data type weighting

2   evaluation step[21]. Following this approach, we conducted a further analysis in which we re-weighted

3   our generic PLN to be more sensitive to functional genomics data most informative to mood

4   instability by considering only phenotypes within the over-arching mouse phenotype ontology

5   (MPO) category *Nervous System* (MP:0003631). The PLN and nervous-PLN (NS-PLN) were built using

6   the same 16 functional genomics datasets described by Honti et al[20], with   64,640,972 and

7   49,656,123 weighted links respectively.

8   Following the approach described by Sandor et al., we identified 'communities' of densely

9   interconnected groups of genes (including at least 20 genes) within each PLN and tested whether

10   any communities were enriched in genes harboured by GWA/subGWA intervals. This test examines

11   how many of these intervals harboured at least one gene belonging to a given Community as

12   compared to randomly shifted intervals equal in gene number. This approach makes no prior

13   hypothesis about the number or nature of genes within each GWA interval.

14   **Definition of GWA and subGWA intervals**

15   The GWAS and subGWA intervals were defined by considering SNPs attaining an association p-value

16   of 5×10^{-8} and 1×10^{-6} identified 6375 (GWA) and 9358 (sub-GWA) SNPs. We then identified the

17   haplotypic block containing each of these of these SNPs using genotypes in the 1000 Genome

18   Project and the python pipeline developed by Brent Pedersen (See URLs). We defined GWA/sub

19   GWA intervals by identifying the most distant block on a chromosome within a region of 500Kb of

20   the lead SNP. We then added an additional 300 kb on either side of the interval to include genes that

21   may be regulated by regulatory variants with effects captured by the lead SNPs. For subGWA

22   regions, we excluded those subGWA intervals harbouring genes present in GWA intervals.

23

1   **Results**

2   **Demographics**

3   In this GWAS sample of 363,705 individuals without a history of psychiatric disorder, 43.2% reported

4   mood instability (n = 157,039) and the rest did not (n = 206,666). There was a higher proportion of

5   females amongst the mood instability cases than in controls (55.4% versus 51.2% respectively), and

6   the mean age of cases was lower than for controls (55.8 years versus 57.7 years).

7   **GWAS findings**

8   We detected 46 loci across the genome with $p < 5 \times 10^{-8}$ (Figure 1 and Table S1) and an estimated SNP

9   heritability (h^2) of 0.09 (S.E. = 0.02). The heritability estimate has increased from the previous GWAS

10   by approximately 2% (Previous h^2 = 0.07, SE = 0.007). We attribute the increase in SE to the differing

11   methodologies. The distribution of test statistics was consistent with a polygenic contribution to risk

12   (λ_{GC} = 1.21; λ_{1000} = 1.001; LDSR intercept = 1.041; SE = 0.006). In addition, to help validate the four

13   loci identified in the previous mood instability GWAS, we tested the top SNP from each locus in a

14   sub-sample that excluded individuals in the previous smaller GWAS and those individuals related to

15   another Biobank participant (n = 169,857). All four SNPs were associated with mood instability after

16   Bonferroni correction ($\alpha < 0.0125$, Table S2). We also note that the directions of effect were the

17   same as for the previous GWAS findings.

18   **Gene-level and Gene Set Analysis**

19   244 significant genes were detected by MAGMA (Supplementary Table S3) and FUMA gene analysis.

20   The Gene Set Analysis returned 6 enriched gene sets that met the threshold for significance after

21   Bonferroni correction (Supplementary Table S4). Of these, 4 sets were related to brain development

22   and differentiation of neurons, glial cells and astrocytes or neurogenesis. Other enriched sets

23   included the Nikolsky breast cancer 16q24 amplicon genes and the prepulse inhibition gene sets.

24

1   **Tissue expression analysis**

2   MAGMA tissue expression analysis identified 11 tissue categories, all of which were in the brain

3   (Figure S1). Indeed, all sampled brain areas except substantia nigra showed enrichment (i.e., frontal

4   and anterior cingulate cortex, basal ganglia, hippocampus, amygdala, hypothalamus and

5   cerebellum). **Genetic correlations**

6   Genetic correlations were calculated between mood instability and five psychiatric phenotypes of

7   interest (Table 1). All genetic correlations remained significant after FDR correction (q < 0.05). The

8   largest correlations were with MDD (r_g = 0.74, q = $8.50*10^{-157}$) and anxiety (r_g = 0.64, q = $8.08*10^{-6}$).

9   PTSD had a moderate correlation with mood instability (r_g = 0.32, q = $1.23*10^{-2}$) and both

10   schizophrenia and bipolar disorder had weak but significant correlations (schizophrenia r_g = 0.14, q =

11   $1.60*10^{-5}$, BD r_g = 0.09, q = 0.0012).

12   **eQTL analysis**

13   Nine of the GWAS loci showed significant eQTLs, many potentially with specific isoforms

14   (summarised in Table S5 and presented in full in Supplementary Table S6). The strongest evidence of

15   association with expression levels was for rs669915, an eQTL located within a region of strong

16   linkage disequilibrium (LD) in chromosome 17q21 resulting from the existence of a 900kb inversion

17   polymorphism that is common in European populations[22]. The extended region of LD across this

18   portion of the chromosome makes it challenging to identify causal SNPs or the genes they regulate.

19   The rs669915 eQTL was most strongly associated with expression of *LRRC37A4P* (LIBD dataset

20   minimum q = 1.96×10^{-99}; CMC dataset q = 3.99×10^{-65}), an expressed pseudogene, but there are

21   many alternative candidates for genes regulated by this SNP, including *MAPT* and *CRHR1*, for which

22   it was also an eQTL. (Supplementary Table S5).

23   The chromosome 17q21 inversion polymorphism has itself been reported to affect the expression of

24   genes in this region[23]. We therefore investigated whether rs669915 might 'tag' the expression

10

1 effects mediated by the inversion polymorphism in our sample. Using the method of de Jong and

2 colleagues, we constructed genetic principal components (GPCs) from SNPs within the region

3 between base positions 40,850,001 and 41,850,000 on chromosome 17. A plot of the first two GPCs

4 is shown in (Figure S2) and reveals three distinct clusters of individuals, each representing one of the

5 three inversion polymorphism genotypes, H1/H1 (right-most cluster; n = 162,113), H1/H2 (middle

6 cluster; n = 158,506) and H2/H2 (left-most cluster; n = 38,597). The H1 inversion allele had a

7 population frequency of 0.32, far higher than the frequency reported by de Jong. In linear regression

8 analyses, there was no association between mood instability phenotype and inversion genotype

9 using a model of additive allelic effects (no. of H2 alleles) and adjusting for age, sex and genotyping

10 array (p = 0.835).

11 **Nervous system PLN analyses (NS-PLN)**

12 Amongst both GWA and subGWA gene sets, we found a disproportionate aggregation of genes

13 within only one community, Community 26 within the NS-PLN (21 GWAS loci including at least one

14 gene, q = 0.011; 25 "subGWAS "loci including at least one, q = 0.018) (**Fig 2A**). Examining the entire

15 NS-PLN Community 26 gene, we found that it was significantly enriched in genes, whose unique 1:1

16 orthologues in the mouse when disrupted induce abnormities in synaptic transmission (Mouse

17 Phenotype Ontology term MP:0003635; q = $2.77e^{-118}$, 75 genes expected vs 259 gene observed).

18 However, we did not find evidence that the unique mouse orthologues of mood instability GWA and

19 subGWA genes that belonged to Community 26 were enriched for any particular mouse phenotype.

20 Nonetheless, we found that the 37 and 35 GWA and subGWA genes present in the Community 26

21 were highly functionally connected with other Community 26 genes annotated with abnormal

22 synaptic transmission phenotype term (Figure 2B).

23

11

1 **Discussion**

2 **Main findings**

3 These analyses represent the largest genetic study of mood instability to date. Forty six unique loci

4 associated with mood instability were identified, with a heritability estimate of 9%. Our findings

5 confirm the four loci identified in our initial GWAS on the UK Biobank interim data release[4] and are

6 further validated by tissue expression analyses (enrichment for 11 brain regions) and pathway

7 analyses (6 enrichment pathways, 4 of which relate to the development and differentiation of

8 neurons). The large number of individuals in this study also provided substantial power to detect

9 genetic correlations with psychiatric traits via LDSR. All five psychiatric traits assessed had a

10 significant genetic correlation with mood instability. Some of these correlations were strong

11 (particularly for MDD and anxiety) but others were weaker than expected. For example, the genetic

12 correlation between mood instability and BD was only 9%, perhaps suggesting that the mood

13 instability phenotype in this study differs from the affective instability that is a core feature of BD.

14 **Biology of mood instability**

15 Loci associated with mood instability included genes that are involved across a variety of

16 biochemical pathways, as well as brain development and function. For example, several gene

17 products localised to the synapse. *PLCL1* and *PLCL2* are involved in GABA signalling[24] and melatonin

18 signalling respectively, and *RAPSN* assists in anchoring nicotinic acetylcholine receptors at synaptic

19 sites[25]. *PLCL1* has already been identified in a GWAS of schizophrenia[26] and *PLCL2* has been shown to

20 be upregulated in bipolar disorder[27]. Additionally, we identified *CALB2* which has many biological

21 functions, including a role in modulating neuronal excitability[28]. Both *DCC* (identified in the previous

22 mood instability GWAS) and *BSN* facilitate the release of neurotransmitters within the active zone of

23 some axons[29]. *BSN* has also been shown to be associated with schizoaffective disorder via GABA

24 signalling[30]. *FARP1* promotes dendritic growth[31] and, although it has so far not been directly linked

12

1   to psychiatric disorders, it has been shown to regulate dendritic complexity[32](reduced dendritic

2   complexity is a recognised feature of schizophrenia[33]).

3   We identified several developmental genes, including *NEGR1*[34], *RARB*[35] and *EPHB1*[36], and

4   transcription factors such as *HIVEP2* (loss of function of which causes intellectual disability[37]) and

5   *TCF4* (previously associated with schizophrenia[38]). *NEGR1* was identified by 23andMe within their

6   GWAS of MDD[39] and increased levels of NEGR1 protein in spinal fluid have been identified in both

7   MDD and BD[40]. *RARB* is involved in retinoic acid synthesis pathways that have been associated

8   depressive symptoms in mice[41] and has also been found to have increased expression in patients

9   with schizophrenia[42]. The methylation state of the *EPHB1* gene has been linked to MDD[43] and SNP-

10   based analyses have identified association between *EPHB1*'s and symptoms of schizophrenia[44].

11   We also found association with several genes involved in mitochondrial energy production, such as

12   *NDUFAF3*, *NDUFS3*, *PTPMT1*, *KBTBD4* and *MTCH2*, suggesting that part of the physiology of mood

13   instability may relate to energy dysregulation.

14   In addition to protein coding genes, several loci were identified in regions containing non-coding

15   protein sequences such as *AC019330.1*, *AC133680.1*, *RP11-6N13.1* and *RP11-436d23.1*. Additionally,

16   eQTL analyses identified three more possible non-coding genes (*RP11-481A20.10*, *RP11-481A20.11*

17   and *FAM85B*) suggesting a possible RNA interference or post-transcriptional regulation basis to

18   mood instability.

19   Furthermore, the eQTL analyses highlighted the 17q21 inversion. Our principal component analysis

20   of this region did not detect a significant association leading us to conclude that it is the SNPs in the

21   region (not the inversion itself) driving the association. It is possible that lead SNPs may tag,

22   enhancer RNA or eRNA which we were unable to detect here (LIBD data was generated using poly-A

23   RNA and so targets messenger RNA). However, our findings are consistent with a recent report

13

1   implicating dopamine neuron-enriched enhancer activity in this region in several dopamine-related

2   psychiatric and neurological conditions[45].

3   Genes within regions associated with mood instability were functionally associated with synaptic

4   transmission, a key pathway for psychiatric disorders, albeit this functional association was only

5   detectable after focussing our gene network towards data types most informative for mammalian

6   nervous system phenotypes. Among the genes lying within associated loci that contribute to this

7   functional association are several interesting candidate genes. *HTR4* is a member of the family of

8   serotonin receptors and implicated in depression and its treatment[46, 47]. *MCHR1*, melanin

9   concentrating hormone receptor 1, is a G protein-coupled which binds melanin-concentrating

10   hormone. *MCHR1* can inhibit cAMP accumulation and stimulate intracellular calcium flux, and may

11   be involved in the neuronal regulation of food consumption[46] and the locus showed association with

12   schizophrenia in a Danish sample[49].

13   **Strengths and Limitations**

14   As noted above, this is the largest GWAS of a mood instability phenotype to date and has

15   successfully identified new loci, eQTLs, tissues, genetic correlations and gene network enrichments.

16   However, there are several limitations, most notably the use of a single question to define mood

17   instability, and the lack of objective verification of this phenotype. There are more detailed

18   suggested measurement scales for mood instability, such as that developed by Chaturvedi and

19   colleagues[50]. In the future, the use of these more comprehensive assessments in large samples may

20   provide some clarification of our findings, for example the lack of strong genetic correlation between

21   mood instability and BD. Nevertheless, the single question approach to mood instability has been

22   widely used, and shown to identify robust associations with a range of health outcomes and

23   disorders[1, 3]. Similarly, exclusions for psychiatric disorder were based on self-report.

1   We validated the top hits of the previous GWAS however the cohort used was not truly independent

2   (it was also part of the UK Biobank cohort). It would be of interest to replicate the 46 loci identified

3   here should sufficiently large independent replication cohorts become available in the future. As

4   well as replication of the loci, further analysis of sex specific differences would be of interest because

5   mood instability was more common in females than males. Although this difference was relatively

6   small, our reported analyses were adjusted for sex and these differences are similar to those

7   reported elsewhere[51].

8   It is also important to note that direct links between genetic risk loci and network constituents in the

9   PLN analysis will have to await the release of more completely annotated gene databases. The

10   incompleteness of phenotypic annotations is likely to explain why the genes identified in the PLN

11   analysis don't have corresponding organismal or physiological phenotypes, but the fact that there

12   were strong functional associations between the genes in the network we detected and mouse

13   orthologues that have the synaptic transmission phenotype annotation suggests that the mood

14   instability genes will also reveal this phenotype when more completely annotated databases

15   become available.

16   Finally, we note the large difference in frequencies of the inversion polymorphism on chromosome

17   17 from that reported by De Jong[23]. This difference could be due to the populations sampled to

18   estimate the frequency or over representation in those who joined UK Biobank. It is however

19   important to note that the inversion itself would be likely to contribute only a small proportion of

20   the mood instability phenotype, such that even larger sample sizes than were used here would be

21   needed to detect a correlation where one exists..

22   **Conclusion**

23   In summary, with a tripling in sample size from the previous GWAS, we identified substantially more

24   associations with mood instability in the UK Biobank cohort[4]. Future analyses of the precise roles

1   that these associations play in the clinical expression of mood instability will be relevant for the wide

2   range of psychiatric phenotypes in which mood instability occurs. We have also been able to more

3   confidently place these GWAS findings within relevant biological contexts and some of the loci and

4   pathways identified may represent candidates for future novel drug development.

5   Our findings are also of interest in the context of precision medicine innovations for mental health. It

6   is possible that polygenic risk scores derived from this work could be applied to clinical populations

7   to conduct pharmacogenomics studies and to inform patient stratification approaches. Overall, we

8   hope that our findings will stimulate further research on the biology and treatment of mood

9   instability across a range of mood and psychotic disorders.

10   **Conflict of interest**

11   The authors have no conflicts of interest to declare.

12   **URLs**

13   FUMA - http://fuma.ctglab.nl/

14   Python pipeline developed by Brent Pedersen - https://gist.github.com/brentp/5050522

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

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

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

18

Acknowledgements

17

1   Regional Development Agency. UK Biobank has also had funding from the Welsh Assembly

2   Government and the British Heart Foundation. Data collection was funded by UK Biobank.

3   The summary statistics of the GWAS are available upon request by contacting the corresponding

4   author.

1   **References**

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

4
5   2.   Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC.
6   *BMC Med* 2013; **11:** 126.

7
8   3.   Broome MR, Saunders KEA, Harrison PJ, Marwaha S. Mood instability: significance, definition
9   and measurement. *The British journal of psychiatry : the journal of mental science* 2015;
10   **207**(4)**:** 283-285.

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

16
17 5.   Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J *et al.* UK biobank: an open access
18   resource for identifying the causes of a wide range of complex diseases of middle and old
19   age. *PLoS Med* 2015; **12**(3)**:** e1001779.

20
21 6.   Biobank U. Genotyping of 500,000 UK Biobank participants. Description of sample
22   processing workflow and preparation of DNA for genotyping. 2015;11 September 2015.

23
24 7.   Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM *et al.* Efficient
25   Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*
26   2015; **47**(3)**:** 284-290.

27
28 8.   Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale
29   datasets. *Nature Genetics* 2018.

30
31 9.   Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation
32   of genetic associations with FUMA. *Nature Communications* 2017; **8**(1)**:** 1826.

33
34 10. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of
35   GWAS Data. *PLOS Computational Biology* 2015; **11**(4)**:** e1004219.

36
37 11. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al.* Gene set
38   enrichment analysis: A knowledge-based approach for interpreting genome-wide expression
39   profiles. *Proceedings of the National Academy of Sciences* 2005; **102**(43)**:** 15545.

40

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.

20

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.

1 **Figure 1 Manhattan and QQ plot of mood instability GWAS.**

2 **Table 1 Genetic correlations of mood instability with psychiatric phenotypes.** r_g = genetic

3 correlation with mood instability, S.E. = standard error of the genetic correlation, Z = the test

4 statistic, p = the p value, q the False discovery rate corrected p value. MDD = major depressive

5 disorder, PTSD = post-traumatic stress disorder.

6 **Figure 2 Different Mood associated genetic risk variants converge in a nervous specific gene**

7 **network.** (A) Enrichments of gene functional communities from a generic PLN and from a Nervous-

8 System (NS) PLN within Mood-GWA and subGWA loci (see Methods). The Community ID is given first

9 in the descriptor followed by the number of genes within that community. Only communities formed

10 from over 20 genes are shown. (B) Gene subnetwork of Community 26 from NS-PLN showing

11 functional associations between genes residing in Mood-associated GWA (red squares) and subGWA

12 (orange squares) intervals and genes whose unique mouse orthologues are annotated with

13 *abnormal synaptic transmission phenotype* (cyan squares). To increase clarity, only genes with

14 *abnormal synaptic transmission phenotype* annotation with at least three functional links to genes

15 residing in GWA and subGWA regions are shown. The colour of the link connecting two genes

16 indicates the strongest information source supporting the functional association.
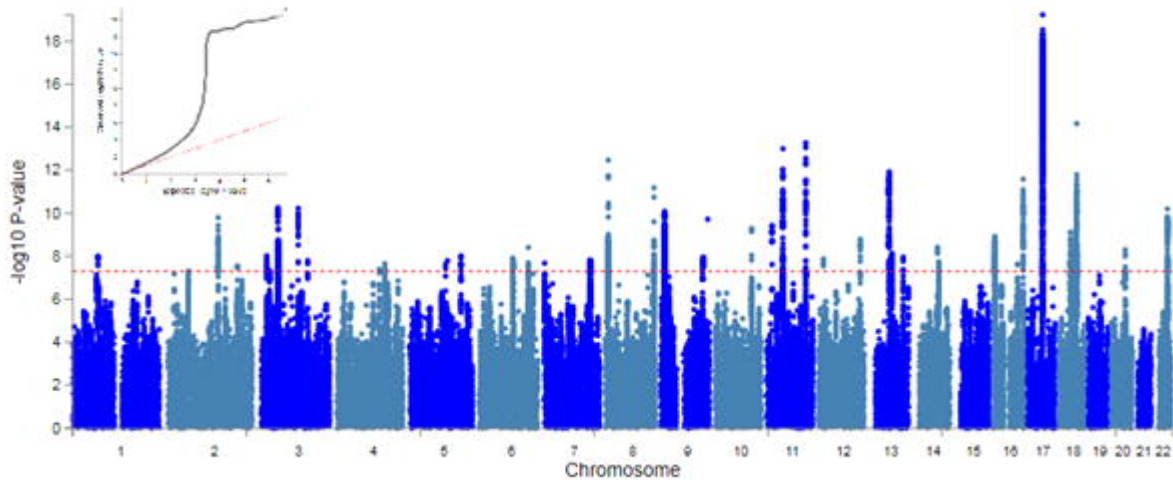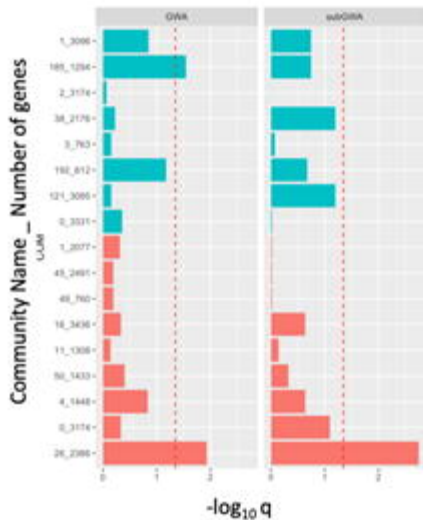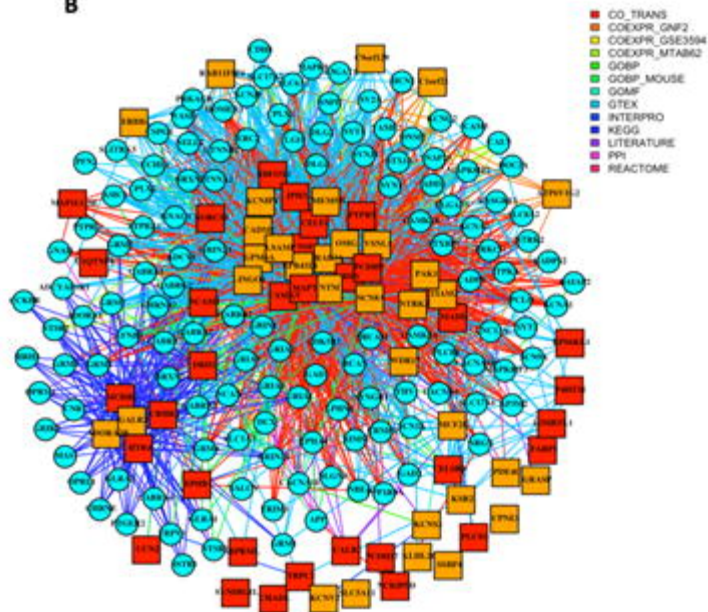
24

Figure 1.

Figure 2.

Figure S1.

Figure S2.