Ward, Joey (2020) *Genomic investigations of psychiatric conditions via research domain criteria traits.* PhD thesis.

https://theses.gla.ac.uk/83975/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

Genomic Investigations of Psychiatric Conditions via Research Domain Criteria Traits

by Joey Ward

“There is little reason to believe that these diagnostic

categories are valid.”

A comment about the DSM diagnostic categories on the first page

of a leading psychiatry textbook (Sadock, 2000).

1 Contents

2 Summary

A major challenge of psychiatry is to be able to tell who will respond best to which treatment before they start. This will save time for the patient and be beneficial to wider society. It is widely assumed that an individual's response to treatment will be due to their genetics. The most common type of genetic analysis in psychiatric genetics is that of comparing those who have a clinical diagnosis of a psychiatric morbidity to controls. The focus of this thesis, however, is to investigate the genetics of phenotypes that are features of several psychiatric diagnoses. This is known as a research domain classification (RDoC) approach. The main benefit to this approach is that is looks at traits that cut across traditional diagnostic boundaries and these traits can also apply to the general population and as such require less effort to obtain larger sample sizes with more detailed phenotyping. The thesis shows a range of differing techniques to identify genomic regions for a variety of traits as well as a range of differing downstream analyses. As the thesis progresses the techniques used become more sophisticated reflecting the progress that is being made in the field of genetic research.

The thesis begins with a meta-analysis of three treatment cohorts to determine whether genetic loading for a psychiatric morbidity and a personality trait - Major Depressive Disorder (MDD) and neuroticism, respectively - can be used to predict response to antidepressants (published in PLOS one). The paper uses a polygenic risk scoring (PRS) approach to calculate an individual's genetic loading for a trait using pruning and thresholding (P&T) methodology to see if higher genetic loading for these traits resulted in poorer outcomes for those taking a selective serotonin reuptake inhibitor (SSRI) antidepressant. The outcome measure is percentage reduction in Hamilton Depression (HAMD) score. The analysis is performed in three cohorts and the results combined using an inverse variance weighted meta-analysis. The results, although largely not statistically significant showed that greater genetic loading for both MDD and neuroticism correlated with poorer response to SSRIs.

This leads onto the first genome-wide association study (GWAS). The second and third papers investigate the same mood instability phenotype, that of a single item question on whether the

participant thought their mood often goes up and down (published in Translational Psychiatry and Molecular Psychiatry, respectively). Each paper uses a different methodology due to the techniques that were available at the time. Firstly, logistic regression and the then in a BOLT-LMM setting which allows for maximising of the sample size via use of a genetic relationship matrix. The first paper, which identifies 4 loci, uses downstream analytical techniques such as PRS analysis and linkage disequilibrium score regression (LDSR) to validate the use of a simple, easily obtainable mood instability phenotype. The second mood instability paper, due to its larger sample size and identification of forty-six genomic loci, also uses other techniques such as phenotype linkage network (PLN) analysis and expression quantitative trait loci (eQTL) analysis to further contextualise the results and identified a community of genes containing serotonin and melatonin receptors.

The fourth paper is an analysis of suicidality and was the first paper to identify areas of the genome that may drive suicidal behaviour (published in EBioMedicine). This paper uses a cumulative link function to analyse an ordered ordinal phenotype that combines self-harm and suicidal behaviours to identify 3 loci. Validation of the phenotype was performed through PRS analysis showing how those who had committed suicide had higher genetic loading for the suicidality phenotype than controls who had reported no suicidal ideation whatsoever. Then the paper explores how loading for this phenotype associates with psychiatric outcomes.

The fifth and final paper uses a measure of anhedonia for the genetic analysis and correlates risk scores of this phenotype with brain structure and function. As with the second mood instability GWAS this approach uses BOLT-LMM to maximise statistical power and sample size which led to the identification of 11 independent loci. This paper also uses a newer approach to polygenic risk scoring, that of LDpred. This newer method is superior to that of P&T as only a single risk score is generated and as it makes use of more of the available information from the GWAS summary statistics, generates a more statistically powerful score. These LDpred anhedonia scores correlated with total grey matter volume and the volume of 4 out of 15 regions of interest previously associated with anhedonia as well as two brain integrity measures in those same 15 regions.

One of the themes running through the five publications (in addition to the evolution of different methodological approaches) is the potential advantage of studying psychopathological traits rather than formal diagnostic categories. As alluded to by the opening quote in the thesis from a leading psychiatry textbook by Sadock, such an approach may be more useful than identifying genetic variants associated with Diagnostic and Statistical Manual of Mental Disorders (DSM) 5 diagnoses.

3 Introduction

3.1 Background

We have now entered an unprecedented era of discovery in the field of genetic epidemiology in mental health, driven by huge datasets and global consortia. As we reach sample sizes numbering into the hundreds of thousands, it becomes possible to begin to deconstruct highly polygenic psychiatric disorders.

However, the best way to accomplish this is currently not clear. Large consortia such as the Psychiatric Genomics Consortium (PGC) tend to compare those with categorical diagnoses of a psychiatric disorder to those without a diagnosis (Major Depressive Disorder Working Group of the Psychiatric, 2013, Schizophrenia Working Group of the Psychiatric Genomics, 2014, Duncan et al., 2018, Wray et al., 2018, Sklar, 2011). The issue here is that psychiatric disorders have always been defined by their symptomatic profile. For example, symptoms and signs are grouped together for a clinical diagnosis but it is highly likely that there are multiple genetic pathways that could lead to any given symptom or symptom cluster. As such, traditional case/control genetic analyses may not be the best way to identify the underlying biology. Additionally, within the broad clinical diagnosis category of MDD, and other psychiatric conditions, there will be a spread of risk alleles for different symptoms or clusters of symptoms and each individual will likely only have a subset of these alleles. By grouping all these symptoms together into a single phenotype many true genetic signals may be missed. This is exemplified by the fact that genome-wide association studies (GWAS) for MDD until recently failed to identify any associated region of the genome and the most recent GWAS of MDD (in over 480,000 people) identified only forty-four associated loci (Wray et al., 2018).

Additionally, if a case/control GWAS identifies a region of the genome associated with a diagnostic category such as bipolar disorder, it is not possible to say which of the associated genetic variants map to different aspects of bipolar psychopathology. For example, it is not possible to know which of the variants have more of an influence on mania and which are related to depression. One of the long-term goals of understanding the genetic aetiology of a condition is the development of new drugs by identifying new potential drug targets. This lack of genotype-phenotype specificity mitigates against translational research efforts such as drug development (for example, it is currently not possible to say which of the genetic variants associated with bipolar disorder would be the best targets for a new anti-manic drug, or indeed a mood stabiliser).

One potential application of the genetic analysis of psychiatric disorders is to stratify patients for appropriate treatment, although accomplishing this task is still a long way away. As noted above, psychiatric disorders are diagnosed by grouping together different symptoms. The causes of those symptoms may vary considerably from person to person. For example, a third of patients diagnosed with MDD and treated with SSRIs do not respond to treatment (Rush et al., 2006) and, as such, may require a different pharmacological treatment approach. The emerging field of stratified medicine aims to deliver "the right treatment to the right patient at the right time" and a major component of this approach is the use of polygenic risk scores (PRS) as potential prediction of treatment response. To date, there has been some success for example the work of Amare et al. has shown how PRS scores for Schizophrenia, HLA Antigen, MDD and depressive symptoms can be used to predict lithium response in patients with bipolar disorder (Amare et al., 2018a, Amare et al., 2018b). In both these studies greater genetic loading correlated with poorer response to lithium treatment.

Over the course of this thesis I assess the usefulness of an approach to genetic discovery that builds on the research domain classification or 'RDoC' approach (Cuthbert and Insel, 2013). These traits cut across traditional diagnoses and are also present to a greater or lesser degree within the general population. I focus on the traits of mood instability, anhedonia and suicidality. These analyses shed light on the genetic basis of the traits and help contextualising the work of others who have previously identified the same loci as associated with a range of psychiatric outcomes.

3.2 Cohorts

Two cohorts were used for the published papers. Here I give an overview of both.

3.2.1 International SSRI Pharmacogenomics Consortium

The International SSRI Pharmacogenomics Consortium (ISPC) (Biernacka et al., 2015) (paper one) is an attempt to discover genetic variants associated with response to SSRIs in MDD. The consortium was made up of just fewer than 1000 subjects for whom genetic and phenotypic data were available, collected across seven sites in Europe, North America and Asia. Only about a third of the sample were of European origin (Biernacka et al., 2015) and it is this third that were analysed in paper one as the summary statistics used to create the PRS were also derived from European ancestry individuals.

3.2.2 UK Biobank

The UK Biobank cohort (Sudlow et al., 2015) is a large general population cohort with over half a million subjects with a comprehensive range of genetic and phenotypic data collected. UK Biobank data are used in papers two to five. Composed mainly of individuals of European ancestry over the age of 40, this cohort has allowed genetic analysis on an unprecedented scale.

3.3 Mood Disorders

3.3.1 Major Depressive Disorder (MDD)

According to the DSM 5 (DSM-5, 2013), for a clinical Diagnoses of MDD an individual must present five or more of the following symptoms during the same two week period:

1. Depressed mood for most of the day, nearly every day
2. Diminished pleasure in all, or almost all, activities for most of the day, nearly every day (anhedonia)
3. Decrease in appetite and loss of weight when not on a diet or increase in weight
4. Slowing of cognitive and physical processes
5. Insomnia or hypersomnia nearly every day
6. Fatigue nearly every day

7. Excessive and/or inappropriate feelings of guilt nearly every day and/or feelings of worthlessness nearly every day
8. Diminished concentration or indecisiveness nearly every day
9. Thoughts of death and/or suicidal ideation

It is important to note that either points 1 or 2 (low mood and anhedonia), as core symptoms, must be present and the symptoms cannot be attributed to substance abuse or other medical condition such as dementia.

3.3.2 Bipolar Disorder

Bipolar Disorder is a mood disorder where individuals experience recurrent depressive phases and manic phases. For clinical diagnosis (DSM-5, 2013) manic phases must last at least 7 days and comprise:

1. Feelings of euphoria
2. Reduced need for sleep
3. Increased sexual desire
4. Hallucinations and/or delusions
5. Marked increase in energy
6. Engage in risky or reckless behaviour

Several subtypes of bipolar disorder have been defined in DSM 5 that distinguish between severity and frequency of the manic and depressive phases (DSM-5, 2013). Bipolar I is defined as mania plus depression and bipolar II is hypomania plus depression.

3.4 Phenotypes

The main phenotypes investigated in each of the papers and the rationale for studying them, are described below.

3.4.1 Hamilton Depression Ratings Scale (HAMD)

The Hamilton depression rating scale (Hamilton, 1960) (HAMD) assesses depressive symptom severity. It was developed in the 1950s as a tool for measuring change in depressive symptoms within pharmacological clinical trials. Paper one investigates how change in HAMD scores during treatment with an SSRI might correlate with genetic loading for both MDD and the depression-related trait of neuroticism. This paper is relevant in today's emerging field of stratified medicine as it is widely assumed that failure to respond to pharmaceutical treatment is due to individual genetic variability.

3.4.2 Mood Instability

A core feature of many psychiatric disorders is "mood instability", which can be defined as a subjective perceived inability to adequately regulate internal mood states. Both paper two and paper three make use of a single item self-report measure in the UK Biobank cohort to define mood instability based on the baseline response to the question "*Does your mood often go up and down?*" (Biobank data variable 1920) as the outcome variable of a genome wide association study. Determining genetic variants that affect mood regulation could be useful in identifying new drug targets that could be used to treat multiple psychiatric conditions, most notably bipolar disorder.

3.4.3 Suicidality

Identifying genes linked to suicide in large genetic studies has always been difficult because suicide is a relatively rare event and the process of obtaining consent from people at risk of suicide is challenging. It is however possible is to create a scale of increasingly severe suicidal thoughts and behaviours, such as self-harm and attempted suicide. Paper four makes use of such a scale constructed from the UK Biobank dataset. As suicidal ideation is not present in all psychiatric disorders it would be of great benefit to patients and clinicians to be able to determine which

patients may be at greater risk of self-harm or suicide. Paper four lays the foundation for the development of such patient stratification by establishing a genetic basis for suicidality.

3.4.4 Anhedonia

Anhedonia is defined as an inability to feel pleasure in normally pleasurable activities and is a core symptom of MDD and other conditions, such as Parkinson's disease. Paper four uses a novel GWAS approach where anhedonia is defined on an ordinal scale. Participants were asked at baseline "*Over the last two weeks, how often have you been bothered by little interest or pleasure in doing things?*" (Biobank data variable 2060) with responses measured in the approximate number of days. One potential advantage of this is that any genomic regions associated with anhedonia should be relevant to fundamental pleasure and reward pathways in the brain, rather than heterogeneous diagnostic categories such as MDD.

3.5 Techniques

3.5.1 Primary techniques

This section describes the techniques and methodologies that I developed and delivered during my research fellowship, including the designing of pipelines, running analyses and analysing and interpreting outputs.

3.5.1.1 Polygenic Risk Scoring

Polygenic risk scoring (PRS) is a method for estimating an individual's genetic loading for a trait using GWAS summary statistics (Dudbridge, 2013). PRS are used extensively throughout the five publications. The first four papers use the pruning & thresholding (P&T) method and are created by firstly determining the linkage disequilibrium (LD) structure of the genome by grouping single nucleotide polymorphisms (SNPs) into clumps, or blocks, using certain LD parameters, then selecting the SNP with the lowest p value from each clump that is below the threshold of the score. For all the SNPs used in the score, the number of risk alleles an individual has at that SNP is multiplied by the effect size for that SNP from a given GWAS, summed across all SNPs used in the score, then averaged. This process makes use of the fact that not all SNPs that affect any given

phenotype may reach genome wide significance in a GWAS of that trait. As such, several scores are created at ever less stringent p value thresholds, allowing for more SNPs to be incorporated into the score.

The second method, LDpred (Vilhjálmsson et al., 2015) used in paper five, works on a similar principal but instead of using the SNP with the lowest p value an expectation function is used to calculate the expected beta coefficient of the clump given the beta coefficients that have been measured. Using their infinitesimal model creates a single score per trait. As more information is used in generating this score than in the P&T methodology the LDpred scores are more powerful than even low threshold P&T risk scores.

These risk scores can then be used as predictors either for the same trait in a different cohort as a form of validation of the GWAS, or as predictors of a different phenotype/outcome. Significant correlations of cross-trait PRS analyses suggest that the underlying genetic architecture of the two traits is to some extent shared and may have multiple genes and/or pathways in common.

3.5.1.2 Random Effects Meta-analysis

When a similar experiment has been performed in several different cohorts it is possible to combine the results into a single result using only the results of the separate experiments. The most common method used is known as the "inverse variance-weighted random effects model", as implemented in paper one. Here variance is defined as simply 1 divided by the standard error of the beta coefficient, not the more commonly used definition of the average of the squared differences from the mean. This has the effect of giving components of the meta-analysis with smaller standard error a larger weighting in the model than those with larger standard error. This is because analyses with smaller standard errors are closer to the true value than those with large standard errors (implying low accuracy). A random effects model is more often chosen than a fixed effects model because it assumes the outcome is arrived at by different means (fixed effects models assume that the outcome in different analysis is achieved via the same route). In the context used in this thesis (HAMD scores) we know that the same score may be arrived at differently in two individuals.

Additionally, different groups of depressed patients may have different genetic factors causing depression. As such, a random effects model is more appropriate than a fixed effects model.

3.5.1.3 Genome Wide Association Study

GWAS identifies regions of the genome that associate with phenotypes. This methodology tests each SNP in turn by looking at how the phenotype changes in individuals with differing copies of the minor allele (MA). To use an extreme example, imagine a binary phenotype and one bi-allelic SNP. First, we look at those who have no copies of the MA and none of them have the phenotype. In those with one copy of the MA, half of them have the phenotype and in those with two copies of the MA all have the phenotype. We would then conclude that this SNP does associate with the phenotype. In the real world, however, the changes in phenotype are a lot more subtle.

As the number of SNPs tested can be in the millions, a more stringent p value threshold is used and conventionally significance is achieved when the p value of a SNP drops below $5*10^{-8}$. This threshold is derived not simply as a Bonferroni correction (i.e. 0.05 divided by the number of tests performed) but by the number of separate LD blocks in the human genome of those with European ancestry, currently estimated at around 1 million blocks. As we shall see, a GWAS can be performed on a variety of traits from case/control analyses, within a logistic setting, and using linear and ordinal phenotypes within a BOLT-LMM setting (Loh et al., 2015b).

The use of BOLT-LMM, which was designed for use in datasets the size of UK Biobank (Loh et al., 2018), is a far more powerful tool for genetic analysis than logistic regression for two main reasons. Firstly, it uses a genetic relationship matrix (GRM) (Loh et al., 2015b) to account for both relatedness between individuals as well as population stratification of the sample and secondly, it creates a global model of genotyped SNPs which more closely resembles the genetic architecture of a traits than testing SNPs one at a time. It has been shown to work well investigating the genetic architecture of psychiatric pathologies such as schizophrenia (Loh et al., 2015a).

Conventionally, BOLT-LMM only applies to linear phenotypes but in paper three I use it to analyse mood instability as a binary phenotype. This is acceptable because BOLT-LMM treats these

phenotypes as a linear phenotype (where the only points on the scale are 0 and 1). Similarly, with the anhedonia GWAS in paper five BOLT-LMM assesses an ordered ordinal phenotype as a linear phenotype. Paper four also uses an ordered ordinal phenotype (suicidality) but it was not possible to treat this as a linear phenotype because the distance between points was not measurable, i.e. there was no way of quantifying the distances between the phenotypes of "*thought life not worth living*" and "*actual self-harm*" and "*attempted suicide*". As such, a cumulative link function was employed for the main analysis in paper four. The link function works as an extension of a binary logistic regression computing the odds of being 0 to being 1 then from 1 to 2 and so on up the scale. This eliminates the need for knowing the distance between points on the suicidality scale.

3.5.1.4 Linkage Disequilibrium Score Regression

Linkage disequilibrium score regression (LDSR) is used in papers two to five and serves several purposes. Firstly, using only GWAS summary statistics it can partition the separate contributions of polygenic effects and other confounding factors. This allows for SNP heritability estimates for the trait to be made and inflation of the test statistics to be calculated. It does this by regression of a SNPs LD score against its test statistic from a given GWAS. Additionally, it can be used to estimate genetic correlations between traits. This is accomplished by calculating the deviation between χ^2 statistics of the two different phenotypes to that which would be expected under the null hypothesis.

3.5.2 Secondary techniques

There are some techniques used in the published papers that were performed by someone else. Here I provide a brief overview of these techniques and how they added value to the primary focus of the GWAS papers.

3.5.2.1 Expression Qualitative Trait Loci analysis

GWAS are good for identifying regions of the genome that associate with phenotypes and for calculating heritability estimates but tell us nothing about function: further downstream analyses are required. One of these is expression qualitative trait loci (eQTL) analyses. These work by identifying SNPs in associated loci that are known to impact on the expression of genes, either by

being part of a genotype-specific gene expression pattern or in very high LD with one, using databases such as GTEx (Consortium, 2013). The GTEx database uses global RNA expression from individual tissues and variations in the genome to identify genetic variants that correlate with the amount of each of the different RNAs found within each tissue.

EQTL analysis is assumed that most of the variations of a phenotype are caused by changes in expression of genes rather than actual alterations to protein coding sequences due to the high frequency of intronic and intergenic mutations compared to exonic mutations. By determining which genes are impacted by eQTLs in the region this approach can aid in identifying the genes that are affecting the phenotype and, by extension, the biological mechanism which would give rise to the phenotype.

3.5.2.2 Phenotype linkage network analysis

Another useful type of downstream analysis of GWAS summary statistics is that of phenotype linkage networks (Honti et al., 2014) used in paper three. These networks are created by combining data from a variety of sources from –omics data sources such as protein-protein interactions, GO terms, the KEGG pathway database, and mouse orthologue databases. Biases can arise from such methods, such as gene coding sequence length but these issues can be overcome and allows for determining if the summary statistics from a given GWAS show enrichment in any given biological pathway or network.

3.5.2.3 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a technique used to analyse biological structures, and in the context of paper five the brain. In paper five we used two common structural MRI techniques. Diffusion tensor magnetic resonance imaging provides measures of white matter tractography by determining the extent and direction of water diffusion along axons in white matter tracts, as water molecules will travel along an axon more freely than across an axon. This allows for the determination of the primary direction of diffusion along a tract axon. From this, two variables can be derived called fractional anisotropy (Chenevert et al., 1990) and the other is mean diffusivity (Le Bihan et al., 1986). Fractional anisotropy is the proportion of water molecules that are traveling in

the expected direction (i.e. along an axon as opposed to across an axon), where higher values are considered a marker of 'better' white matter tract integrity. Mean diffusivity is a measure of the average rate of diffusion of water in the brain in all three axes, where higher values are considered 'worse'.

Volumetric image analysis is used to calculate the volume of brain tissues (i.e., grey matter, white matter, cerebrospinal fluid), and of individual regions of interest. There are multiple methods for calculating these volumes from manually specifying points on an image to more autonomous methods. In UK Biobank, an automated method (FMRIB's Automated Segmentation Tool (Woolrich et al., 2009)) was used to derive total white matter, grey matter and cerebrospinal fluid volumes, and volumes of 139 regions of interest from the Harvard-Oxford cortical/subcortical atlas. Paper five uses a PRS approach to see whether the volumes of specific brain regions as well as their fractional anisotropy and mean diffusivity correlate with a person's genetic loading for anhedonia.

3.6 Relevant concepts

3.6.1 Principal Genetic Components

Principal genetic components (PGCs), used extensively throughout the thesis, are covariates that allow for high dimensional data, in this context SNP data, to be compressed into lower orders of data without losing the information. Adjusting models for principal components in genetic analysis accounts for cryptic relatedness, or population stratification, between individuals used in the analyses (Price et al., 2006). The use of PGCs as covariates in the models accounts for the natural variation in allele frequencies in different locations reducing the chance of a type 1 error.

It is possible to generate as many principal components as there are SNPs but most of the variation will always be in principal components one and two. In the papers I adjust for the first eight principal components as a conservative measure.

3.6.2 Relatedness

Another important issue that requires consideration in genetic analyses is that of relatedness. It is important to make sure that no two individuals in a sample are too closely related as it leads to bias

in the results. If there are closely related individuals in the sample, appropriate consideration of this is necessary. The two different methodologies used are that of filtering by relatedness using kinship coefficients and using a genetic relationship matrix (GRM).

A kinship coefficient is a number that states how closely two individuals are related. The closer related two individuals are, the closer their kinship coefficient is to 1 although it is important to note that a kinship coefficient of greater than 0.5 implies inbreeding. More reasonable examples of kinship coefficients are that of sibling-sibling and parent-sibling, both of which are 0.25, and monozygotic twins who have a kinship coefficient of 0.5. Kinship coefficients are used in papers two and four for excluding one individual at random from each pair of related individuals with valid phenotypes. The cut-off for deciding if two individuals are related was 0.042, which is the equivalent of second cousins.

An alternative approach used in papers two and five is that of a GRM used by BOLT-LMM. Here the model accounts for relatedness using a matrix of SNPs that are not in LD to establish the degree of relatedness. The benefit of this method is that it allows for the sample size to be maximised because no one is filtered out for being too closely related to someone else in the sample.

3.6.3 Genotyping chip

The UK Biobank cohort was genotyped on two different genotyping chips (Biobank, 2018). Approximately 50,000 people were genotyped on the UK BiLEVE Axiom array chip whilst the remainder were genotyped on the UK Biobank Axiom Array chip. Although they had over 98% of SNPs in common these slight differences lead to a difference in the results of the final imputed dataset. To control for these differences, and to stop variation between the chips being falsely attributed to the phenotype, all of the GWAS in this thesis were adjusted for genotyping chip.

4 Papers

4.1 Polygenic risk scores for major depressive disorder and neuroticism as predictors of antidepressant response: Meta-analysis of three treatment cohorts

PLOS ONE

RESEARCH ARTICLE

Polygenic risk scores for major depressive disorder and neuroticism as predictors of antidepressant response: Meta-analysis of three treatment cohorts

Joey Ward[1]*, Nicholas Graham[1], Rona J. Strawbridge[1,2], Amy Ferguson[1], Gregory Jenkins[3], Wenan Chen[4], Karen Hodgson[5], Mark Frye[3], Richard Weinshilboum[3], Rudolf Uher[6], Cathryn M. Lewis[5], Joanna Biernacka[3], Daniel J. Smith[1]

1 Institute of Health And Wellbeing, University of Glasgow, Glasgow, Scotland, **2** Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden, **3** Mayo Clinic, Rochester, MN, United States of America, **4** St. Jude Children's Research Hospital, Memphis, TN, United States of America, **5** King's College London, London, England, **6** Dalhousie University, Halifax, Canada

* joey.ward@glasgow.ac.uk

Check for updates

OPEN ACCESS

Citation: Ward J, Graham N, Strawbridge RJ, Ferguson A, Jenkins G, Chen W, et al. (2018) Polygenic risk scores for major depressive disorder and neuroticism as predictors of antidepressant response: Meta-analysis of three treatment cohorts. PLoS ONE 13(9): e0203896. https://doi.org/10.1371/journal.pone.0203896

Editor: Toni-Kim Clarke, University of Edinburgh, UNITED KINGDOM

Received: June 19, 2018

Accepted: August 29, 2018

Published: September 21, 2018

Copyright: © 2018 Ward et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The GENDEP genetic and phenotype data is available from GENDEP - http://gendep.iop.kcl.ac.uk/. The ISPC genetic and phenotype data is available from the ISPC - https://www.pharmgkb.org/page/ispc.

Funding: This work received support from Royal College of Physicians of Edinburgh JMAS Sims Fellowship, http://www.rcpe.ac.uk/college/jmas-sim-fellowship, UKRI Innovation- HDR-UK Fellowship (Grant MR/S003061/1 to Dr Rona J

Abstract

There are currently no reliable approaches for correctly identifying which patients with major depressive disorder (MDD) will respond well to antidepressant therapy. However, recent genetic advances suggest that Polygenic Risk Scores (PRS) could allow MDD patients to be stratified for antidepressant response. We used PRS for MDD and PRS for neuroticism as putative predictors of antidepressant response within three treatment cohorts: The Genome-based Therapeutic Drugs for Depression (GENDEP) cohort, and 2 sub-cohorts from the Pharmacogenomics Research Network Antidepressant Medication Pharmacogenomics Study PRGN-AMPS (total patient number = 760). Results across cohorts were combined via meta-analysis within a random effects model. Overall, PRS for MDD and neuroticism did not significantly predict antidepressant response but there was a consistent direction of effect, whereby greater genetic loading for both MDD (best MDD result, p < 5*10–5 MDD-PRS at 4 weeks, β = -0.019, S.E = 0.008, p = 0.01) and neuroticism (best neuroticism result, p < 0.1 neuroticism-PRS at 8 weeks, β = -0.017, S.E = 0.008, p = 0.03) were associated with less favourable response. We conclude that the PRS approach may offer some promise for treatment stratification in MDD and should now be assessed within larger clinical cohorts.

Introduction

Major Depressive disorder (MDD) is a leading cause of disability worldwide [1]. Antidepressants such as Selective Serotonin Reuptake Inhibitors (SSRIs) are first line treatments for MDD but up to one third of patients do not respond satisfactorily [2, 3]. There are currently no robust methods for predicting whether an individual patient will respond well to SSRIs and

Strawbridge), MRC Doctoral Training Programme (Grant MR/K501335/1 to Ms Amy Ferguson), National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London to Prof Cathryn Lewis.The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

there is often a lag period of several weeks before clinical response, making decisions on switching to a different class of antidepressant difficult. Individual genetic variation may dictate likelihood of response to SSRIs [4] and, as such, stratifying patients into sub-groups based on genetic profiles may allow for more efficient targeting of treatment.

Polygenic risk scoring (PRS) [5] is a method which allows an individual's genetic loading for a trait to be calculated using genome-wide single nucleotide polymorphism (SNP) data and the output of genome-wide association study (GWAS) summary statistics from another study of the same or related phenotype. As current GWAS results do not capture the full extent of genetic effects on any given trait, typically a series of scores are created at different association p-value cut offs, allowing for the capture of more variance than that explained by only genome-wide significant loci. Additionally, as the underlying genetic architecture of the trait is unknown creating a range of scores can allow for the optimum p value threshold to be determined, should one detect a significant correlation.

It has been shown that a PRS can be of clinical use in predicting traits in independent samples. For example, for coronary heart disease, PRS improved the 10 year risk prediction in those over age 60 [6]. PRS approaches can also predict response to treatment, as demonstrated recently with an association between PRS for schizophrenia and less favourable response to lithium in bipolar disorder [7]. Here we test the hypothesis that PRS for MDD and PRS for neuroticism are associated with less favourable response to SSRIs, specifically citalopram and its active S-enantiomer escitalopram, in patients with MDD. Neuroticism is of particular interest in this regard because it has a known association with both serotonergic neurotransmission [8] and response to antidepressants [9, 10], and those with higher phenotypic neuroticism are less likely to respond as well to antidepressant therapy [11].

The analysis investigated three cohorts, GENDEP, AMPS-1 and AMPS-2 separately and then combine the results via meta-analysis.

Methods

Cohort descriptions, genotyping and imputation

The Pharmacogenomics Research Network Antidepressant Medication Pharmacogenomics Study (PGRN-AMPS) is a study of citalopram/escitalopram for treatment of MDD performed at the Mayo Clinic. An initial batch of 530 subjects (N = 499 subjects of European ancestry that passed quality control) was genotyped for a pharmacogenomics GWAS of SSRIs [12]. An additional 229 patients recruited in the PGRN-AMPS were subsequently genotyped for the International SSRI Pharmacogenomics Consortium (ISPC) GWAS [13]. Depressive symptoms were assessed on the Hamilton Depression Rating Scale (HAMD) with a maximum score of 51, a scale developed to rate both the psychiatric as well as the psychomotor and somatic symptoms of the condition[14]. Full genotyping and imputation of these cohorts (here referred to as AMPS-1 and AMPS-2) have been described previously [12, 13].

Genome Based Therapeutic Drugs for Depression (GENDEP) is a cohort of 868 individuals, recruited from across Europe, treated with two classes of antidepressants: escitalopram (an SSRI) and nortriptyline (a tricyclic antidepressant). For the purposes of this study, only those patients in GENDEP treated with an SSRI were assessed (n = 267). Depressive symptoms were assessed on the 10-item Montgomery-Asberg Depression Rating Scale (MADRS) with a maximum score of 60, with measurements taken weekly for 12 weeks from baseline. MADRS differs from HAMD in that it focuses exclusively on the psychiatric symptoms only and not the accompanying psychomotor and somatic symptoms of MDD [14]. Full genotyping and imputation methodology in GENDEP is described in previous reports [15].

Principal component generation and PRS construction

Principal genetic components were derived using PLINK. For all models the top 4 principal components were used as covariates in the model to account for hidden population structure. To ensure that an ethnically homogeneous sample was used in the AMPS-1 and AMPS-2 cohorts those whose Principal genetic components 1 to 4 were outside two standard deviations from the mean were excluded as outliers.

PRS were constructed via PLINK [16] with SNP weights based on outputs from the Smith et al. (2016) neuroticism GWAS [17] and the "probable MDD" phenotype of Howard et al (2018) MDD GWAS from UK Biobank[18]. SNPs were filtered by MAF < 0.01, HWE $p<1^{*}10^{-6}$ and imputation score < 0.8 before Linkage Disequilibrium (LD) clumping. SNPs were clumped using LD parameters of $r^2 > 0.05$ in a 500kb window. Selection of SNPs for each clump was based on which SNP had the lowest p value. If 2 SNPs in a clump had the same P value the SNP with the largest beta coefficient was selected. The scores generated were average scores with no-mean-imputation flag. Six profile scores were created for each trait using p value cut offs of $p < 5^{*}10^{-8}$, $p < 5^{*}10^{-5}$, $p < 0.01$, $p < 0.05$, $p < 0.1$ and $p < 0.5$. Risk scores were then standardised to mean = 0, SD = 1[19].

Due to low numbers and therefore the potential for noise within outcome data, instead of assessing change in outcomes across the full range of polygenic scores we chose to investigate only the difference between the extreme ends of the PRS scale. To do this, we split the standardised scores into quintiles and looked at the difference between the top and bottom quintile of each PRS p-value cut off within each cohort. For the GENDEP cohort the top and bottom quintile from each centre was selected to account for variation between recruitment centres. It is also important to note that an individual may be in the top quintile for one PRS P-value cut-off but not in another. As such, the two fifths of individuals used in each regression will change depending on the PRS p value cut off used.

Phenotype definition

For all three cohorts the primary outcome of interest was percentage change in depression score from baseline at four weeks. This was calculated by subtracting the score at four weeks from baseline, and dividing this difference by the score at baseline. A secondary outcome at eight weeks was also assessed, calculated using the same method. To be included in the analysis, an individual had to have a score recorded at baseline, four weeks and eight weeks.

Statistical modelling

Modelling was performed in R using the lm function. All models were adjusted for age, sex and the first 4 principal components. The GENDEP models were additionally adjusted for recruitment centre which was treated as a factor variable. The R^2 for the PRS term of the model was derived using the methodology described in Selzam et al[20]. Due to the results being largely null we did not perform any correction for multiple testing.

Meta-analysis

A random effects Meta-analysis was performed using the rma.uni function of the metaphor package with method set to "REML"[21].

Results

Demographic and clinical characteristics of the three cohorts (GENDEP, AMPS-1, and AMPS-2) are presented in Table 1. The percentage female and age range of the three cohorts were broadly

Table 1. Demographic and clinical characteristics.

Cohort	Total N	N used per regression	N female of total N (%)	Age of total N, mean (SD)	Baseline score*, mean (SD)	4 week score*, mean (SD)	8 week score*, mean (SD)	% drop in mean score at 4 weeks from mean score at baseline	% drop in mean score at 8 weeks from mean score at baseline
AMPS-1	357	142	229 (64.1)	40.9 (13.5)	22 (4.88)	11.9 (6.7)	8.83 (5.92)	46	60
AMPS-2	138	55	85 (61.6)	40.1 (13.6)	21.2 (5.14)	12 (5.84)	9.14 (6.41)	43	57
GENDEP	265	106	170 (64.2)	42.3 (11.8)	28.3 (6.16)	18.7 (8.2)	14.2 (8.89)	34	50

*score rating is HAMD for AMPS-1 and AMPS-2 and MADRS for GENDEP.

https://doi.org/10.1371/journal.pone.0203896.t001

similar. The scores at baseline, 4 week and 8 week time points in AMPS-1 and AMPS-2 show a similar trend with a similar percentage drop at 4 and 8 week time points. The baseline scores of the GENDEP cohort are higher than in AMPS-1 and AMPS-2 due to the cohort being scored using MADRS and not HAMD as is the case with AMP-1 and AMPS-2. At Both the 4 week and 8 week time point the GENDEP cohort showed a smaller percentage reduction than in the AMPS-1 and AMPS-2. This difference may be explained by the differing depression measures picking up on differing aspects of MDD, differing healthcare settings and levels of severity at baseline. The within cohort drop from baseline at both 4 and 8 weeks was statistically significant for all three cohorts.

For neuroticism PRS in GENDEP, AMPS-1 and AMPS-2 the number of SNPs in each risk score were similar between cohorts across all p-value cut-offs (S1 Table). For the MDD risk scores the number of SNPs were similar between cohorts in the lower p value thresholds but diverged at the higher p value cut-offs. These differences arise mainly due to the differences in imputation coverage and the differing ethnicities and their impact on LD block estimation.

Individual study analyses

The results of all the individual study analyses can be found in S2–S4 Tables. Two of the models returned nominally significant results, both of which were in the AMPS-2 cohort (Table 2). They were neuroticism $p < 0.5$ PRS at four weeks ($\beta = -0.04$, $p = 0.02$) and neuroticism $p < 0.5$ at eight weeks ($\beta = -0.039$, $p = 0.03$). Of particular note is the R^2 of the PRS term of the significant models which accounts for approximately 10% of the variance. Note, however, that these results would not pass correction for multiple testing.

Although we were unable to reject the null hypothesis in the rest of the models, a clear majority (56 of 72 models) identified beta coefficients in the same direction of effect (greater loading for MDD or neuroticism associated with a smaller percentage drop in depression score). Of the 16 positive beta coefficient models, ten were from GENDEP MDD PRS models, three were from GENDEP neuroticism PRS model, two were from AMPS-1 neuroticism PRS models and one was an AMPS-2 MDD PRS models (S2–S4 Tables).

Meta-analysis

Two of the 24 meta-analyses were nominally significant: MDD $p < 5^{*}10^{-5}$ PRS at four weeks ($\beta = -0.02$, $p = 0.009$, $I^2 = 0$); and neuroticism $p<0.1$ PRS at eight weeks ($\beta = -0.017$, $p = 0.03$,

Table 2. Nominally significant individual PRS models (AMSP-2 cohort).

predictor	Time point (weeks)	p	Beta	SE	T Test stat	R^2
Neuroticism p<0.5	4	0.019	-0.044	0.018	-2.42	0.1
Neuroticism p<0.5	8	0.029	-0.039	0.017	-2.26	0.08

https://doi.org/10.1371/journal.pone.0203896.t002

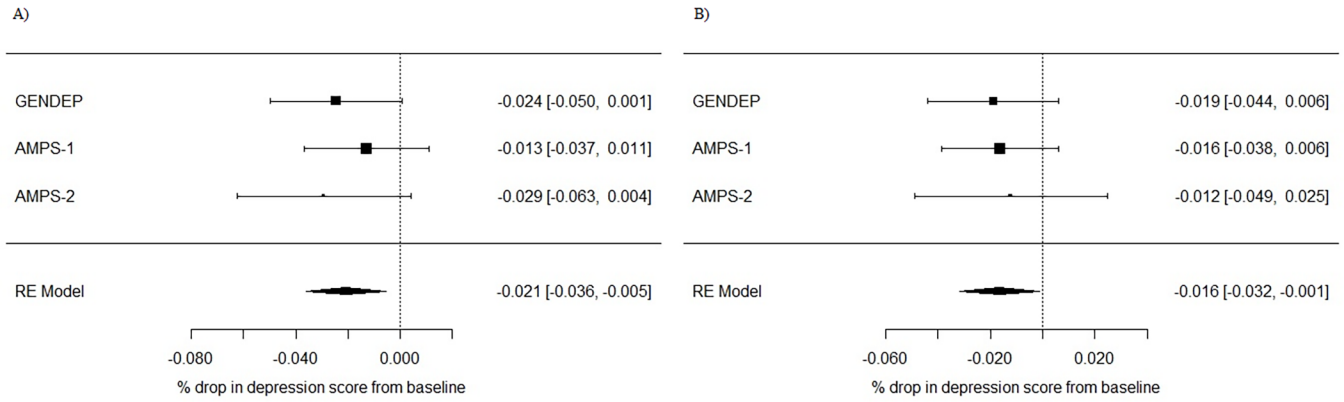


Fig 1. Forest plot of nominally significant meta-analyses. A) $p < 5^{*}10^{-5}$ MDD-PRS at 4 weeks, B) p<0.1 Neuroticism-PRS at 8 weeks.

https://doi.org/10.1371/journal.pone.0203896.g001

$I^2 = 0$) (Fig 1). Neither of these results would survive correction for multiple testing. The direction of effect in all of the meta-analyses was negative (greater genetic loading for MDD and neuroticism associated with a smaller percentage drop in depression score at both four and eight weeks; S5 Table. The forest plots of all other meta-analyses are provided as supplementary material (S1–S4 Figs).

Discussion

Our goal was to assess the extent to which PRS for MDD and PRS for neuroticism were associated with response to SSRIs in patients with MDD. Although most of the findings were null, there was a direction of effect where higher PRS for MDD and higher PRS for neuroticism were associated with less favourable response to SSRIs. It is likely that our analyses were under-powered–replication in larger datasets will therefore be of interest. We estimate that a training sample of approximately 10,000 and a target sample of 5,000 individuals would give 60% power in a PRS of 100,000 SNPs that explain 10% of the variance in the training sample [22]. For the two AMPS-2 nominally significant results the R^2 values of approximately 10%, suggesting that these PRSs could potentially be useful clinically.

This work diverges from previous analyses in these cohorts which have focused on GWAS and candidate gene analyses to identify genetic loci that associate with antidepressant response with the exception of Garcia-Gonzalez et al[23]. However, the outcome is markedly different to the outcome used here. It is possible that the use of PRS is advantageous for clinical use over these methods as it allows for a whole-genome approach instead of focusing on specific SNPs, genes or regions. An individual's response to antidepressants is likely to be influenced by many genetic factors and, as such, candidate gene methodologies will fail to capture polygenic influences. An additional strength of this work is that all three cohorts systematically assessed treatment response at comparable time-points and in the context of the use of the same class of antidepressants, namely SSRIs.

Limitations

Apart from the issue of low power, our methodology was one in which only the extreme ends of genetic loadings were considered. This makes it difficult to translate the findings into a general population setting and routine clinical practice. Further work is needed to assess genetic loadings for MDD and neuroticism within the general population and how these relate to the clinical cohorts described here. The use of different depression rating scales between GENDEP

and the AMPS-1/AMPS-2 may have had some impact on the results as they may have captured different aspects of the depressive phenotype and symptom changes induced by antidepressants. However, I^2 was low in the meta-analyses that achieved nominal significance. Using a consistent depression rating in future would aid in keeping heterogeneity consistently low.

Another limitation was in the estimation of LD blocks in the GENDEP cohort. Due to the cohort being composed of individuals across Europe, treating the group as a whole for estimating which SNPs are in LD may have led to inaccuracies. This could explain why many of coefficients in the GENDEP models showed as positive correlation unlike the models from AMPS-1 and AMPS-2. Principal component analysis of treatment centres showed overlapping clusters but they were not distinct enough to warrant calculating LD in each centre separately. Further work in this area should capture more detail on ethnicity and ancestral background, to allow for more robust determination of LD clumps and more informed decisions on the most appropriate inclusion criteria.

Finally, the result may have been impeded by the use of a single PRS predictor. Recent research has shown that the use of multiple scores covering a variety genetic loadings can explain significantly more variance that that of a single score [24]. As such, incorporation of multiple genetic risk scores for outcomes as complex as antidepressant response may prove more fruitful.

Conclusion

Stratified medicine in psychiatry is still in its infancy. Genotyping is not currently routine practice in clinical settings and the use of PRS to guide the use of SSRIs in MDD remains a long-term goal.

However, with increasingly large and well-phenotyped cohorts available for analysis and more powerful GWAS outputs being produced, we tentatively conclude that more targeted prescribing of anti-depressants in MDD based on genetic profiles is a realistic prospect for the future.

Supporting information

S1 Table. Number of SNPs in each PRS for MDD and Neuroticism.
(XLSX)

S2 Table. Results of the individual regressions in the AMPS-1 cohort.
(XLSX)

S3 Table. Results of the induvial regressions in the AMPS-2 cohort.
(XLSX)

S4 Table. Results of the induvial regressions in the GENDEP cohort.
(XLSX)

S5 Table. Results of the meta-analyses.
(XLSX)

S1 Fig. MDD PRS meta-analysis results at 4 weeks. A) $p < 5^{*}10^{-8}$, B) $p < 5^{*}10^{-5}$, C) $p < 0.01$, D) $p < 0.05$, E) $p < 0.1$, F) $p < 0.5$.
(TIF)

S2 Fig. MDD PRS meta-analysis results at 8 weeks A) $p < 5^{*}10^{-8}$, B) $p < 5^{*}10^{-5}$, C) $p < 0.01$, D) $p < 0.05$, E) $p < 0.1$, F) $p < 0.5$.
(TIF)

S3 Fig. Neuroticism PRS meta-analysis results at 4 weeks A) $p < 5^{*}10^{-8}$, B) $p < 5^{*}10^{-5}$, C) $p < 0.01$, D) $p < 0.05$, E) $p < 0.1$, F) $p < 0.5$.
(TIF)

S4 Fig. Neuroticism PRS meta-analysis results at 8 weeks A) $p < 5^{*}10^{-8}$, B) $p < 5^{*}10^{-5}$, C) $p < 0.01$, D) $p < 0.05$, E) $p < 0.1$, F) $p < 0.5.0$.
(TIF)

Acknowledgments

Disclaimer: This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Author Contributions

Conceptualization: Joey Ward, Daniel J. Smith.

Data curation: Karen Hodgson, Joanna Biernacka, Daniel J. Smith.

Investigation: Joey Ward.

Methodology: Joey Ward, Nicholas Graham, Rona J. Strawbridge, Amy Ferguson, Wenan Chen, Cathryn M. Lewis, Joanna Biernacka, Daniel J. Smith.

Resources: Daniel J. Smith.

Writing – original draft: Joey Ward.

Writing – review & editing: Joey Ward, Nicholas Graham, Rona J. Strawbridge, Amy Ferguson, Gregory Jenkins, Wenan Chen, Karen Hodgson, Mark Frye, Richard Weinshilboum, Rudolf Uher, Cathryn M. Lewis, Joanna Biernacka, Daniel J. Smith.

References

1. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL, et al. Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. PLOS Medicine. 2013; 10(11):e1001547. https://doi.org/10.1371/journal.pmed.1001547 PMID: 24223526
2. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry. 2006; 163(11):1905–17. Epub 2006/11/01. https://doi.org/10.1176/ajp.2006.163.11.1905 PMID: 17074942.
3. Linde K, Kriston L, Rucker G, Jamil S, Schumann I, Meissner K, et al. Efficacy and acceptability of pharmacological treatments for depressive disorders in primary care: systematic review and network meta-analysis. Ann Fam Med. 2015; 13(1):69–79. Epub 2015/01/15. https://doi.org/10.1370/afm.1687 PMID: 25583895; PubMed Central PMCID: PMCPMC4291268.
4. Peterson K, Dieperink E, Anderson J, Boundy E, Ferguson L, Helfand M. Rapid evidence review of the comparative effectiveness, harms, and cost-effectiveness of pharmacogenomics-guided antidepressant treatment versus usual care for major depressive disorder. Psychopharmacology (Berl). 2017; 234(11):1649–61. Epub 2017/05/01. https://doi.org/10.1007/s00213-017-4622-9 PMID: 28456840.
5. Dudbridge FM. Power and Predictive Accuracy of Polygenic Risk Scores. PLOS Genetics. 2013; 9(3): e1003348. https://doi.org/10.1371/journal.pgen.1003348 PMID: 23555274
6. Fernández-Ruiz I. New polygenic risk score improves prediction of CHD. Nature Reviews Cardiology. 2016; 13:697. https://doi.org/10.1038/nrcardio.2016.172

7. International Consortium on Lithium G. Association of polygenic score for schizophrenia and hla antigen and inflammation genes with response to lithium in bipolar affective disorder: A genome-wide association study. JAMA Psychiatry. 2017. https://doi.org/10.1001/jamapsychiatry.2017.3433 PMID: 29121268
8. Frokjaer VG, Mortensen EL, Nielsen FA, Haugbol S, Pinborg LH, Adams KH, et al. Frontolimbic serotonin 2A receptor binding in healthy subjects is associated with personality risk factors for affective disorder. Biol Psychiatry. 2008; 63(6):569–76. Epub 2007/09/22. https://doi.org/10.1016/j.biopsych.2007.07.009 PMID: 17884017.
9. Di Simplicio M, Norbury R, Reinecke A, Harmer CJ. Paradoxical effects of short-term antidepressant treatment in fMRI emotional processing models in volunteers with high neuroticism. Psychol Med. 2014; 44(2):241–52. Epub 2013/04/20. https://doi.org/10.1017/S0033291713000731 PMID: 23597109.
10. Katon W, Unutzer J, Russo J. Major depression: the importance of clinical characteristics and treatment response to prognosis. Depress Anxiety. 2010; 27(1):19–26. Epub 2009/10/03. https://doi.org/10.1002/da.20613 PMID: 19798766.
11. Steffens DC, McQuoid DR, Smoski MJ, Potter GG. Clinical outcomes of older depressed patients with and without comorbid neuroticism. Int Psychogeriatr. 2013; 25(12):1985–90. Epub 2013/08/15. https://doi.org/10.1017/S1041610213001324 PMID: 23941723; PubMed Central PMCID: PMCPMC3830609.
12. Ji Y, Biernacka JM, Hebbring S, Chai Y, Jenkins GD, Batzler A, et al. Pharmacogenomics of selective serotonin reuptake inhibitor treatment for major depressive disorder: genome-wide associations and functional genomics. The pharmacogenomics journal. 2013; 13(5):456–63. https://doi.org/10.1038/tpj.2012.32 PubMed PMID: PMC3941038. PMID: 22907730
13. Biernacka JM, Sangkuhl K, Jenkins G, Whaley RM, Barman P, Batzler A, et al. The International SSRI Pharmacogenomics Consortium (ISPC): a genome-wide association study of antidepressant treatment response. Transl Psychiatry. 2015; 5:e553. https://doi.org/10.1038/tp.2015.47 PMID: 25897834
14. Carneiro AM, Fernandes F, Moreno RA. Hamilton depression rating scale and montgomery–asberg depression rating scale in depressed and bipolar I patients: psychometric properties in a Brazilian sample. Health and Quality of Life Outcomes. 2015; 13:42. https://doi.org/10.1186/s12955-015-0235-3 PubMed PMID: PMC4391145. PMID: 25889742
15. Uher R, Perroud N, Ng MYM, Hauser J, Henigsberg N, Maier W, et al. Genome-Wide Pharmacogenetics of Antidepressant Response in the GENDEP Project. American Journal of Psychiatry. 2010; 167(5):555–64. https://doi.org/10.1176/appi.ajp.2009.09070932 PMID: 20360315
16. 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):559–75. PubMed PMID: PMC1950838. https://doi.org/10.1086/519795 PMID: 17701901
17. Smith DJ, Escott-Price V, Davies G, Bailey MES, Colodro-Conde L, Ward J, et al. Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. Molecular Psychiatry. 2016; 21(6):749–57. https://doi.org/10.1038/mp.2016.49 PubMed PMID: PMC4879189. PMID: 27067015
18. Howard DM, Adams MJ, Shirali M, Clarke T-K, Marioni RE, Davies G, et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nature Communications. 2018; 9(1):1470. https://doi.org/10.1038/s41467-018-03819-3 PMID: 29662059
19. Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. Genome Medicine. 2017; 9:96. https://doi.org/10.1186/s13073-017-0489-y PubMed PMID: PMC5683372. PMID: 29132412
20. Selzam S, Krapohl E, von Stumm S, O'Reilly PF, Rimfeld K, Kovas Y, et al. Predicting educational achievement from DNA. Molecular Psychiatry. 2016; 22:267. https://doi.org/10.1038/mp.2016.107 https://www.nature.com/articles/mp2016107#supplementary-information. PMID: 27431296
21. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software; Vol 1, Issue 3 (2010). 2010. https://doi.org/10.18637/jss.v036.i03
22. Palla L, Dudbridge F. A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. Am J Hum Genet. 2015; 97(2):250–9. Epub 2015/07/21. https://doi.org/10.1016/j.ajhg.2015.06.005 PMID: 26189816; PubMed Central PMCID: PMCPMC4573448.
23. Garcia-Gonzalez J, Tansey KE, Hauser J, Henigsberg N, Maier W, Mors O, et al. Pharmacogenetics of antidepressant response: A polygenic approach. Prog Neuropsychopharmacol Biol Psychiatry. 2017; 75:128–34. Epub 2017/02/06. https://doi.org/10.1016/j.pnpbp.2017.01.011 PMID: 28159590.
24. Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, et al. Multi-polygenic score approach to trait prediction. Molecular Psychiatry. 2017. https://doi.org/10.1038/mp.2017.163 https://www.nature.com/articles/mp2017163#supplementary-information. PMID: 28785111

4.1.1 Appendix A

4.1.1.1 Table S1. Number of SNPs in each PRS for MDD and Neuroticism.

cohort	AMPS 1 and 2			GENDEP			Absolute difference
	PRS	**p value cut off**	**number of SNP**	**PRS**	**p value cut off**	**number of SNP**	
	MDD	$5*10^{-8}$	24	MDD	$5*10^{-8}$	24	0
		$5*10^{-5}$	1096		$5*10^{-5}$	1216	120
		0.01	91778		0.01	115428	23650
		0.05	400714		0.05	505328	104614
		0.1	757322		0.1	955156	197834
		0.5	3275666		0.5	4162482	886816
	Neuroticism	$5*10^{-8}$	2066	Neuroticism	$5*10^{-8}$	1938	128
		$5*10^{-5}$	9760		$5*10^{-5}$	8016	1744
		0.01	183022		0.01	167242	15780
		0.05	613656		0.05	588642	25014
		0.1	1057592		0.1	1036136	21456
		0.5	3951360		0.5	4060670	109310

4.1.1.2 **Table S2.** Results of the individual regressions in the AMPS-1 cohort.

Time point (week)	predictor	P	Beta	SE	T stat	R^2
4	mdd $p<5*10^{-8}$	0.24	-0.014	0.0118	-1.17	0.01
	mdd $p<5*10^{-5}$	0.03	-0.013	0.0122	-1.05	0.008
	mdd p<0.01	0.57	-0.007	0.0122	-0.571	0.002
	mdd p<0.05	0.43	-0.0096	0.012	-0.799	0.005
	mdd p<0.1	0.4	-0.011	0.0125	-0.853	0.005
	mdd p<0.5	0.12	-0.02	0.0127	-1.58	0.017
8	mdd $p<5*10^{-8}$	0.19	-0.015	0.0116	-1.31	0.012
	mdd $p<5*10^{-5}$	0.18	-0.017	0.0122	-1.35	0.013
	mdd p<0.01	0.14	-0.016	0.0107	-1.48	0.015
	mdd p<0.05	0.49	-0.0077	0.0111	-0.695	0.003
	mdd p<0.1	0.47	-0.0078	0.0108	-0.72	0.004
	mdd p<0.5	0.11	-0.018	0.0111	-1.59	0.017
4	neuroticism $p<5*10^{-8}$	0.77	-0.0036	0.0125	-0.288	0.0006
	neuroticism $p<5*10^{-5}$	0.46	-0.0095	0.0128	-0.744	0.004
	neuroticism p<0.01	0.85	-0.0024	0.0126	-0.192	0.0003
	neuroticism p<0.05	0.39	-0.011	0.0124	-0.87	0.005
	neuroticism p<0.1	0.7	-0.0048	0.0124	-0.385	0.001
	neuroticism p<0.5	0.44	0.0091	0.0118	0.769	0.004
8	neuroticism $p<5*10^{-8}$	0.88	-0.0017	0.0115	-0.146	0.0002
	neuroticism $p<5*10^{-5}$	0.87	-0.0018	0.0115	-0.159	0.0002
	neuroticism p<0.01	0.68	-0.0049	0.0117	-0.42	0.001
	neuroticism p<0.05	0.17	-0.016	0.0116	-1.39	0.013
	neuroticism p<0.1	0.15	-0.016	0.0113	-1.43	0.015
	neuroticism p<0.5	0.95	0.0007	0.0105	0.0661	0.00003

4.1.1.3 Table S3. Results of the induvial regressions in the AMPS-2 cohort.

Time point (week)	predictor	P	Beta	SE	T stat	R^2
4	mdd $p<5*10^{-8}$	0.8	0.004	0.02	0.25	0.001
	mdd $p<5*10^{-5}$	0.09	-0.03	0.02	-1.73	0.039
	mdd p<0.01	0.06	-0.03	0.02	-1.93	0.066
	mdd p<0.05	0.32	-0.01	0.01	-1.01	0.018
	mdd p<0.1	0.6	-0.01	0.02	-0.52	0.005
	mdd p<0.5	0.58	-0.01	0.02	-0.55	0.005
8	mdd $p<5*10^{-8}$	0.78	-0.01	0.02	-0.28	0.001
	mdd $p<5*10^{-5}$	0.25	-0.02	0.02	-1.18	0.024
	mdd p<0.01	0.78	-0.01	0.02	-0.28	0.001
	mdd p<0.05	0.23	-0.02	0.02	-1.22	0.027
	mdd p<0.1	0.4	-0.02	0.02	-0.84	0.011
	mdd p<0.5	0.15	-0.03	0.02	-1.45	0.037
4	neuroticism $p<5*10^{-8}$	0.2	-0.03	0.02	-1.3	0.03
	neuroticism $p<5*10^{-5}$	0.37	-0.01	0.02	-0.91	0.01
	neuroticism p<0.01	0.079	-0.03	0.02	-1.79	0.06
	neuroticism p<0.05	0.14	-0.03	0.02	-1.49	0.04
	neuroticism p<0.1	0.28	-0.02	0.02	-1.1	0.02
	neuroticism p<0.5	0.019	-0.04	0.02	-2.42	0.099
8	neuroticism $p<5*10^{-8}$	0.32	-0.02	0.02	-1	0.016
	neuroticism $p<5*10^{-5}$	0.28	-0.02	0.02	-1.09	0.019
	neuroticism p<0.01	0.07	-0.03	0.02	-1.82	0.05
	neuroticism p<0.05	0.62	-0.01	0.02	-0.49	0.004
	neuroticism p<0.1	0.52	-0.01	0.02	-0.64	0.007
	neuroticism p<0.5	0.02	-0.04	0.02	-2.26	0.08

4.1.1.4 Table S4. Results of the induvial regressions in the GENDEP cohort.

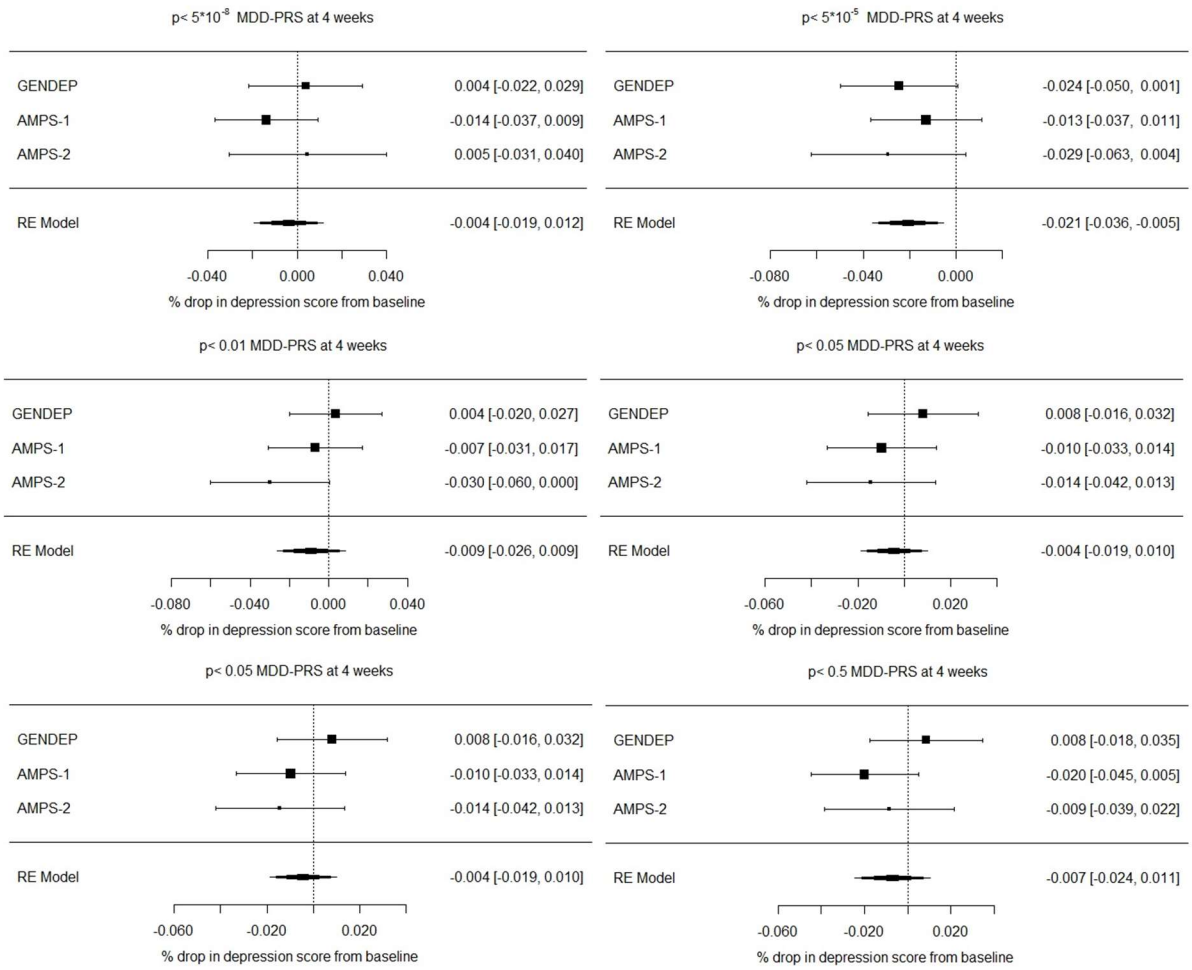
Time point (week)	predictor	P	Beta	SE	T stat	R^2
4	mdd $p<5*10^{-8}$	0.77	0.004	0.013	0.3	0.0008
	mdd $p<5*10^{-5}$	0.06	-0.02	0.013	-1.9	0.026
	mdd p<0.01	0.77	0.003	0.012	0.3	0.0007
	mdd p<0.05	0.51	0.008	0.012	0.66	0.003
	mdd p<0.1	0.62	0.006	0.013	0.5	0.002
	mdd p<0.5	0.52	0.009	0.013	0.64	0.003
8	mdd $p<5*10^{-8}$	0.18	0.019	0.014	1.35	0.02
	mdd $p<5*10^{-5}$	0.49	-0.012	0.017	-0.69	0.004
	mdd p<0.01	0.73	0.006	0.017	0.35	0.001
	mdd p<0.05	0.28	0.018	0.017	1.08	0.009
	mdd p<0.1	0.43	0.013	0.016	0.8	0.005
	mdd p<0.5	0.2	0.02	0.016	1.29	0.01
4	neuroticism $p<5*10^{-8}$	0.71	-0.004	0.011	-0.37	0.001
	neuroticism $p<5*10^{-5}$	0.62	-0.006	0.012	-0.5	0.002
	neuroticism p<0.01	0.29	-0.013	0.012	-1.06	0.009
	neuroticism p<0.05	0.97	-0.001	0.012	-0.04	0.00001
	neuroticism p<0.1	0.33	-0.01	0.012	-0.97	0.008
	neuroticism p<0.5	0.61	-0.006	0.012	-0.5	0.002
8	neuroticism $p<5*10^{-8}$	0.99	0.0001	0.012	0.005	0.0000002
	neuroticism $p<5*10^{-5}$	0.97	0.0005	0.013	0.04	0.00001
	neuroticism p<0.01	0.92	0.002	0.018	0.095	0.00008
	neuroticism p<0.05	0.36	-0.012	0.013	-0.91	0.007
	neuroticism p<0.1	0.15	-0.019	0.013	-1.47	0.02
	neuroticism p<0.5	0.43	-0.01	0.013	-0.79	0.005

4.1.1.5 Table S5. Results of the meta-analyses.

outcome	PRS	beta	SE	Z	P	ci.lb	ci.ub	I^2
% drop in 4 weeks	MDD p < $5*10^{-8}$	-0.004	0.008	-0.49	0.63	-0.02	0.01	0
	MDD p < $5*10^{-5}$	-0.021	0.008	-2.63	0.01	-0.04	-0.01	0
	MDD p < 0.01	-0.009	0.009	-1	0.32	-0.03	0.01	28.2
	MDD p < 0.05	-0.004	0.007	-0.61	0.54	-0.02	0.01	0
	MDD p < 0.1	-0.004	0.008	-0.48	0.63	-0.02	0.01	0
	MDD p < 0.5	-0.007	0.009	-0.77	0.44	-0.02	0.01	21.6
% drop in 8 weeks	MDD p < $5*10^{-8}$	-0.001	0.012	-0.09	0.93	-0.02	0.02	45.3
	MDD p < $5*10^{-5}$	-0.017	0.009	-1.86	0.06	-0.03	0	0
	MDD p < 0.01	-0.009	0.008	-1.08	0.28	-0.02	0.01	0
	MDD p < 0.05	-0.004	0.01	-0.36	0.72	-0.02	0.02	23.9
	MDD p < 0.1	-0.004	0.008	-0.5	0.62	-0.02	0.01	0
	MDD p < 0.5	-0.008	0.014	-0.58	0.57	-0.04	0.02	61.5
% drop in 4 weeks	N p < $5*10^{-8}$	-0.007	0.008	-0.94	0.35	-0.02	0.01	0
	N p < $5*10^{-5}$	-0.009	0.008	-1.17	0.24	-0.02	0.01	0
	N p < 0.01	-0.012	0.008	-1.56	0.12	-0.03	0	3
	N p < 0.05	-0.009	0.008	-1.17	0.24	-0.02	0.01	0
	N p < 0.1	-0.01	0.008	-1.33	0.18	-0.03	0	0
	N p < 0.5	-0.011	0.015	-0.75	0.45	-0.04	0.02	70.4
% drop in 8 weeks	N p < $5*10^{-8}$	-0.004	0.008	-0.49	0.62	-0.02	0.01	0
	N p < $5*10^{-5}$	-0.004	0.008	-0.55	0.58	-0.02	0.01	0
	N p < 0.01	-0.01	0.009	-1.09	0.27	-0.03	0.01	2.3
	N p < 0.05	-0.013	0.008	-1.7	0.09	-0.03	0	0
	N p < 0.1	-0.017	0.008	-2.13	0.03	-0.03	0	0
	N p < 0.5	-0.013	0.011	-1.2	0.23	-0.03	0.01	48

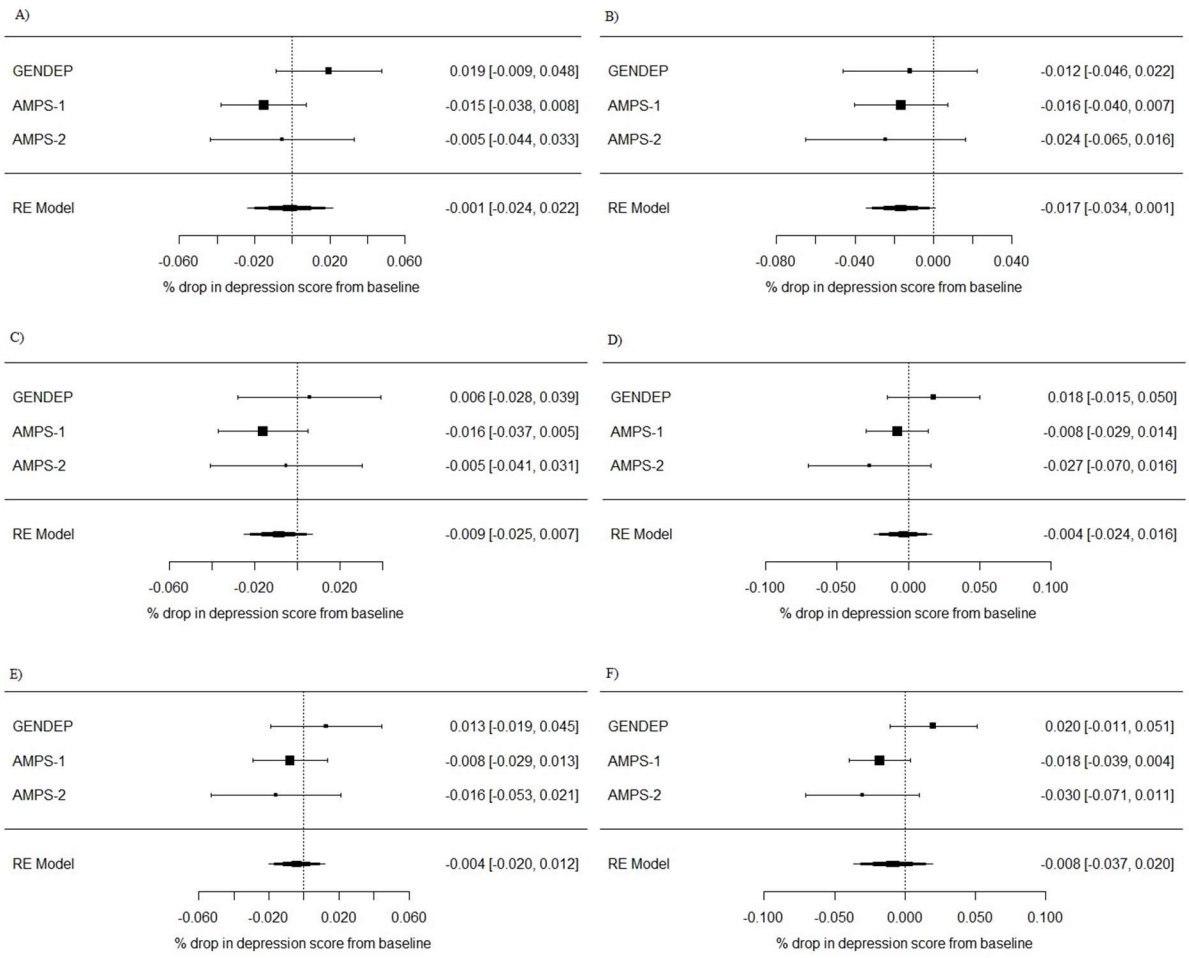
4.1.1.6 **Fig S1.** MDD PRS meta-analysis results at 4 weeks.

A) $p < 5*10^{-8}$, B) $p < 5*10^{-5}$, C) $p < 0.05$, E) $p < 0.1$, F) $p < 0.5$.



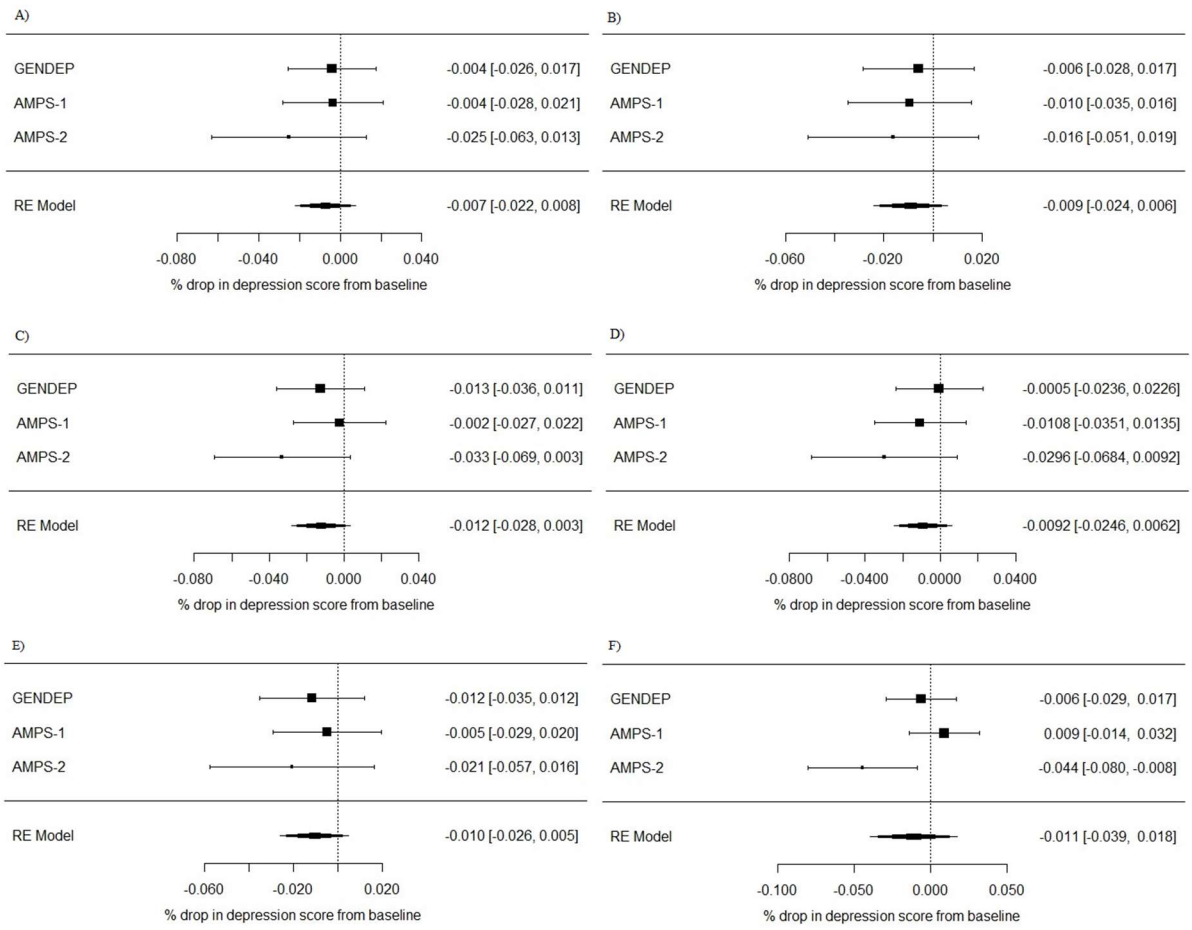
4.1.1.7 **Fig S2.** MDD PRS meta-analysis results at 8 weeks

A) $p < 5*10^{-8}$, B) $p < 5*10^{-5}$, C) $p < 0.05$, E) $p < 0.1$, F) $p < 0.5$.



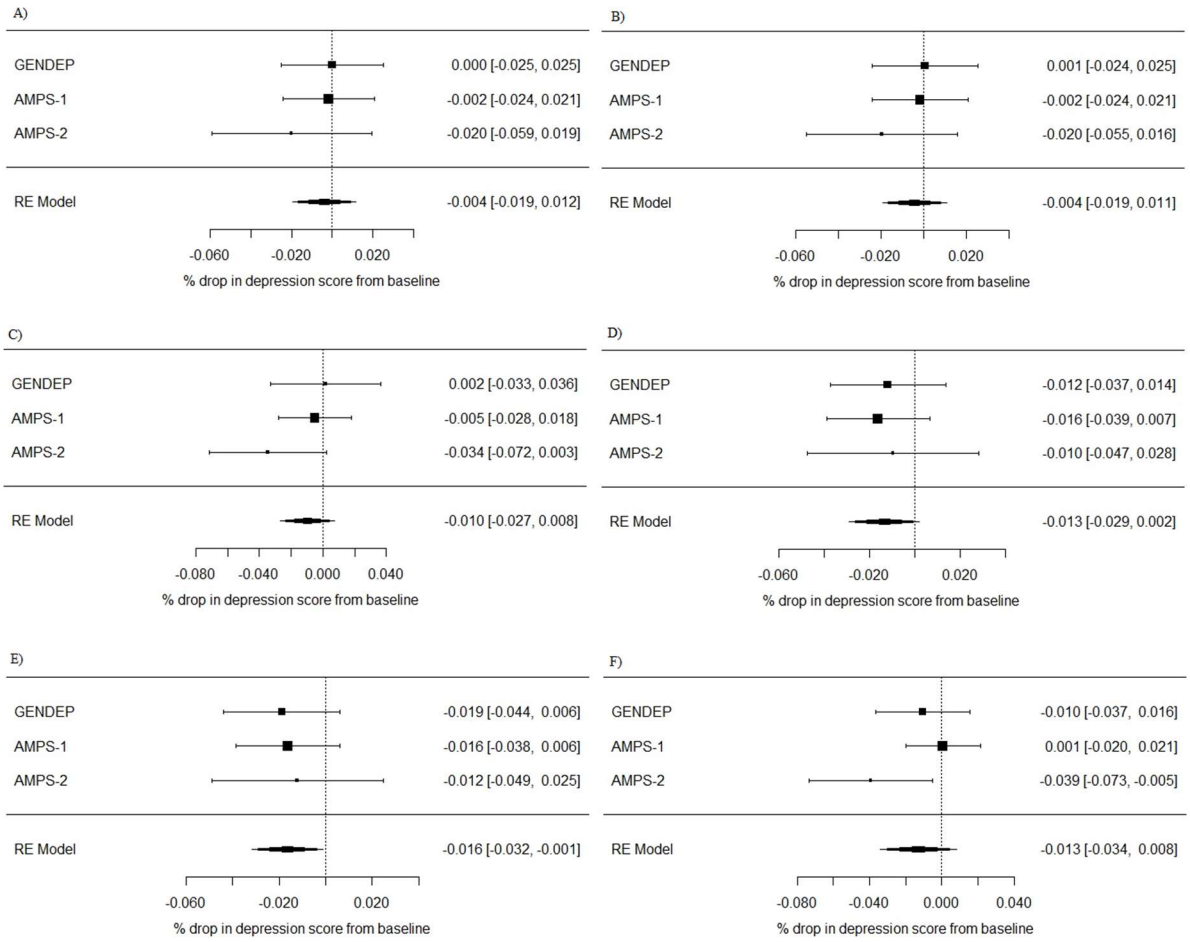
4.1.1.8 **Fig S3.** Neuroticism PRS meta-analysis results at 4 weeks

A) $p < 5*10^{-8}$, B) $p < 5*10^{-5}$, C) $p < 0.05$, E) $p < 0.1$, F) $p < 0.5$.



4.1.1.9 **Fig S4.** Neuroticism PRS meta-analysis results at 8 weeks

A) $p < 5*10^{-8}$, B) $p < 5*10^{-5}$, C) $p < 0.05$, E) $p < 0.1$, F) $p < 0.5$.



4.2 Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia

Ward et al. Translational Psychiatry (2017)7:1264
DOI 10.1038/s41398-017-0012-7

Translational Psychiatry

ARTICLE Open Access

Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia

Joey Ward[1], Rona J. Strawbridge[1,2], Mark E. S. Bailey[3], Nicholas Graham[1], Amy Ferguson[1], Donald M. Lyall[1], Breda Cullen[1], Laura M. Pidgeon[1], Jonathan Cavanagh[1], Daniel F. Mackay[1], Jill P. Pell[1], Michael O'Donovan[4], Valentina Escott-Price[4] and Daniel J. Smith[1]

Abstract

Mood instability is a core clinical feature of affective and psychotic disorders. In keeping with the Research Domain Criteria approach, it may be a useful construct for identifying biology that cuts across psychiatric categories. We aimed to investigate the biological validity of a simple measure of mood instability and evaluate its genetic relationship with several psychiatric disorders, including major depressive disorder (MDD), bipolar disorder (BD), schizophrenia, attention deficit hyperactivity disorder (ADHD), anxiety disorder and post-traumatic stress disorder (PTSD). We conducted a genome-wide association study (GWAS) of mood instability in 53,525 cases and 60,443 controls from UK Biobank, identifying four independently associated loci (on chromosomes 8, 9, 14 and 18), and a common single-nucleotide polymorphism (SNP)-based heritability estimate of ~8%. We found a strong genetic correlation between mood instability and MDD ($r_g = 0.60$, SE = 0.07, $p = 8.95 \times 10^{-17}$) and a small but significant genetic correlation with both schizophrenia ($r_g = 0.11$, SE = 0.04, $p = 0.01$) and anxiety disorders ($r_g = 0.28$, SE = 0.14, $p = 0.04$), although no genetic correlation with BD, ADHD or PTSD was observed. Several genes at the associated loci may have a role in mood instability, including the *DCC netrin 1 receptor* (*DCC*) gene, *eukaryotic translation initiation factor 2B subunit beta* (*eIF2B2*), *placental growth factor* (*PGF*) and *protein tyrosine phosphatase, receptor type D* (*PTPRD*). Strengths of this study include the very large sample size, but our measure of mood instability may be limited by the use of a single question. Overall, this work suggests a polygenic basis for mood instability. This simple measure can be obtained in very large samples; our findings suggest that doing so may offer the opportunity to illuminate the fundamental biology of mood regulation.

Introduction

Mood instability is a common clinical feature of affective and psychotic disorders, particularly major depressive disorder (MDD), bipolar disorder (BD) and schizophrenia[1]. It may also be relatively common in the general population, estimated to affect ~13% of individuals[2]. As a dimensional psychopathological trait, it is potentially a useful construct in line with the Research Domain Criteria approach[3]. Mood instability may be of fundamental importance for understanding the pathophysiology of MDD and BD, as well as conditions such as borderline personality disorder, anxiety disorders,

Correspondence: Daniel J Smith (daniel.smith@glasgow.ac.uk)
[1]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK
[2]Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden
Full list of author information is available at the end of the article

© The Author(s) 2017
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

attention deficit hyperactivity disorder (ADHD) and psychosis[4]. This trait is reported by 40–60% of individuals with MDD[5] and is recognised as part of the prodromal stage of BD[6]. In established BD, it is a clinical feature that independently predicts poor functional outcome[7]. Furthermore, general population twin studies suggest that additive genetic effects account for 40% of the variance in measures of affect intensity and 25% of the variance in affective liability[8].

Population-based studies such as the Adult Psychiatric Morbidity Survey (APMS) have defined mood instability based on responses to a single question, while clinical studies have made use of more detailed rating scales[4]. However, there is a lack of consensus about how best to measure and classify mood instability, and none of the currently available instruments adequately capture intensity, speed and frequency of affective change, or physiological and behavioural correlates. A recent systematic review proposed that mood instability be defined as 'rapid oscillations of intense affect, with a difficulty in regulating these oscillations or their behavioural consequences'[9]. Applying this definition will require the future development and validation of a multidimensional assessment of mood instability, which is currently not available.

Within the UK Biobank population cohort of over 0.5 million individuals[10], the baseline assessment interview contained a question of relevance to mood instability, specifically: *'Does your mood often goes up and down?'* This is similar to the question for mood instability used within the APMS (*'Do you have a lot of sudden mood changes, suffered over the last several years'*). Hypothesising that this simple question taps into pathological mood instability, we predicted that it would be more commonly endorsed by individuals within UK Biobank with MDD and BD, compared to individuals with no psychiatric disorder. Moreover, under the hypothesis that this trait has cross-disorder pathophysiological relevance, we predicted that a genome-wide association study (GWAS) might identify shared genetic liability to mood instability and risk for psychiatric disorders in which disordered mood is a feature, including MDD, BD, schizophrenia, ADHD, anxiety disorder and post-traumatic stress disorder (PTSD). Given the size of the sample, we also aimed to identify loci associated with this measure of mood instability.

Materials and methods

Sample

UK Biobank is a large cohort of more than 502,000 United Kingdom residents, aged between 40 and 69 years[10]. The aim of UK Biobank is to study the genetic, environmental and lifestyle factors that cause or prevent disease in middle and older age. Baseline assessments occurred over a 4-year period, from 2006 to 2010, across 22 United Kingdom (UK) centres. These assessments were comprehensive and included social, cognitive, lifestyle and physical health measures. For the present study, we used the first genetic data release based on approximately one-third of UK Biobank participants. Aiming to maximise homogeneity, we restricted the sample to those who reported being of white UK ancestry (around 95% of the sample).

UK Biobank obtained informed consent 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).

Mood instability phenotype

As part of the baseline assessment, UK Biobank participants completed the 12 items of the neuroticism scale from the Eysenck Personality Questionnaire-Revised Short Form (EPQ-R-S)[11]. One of these items assesses mood instability, namely *'Does your mood often goes up and down?'* Participants responding 'yes' to this question were considered to be cases of mood instability and those responding 'no' were considered controls. From the control sample, we excluded those who reported being on psychotropic medication, and those who reported a physician diagnosis of psychiatric disorder (including MDD, BD, anxiety/panic attacks, 'nervous breakdown', schizophrenia and deliberate self-harm/suicide attempt).

After quality-control steps (detailed below) and exclusions (3679 participants responded 'don't know' and 211 responded 'prefer not to say'), the final sample for genetic analysis comprised 53,525 cases of mood instability and 60,443 controls. Mood instability cases were younger than controls (mean age 55.8 years (SD = 8.05) vs. 57.7 years (SD = 7.74); $p < 0.0001$) and had a greater proportion of females (55.5% vs. 49.6%; $p < 0.0001$).

Genotyping and imputation

In June 2015, UK Biobank released the first set of genotypic data for 152,729 UK Biobank participants. Approximately 67% of this sample was genotyped using the Affymetrix UK Biobank Axiom array (Santa Clara, CA, USA) and the remaining 33% were genotyped using the Affymetrix UK BiLEVE Axiom array. These arrays have over 95% content in common. Only autosomal data were available under the data release. Data were pre-imputed by UK Biobank as fully described in the UK Biobank interim release documentation[12]. Briefly, after removing genotyped single-nucleotide polymorphisms (SNPs) that were outliers, or were multiallelic or of low frequency (minor allele frequency (MAF) < 1%), phasing

was performed using a modified version of SHAPEIT2 and imputation was carried out using IMPUTE2 algorithms, as implemented in a C++ platform for computational efficiency[13, 14]. Imputation was based upon a merged reference panel of 87,696,888 biallelic variants on 12,570 haplotypes constituted from the 1000 Genomes Phase 3 and UK10K haplotype panels[15]. Variants with MAF < 0.001% were excluded from the imputed marker set. Stringent quality control before release was applied by the Wellcome Trust Centre for Human Genetics, as described in UK Biobank documentation[16].

Statistical analyses

Quality control and association analyses

Before all analyses, further quality-control measures were applied. Individuals were removed based on UK Biobank genomic analysis exclusions (Biobank Data Dictionary item #22010), relatedness (#22012: genetic relatedness factor; a random member of each set of individuals with KING-estimated kinship coefficient >0.0442 was removed), gender mismatch (#22001: genetic sex), ancestry (#22006: ethnic grouping; principal component (PC) analysis identified probable Caucasians within those individuals who were self-identified as British and other individuals were removed from the analysis), and quality-control failure in the UK BiLEVE study (#22050: UK BiLEVE Affymetrix quality control for samples and #22051: UK BiLEVE genotype quality control for samples). A sample of 113,968 individuals remained for further analyses. Of these, 53,525 were classed as cases and 60,443 were classified as controls. Genotype data were further filtered by removal of SNPs with Hardy–Weinberg equilibrium $P < 10^{-6}$, with MAF < 0.01, with imputation quality score <0.4 and with data on <90% of the sample after excluding genotype calls made with <90% posterior probability, after which 8,797,848 variants were retained.

Association analysis was conducted in PLINK[17] using logistic regression under a model of additive allelic effects with sex, age, genotyping array and the first eight PCs (Biobank Data Dictionary items #22009.01 to #22009.08) as covariates. Sex and age were included as covariates because cases and controls differed significantly on these measures. Genetic PCs were included to control for hidden population structure within the sample, and the first eight PCs, out of 15 available in the Biobank, were selected after visual inspection of each pair of PCs, taking forward only those that resulted in multiple clusters of individuals after excluding individuals self-reporting as being of non-white British ancestry (Biobank Data Dictionary item #22006). Overall, population structure had little impact on mood instability status. The threshold for genome-wide significance was $p < 5.0 \times 10^{-8}$.

Heritability and genetic correlation between mood instability and psychiatric phenotypes

We applied Linkage Disequilibrium Score Regression (LDSR)[18] to the GWAS summary statistics to estimate SNP heritability (h^2_{SNP}). Genetic correlations between mood instability and MDD, BD, schizophrenia, ADHD, anxiety disorder and PTSD were also evaluated using LDSR[19] (with unconstrained intercept), a process that corrects for potential sample overlap without relying on the availability of individual genotypes[18]. For the MDD, BD, schizophrenia, ADHD, anxiety disorder and PTSD phenotypes, we used GWAS summary statistics provided by the Psychiatric Genomics Consortium (http://www.med.unc.edu/pgc/)[20–25]. Note that for the purposes of these genetic correlation analyses we re-ran the GWAS of mood instability excluding from the cases those 9865 participants who reported being on psychotropic medication, or who self-reported psychiatric disorder (MDD, BD, anxiety/panic attacks, 'nervous breakdown', schizophrenia and deliberate self-harm/suicide attempt). This secondary GWAS output (rather than the primary GWAS reported below) was used for the genetic correlation calculations and for polygenic risk score (PRS) analyses, the rationale being that this was a more conservative approach that would avoid genetic correlations between mood instability and MDD/BD/schizophrenia/ADHD/anxiety disorders/PTSD being driven by a subset of individuals with psychiatric disorder.

PRS analysis of MDD, BD and schizophrenia as predictors of mood instability

PRSs were created using the output of the PCG MDD 29 of 32 cohort GWAS (supplied by the MDD working group of the PGC, http://www.med.unc.edu/pgc/pgc-workgroups), BD GWAS[20] and schizophrenia GWAS[21]. Five PRS were created for each psychiatric phenotype using p value cutoffs of $p < 5 \times 10^{-8}$, $p < 0.01$, $p < 0.05$, $p < 0.1$ and $p < 0.5$, with the exception of MDD for which there were no genome-wide significant SNPs. Ambiguous SNPs, indels (insertion/deletion mutations) and SNPs with an imputation quality score of less than 0.8 were removed. LD clumping was performed via PLINK on a random sample of 10,000 individuals using an $r^2 > 0.05$ in a 250 kb window. SNPs were clumped into sets and filtered, selecting the SNP with the lowest p value from each set. In the event that two or more SNPs from a set had the same p value, the SNP with the largest beta coefficient was used. PLINK was also used to calculate the PRS to produce a per-allele weighted score with no mean imputation.

PRS modelling

Only those subjects who were used for the genetic correlation analyses were used in the PRS analyses (that is,

PRS analyses also excluded from both case and control groups those individuals in UK Biobank with psychiatric disorder). Modelling was performed in R (version 3.1.2) using the glm function. Full sample and age-stratified analysis models were adjusted for age, sex, chip and PGCs 1–8, whereas sex-stratified analysis was not adjusted for sex. Scores were then split into deciles using the ntile function of the dplyr package. Model Nagelkerke r^2 was calculated using the fmsb package.

Results

Mood instability in MDD and BD within UK Biobank

In previous work we have identified individuals within UK Biobank with a probable diagnosis of mood disorder, including cases of MDD (subdivided into single-episode MDD, recurrent moderate MDD and recurrent severe MDD) and BD, as well as non-mood disordered controls[26]. These classifications were independent of response to the mood instability question or other questions from the EPQ-R-S. For the group of participants who could be classified in this way, we assessed the proportion with mood instability within each mood disorder category. All mood disorder groups had a significantly greater proportion of individuals with mood instability compared with the control group (Table 1), in which the prevalence was 35.3%. This proportion was highest in the BD group (74.0%) followed by the three MDD groups (71.7% for recurrent severe MDD, 64.2% for recurrent moderate MDD and 43.7% for single-episode MDD). There were too few UK Biobank participants with a reliable classification of schizophrenia, ADHD, anxiety disorder or PTSD to allow for an assessment of the prevalence of mood instability in these groups.

GWAS of mood instability

The mood instability GWAS results are summarised in Fig. 1 (Manhattan plot), Fig. 2 (QQ plot) and Table 2 (genome-wide significant loci associated with mood instability). Regional plots are provided in Figs. 3a–d.

Overall, the GWAS data showed modest deviation in the test statistics compared with the null ($\lambda_{GC} = 1.13$); this was negligible in the context of sample size (λ_{GC} 1000 = 1.002). LDSR suggested that deviation from the null was due to a polygenic architecture in which h^2_{SNP} accounted for ~8% of the population variance in mood instability (observed scale $h^2_{SNP} = 0.077$ (SE 0.007)), rather than inflation due to unconstrained population structure (LD regression intercept = 0.998 (SE 0.009)).

We observed four independent genomic loci exhibiting genome-wide significant associations with mood instability (Fig. 1, Table 2 and Figs. 3a–d), on chromosome 8 (index SNP rs7829975; *CLDN23 and MFHAS1*), chromosome 9 (index SNP rs10959826; *PTPRD*), chromosome 14 (index SNP rs397852991; *LTBP2, AREL1, FCF1, YLPM1, PROX2, DLST, RPS6KL1, PGF, EIF2B2* and *MLH3*) and chromosome 18 (index SNP rs8084280; *DCC*). In total, there were 111 genome-wide significant

Table 1 Proportion of individuals with mood instability within mood disorder groups, compared to non-mood disordered controls

	Mood instability *N* (%)	Pearson X^2	*P* value
BD	1180 (74.0)	1.0×10^3	<0.001
Recurrent MDD, severe	6303 (71.7)	4.5×10^3	<0.001
Recurrent MDD, moderate	9509 (64.2)	4.4×10^3	<0.001
Single-episode MDD	3403 (43.7)	221.1	<0.001
Non-mood disordered controls	30,844 (35.3)	–	–

BD bipolar disorder, *MDD* major depressive disorder

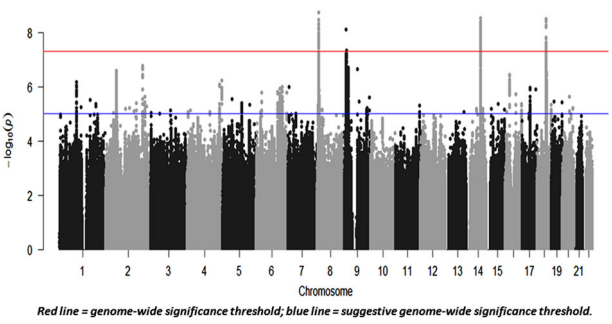


Fig. 1 Manhattan plot of GWAS of mood instability in UK Biobank (n = 113,968)

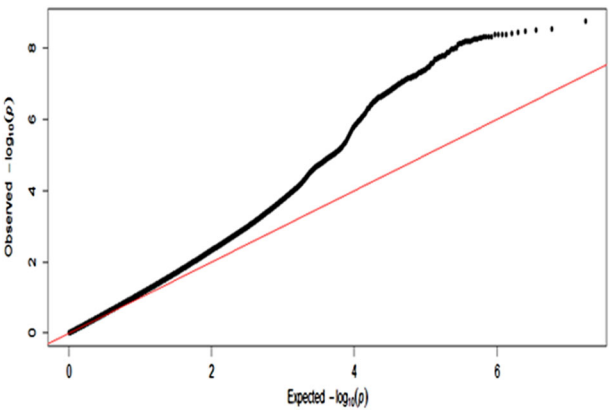


Fig. 2 QQ plot for UK Biobank mood instability GWAS results

Table 2 Genome-wide significant loci associated with mood instability in UK Biobank

Index SNP	Chr	Position	Risk allele/other allele	RAF	Beta (SE)	*P* value	Associated region	Nearby genes
rs7829975	8	8,548,117	A/T	0.516	0.051 (0.0085)	1.8×10^{-9}	8,548,117–8,704,330	*CLDN23, MFHAS1*
rs10959826	9	11,459,410	G/A	0.785	0.060 (0.01)	7.7×10^{-9}	11,459,410–11,701,596	*PTPRD*
rs397852991	14	75,268,920	C/CA	0.606	0.053 (0.0088)	2.98×10^{-9}	75,144,618–75,359,229	*LTBP2, AREL1, FCF1, YLPM1, PROX2, DLST, RPS6KL1, PGF, EIF2B2, MLH3*
rs8084280	18	50,726,749	T/A	0.508	0.050 (0.0085)	3.15×10^{-9}	50,635,119–50,893,647	*DCC*

Shown are LD-independent genome-wide significant SNP associations for mood instability (sorted by genomic position according to NCBI Build 37). Chromosome (Chr) and Position denote the location of the index SNP. The final column indicates protein-coding genes at the associated loci (see regional plots in supplementary information) or, where there are no genes at the associated locus, the nearest gene if less than 1 MB from the locus. *RAF* risk allele frequency, *Beta* logistic regression coefficient for allele1, *SE* standard error for Beta. *P* value the probability of getting the derived test statistic under the null hypothesis

SNPs across all loci. Given the functional alleles that drive association signals in GWAS may not affect the nearest gene, we use the above gene names to provide a guide to location rather than to imply that altered function or expression of those genes are the sources of the association signals.

We also repeated this GWAS for males and females separately (Supplementary Figs. S1 and S2) and for the sample stratified according to median age (age 58 and below, and age 59 and above; Supplementary Figs. S3 and S4). No genome-wide significant loci were observed from these stratified analyses, possibly because of reduced power, apart from the retention of a single genome-wide significant finding at rs8084280 on chromosome 18 (the *DCC* gene) for males only (Supplementary Fig. S1). There was a high degree of genetic correlation between mood instability in males and females ($r_g = 1.02$, SE = 0.09, $p = 2.84 \times 10^{-30}$), and between mood instability in the younger and older subgroups ($r_g = 1.02$, SE = 0.09, $p = 2.67 \times 10^{-27}$).

Within supplementary materials, we also present the results of the secondary GWAS of mood instability that was excluded from the case group of 9865 participants with a psychiatric disorder (Supplementary Table S1). This GWAS was used to assess for genetic correlation between mood instability and MDD, BD, schizophrenia, ADHD, anxiety disorders and PTSD, and for the PRS analyses. Supplementary Table S1 shows that the risk allele frequencies (RAFs) of the index SNPs within the four genome-wide significant loci from the primary GWAS were very similar to the RAFs for these same SNPs within this secondary GWAS: for rs7829975 the RAF was 0.516 vs. 0.523; for rs10959826 it was 0.785 vs. 0.789; for rs397852991 it was 0.606 vs. 0.673; and for rs8084280 it was 0.508 vs. 0.514). However, it should be noted that, perhaps due to a loss of power from excluding 9865 individuals, only one of these four loci retained genome-wide significance (rs7829975 on chromosome 8).

Genetic correlation of mood instability with MDD, schizophrenia, BD, ADHD, anxiety disorder and PTSD

We identified strong genetic correlation between mood instability and MDD ($r_g = 0.60$, SE = 0.07, $p = 8.95 \times 10^{-17}$) and a smaller, but significant, correlation between mood instability and both schizophrenia ($r_g = 0.11$, SE = 0.04, $p = 0.01$) and anxiety disorders ($r_g = 0.28$, SE = 0.14, $p = 0.04$; Table 3). We did not find significant genetic overlap between mood instability and BD ($r_g = 0.01$, SE = 0.05, $p = 0.27$), ADHD ($r_g = 0.14$, SE = 0.11, $p = 0.18$) or PTSD ($r_g = 0.33$, SE = 0.17, $p = 0.06$).

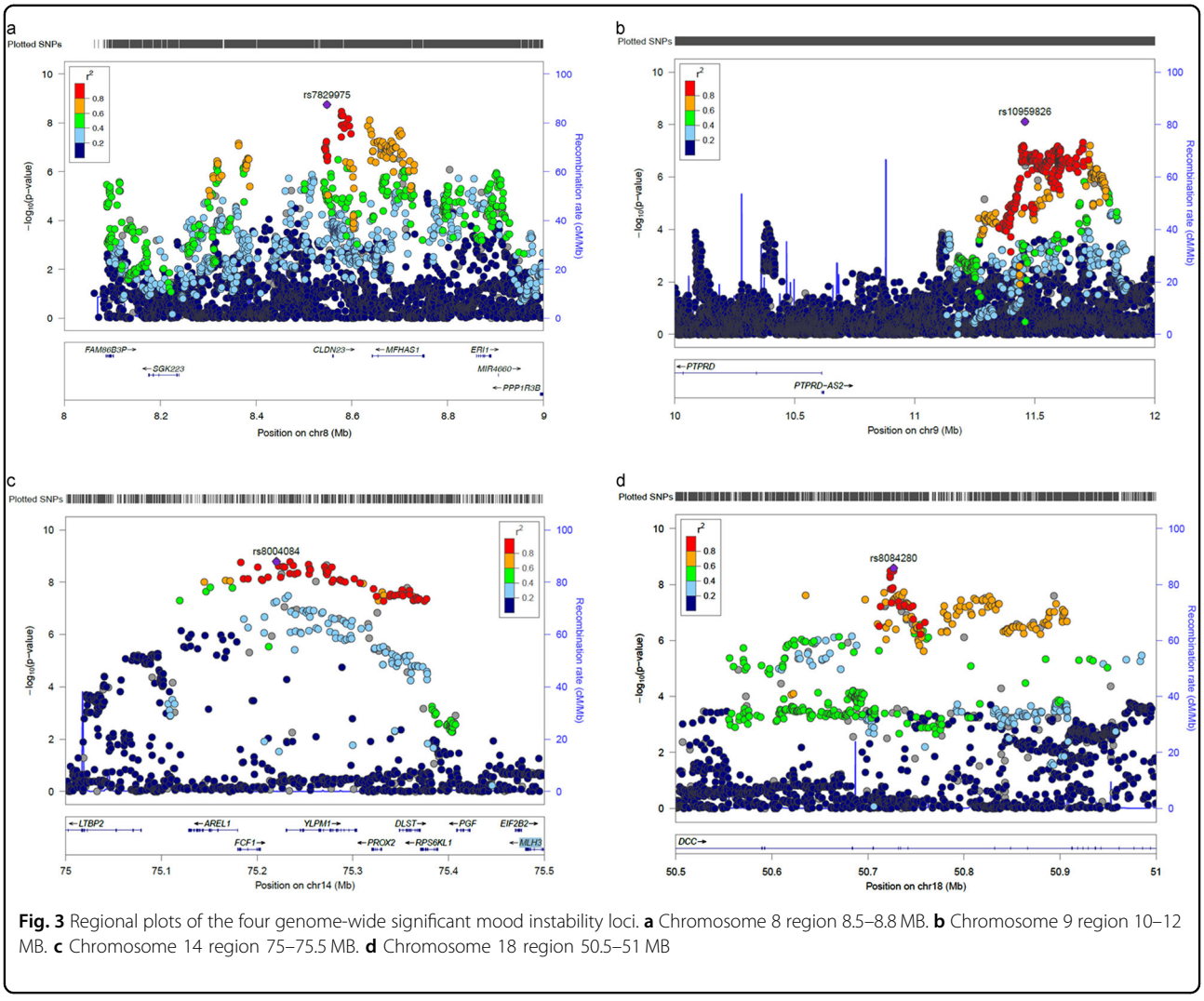


Fig. 3 Regional plots of the four genome-wide significant mood instability loci. **a** Chromosome 8 region 8.5–8.8 MB. **b** Chromosome 9 region 10–12 MB. **c** Chromosome 14 region 75–75.5 MB. **d** Chromosome 18 region 50.5–51 MB

Table 3 Genetic correlation between mood instability and MDD, schizophrenia, BD, PTSD, ADHD and anxiety disorder

Phenotype	Rg	se	z	p	h2 obs	h2 obs se	h2 int	h2 int se	Gcov int	Gcov int se
MDD	0.6	0.07	8.32	8.95×10^{-17}	0.11	0.01	0.99	0.008	−0.0019	0.006
Schizophrenia	0.11	0.04	2.48	0.01	0.25	0.01	1.03	0.01	0.0008	0.007
BD	0.01	0.05	0.27	0.27	0.12	0.01	1.02	0.008	0.0069	0.005
PTSD	0.33	0.17	1.9	0.06	0.10	0.004	0.99	0.007	0.0004	0.005
ADHD	0.14	0.11	1.35	0.18	0.4	0.15	1.01	0.01	0.0046	0.004
Anxiety disorder	0.28	0.14	2.04	0.04	0.06	0.03	1.00	0.01	0.01	0.005

R_g genetic correlation with mood instability, *SE* standard error of the genetic correlation, *Z* the test statistic, h^2 *obs* heritability on the observed scale, h^2 *obs SE* the standard error of the heritability, h^2 *int* intercept of the heritability, h^2 *int SE* standard error of the heritability intercept, *Gcov int* intercept of the genetic covariance, *Gcov int SE* standard error of the genetic covariance intercept, *MDD* major depressive disorder, *BD* bipolar disorder, *PTSD* post-traumatic stress disorder, *ADHD* attention deficit hyperactivity disorder

PRS analysis of MDD, BD and schizophrenia as predictors of mood instability

Using the PRS approach, both MDD and schizophrenia had significant positive correlations with mood instability status (for MDD at $p < 0.5$ PRS threshold: OR = 1.029, 95% CI = 1.02–1.033, $r^2 = 0.023$, $p = 1.00 \times 10^{-34}$ and for schizophrenia at $p < 0.1$ PRS threshold: OR = 1.009, 95% CI = 1.005–1.014, $r^2 = 0.021$, $p = 6.71 \times 10^{-5}$; Supplementary Table S2). There was no evidence of an association between PRS for BD and mood instability. This finding of a positive correlation between PRSs for MDD and schizophrenia and mood instability status (and no such correlation for BD PRS) was consistent across additional analyses stratified for sex and age (Supplementary Tables S3–S6).

Discussion

We have identified four independent loci associated with mood instability within a large population cohort, in what is to date the only GWAS of this phenotype. We also identified a SNP-based heritability estimate for mood instability of ~8%, and a strong genetic correlation between mood instability and MDD, suggesting substantial genetic overlap between mood instability and vulnerability to MDD. There was also a small but significant genetic correlation between mood instability and schizophrenia and between mood instability and anxiety disorders, but no significant genetic correlation with BD, ADHD or PTSD. PRS analyses found a positive correlation between genes for both MDD and schizophrenia and mood instability status, but this was not the case for BD.

The strong genetic correlation between mood instability and MDD is of interest because it is consistent with the hypothesis that at least part of the pathophysiology of MDD might include a reduced capacity to effectively regulate affective states. In support of this is evidence that individuals with MDD tend to have maladaptive responses to intense emotions, responding with worry, rumination and self-criticism, which can then exacerbate negative emotional states[27]. This maladaptive pattern of responses is also consistent with our finding of a small but significant genetic correlation between mood instability and both anxiety disorder and schizophrenia.

The lack of genetic correlation between mood instability and BD was unexpected, given that mood instability is considered a core deficit in BD[4] and was more common in our BD cases than MDD cases. Similarly, a genetic correlation between mood instability, ADHD and PTSD might have been anticipated. This lack of correlation between mood instability and BD/ADHD/PTSD is difficult to account for, but might be explained by the relatively underpowered nature of the BD, ADHD and PTSD GWAS analyses, compared to the analyses used for MDD and schizophrenia. It is worth noting that, although not significant, the magnitude of the genetic correlation between mood instability and ADHD was 0.14. Similarly, the genetic correlation between mood instability and PTSD was not significant but had a magnitude of 0.33.

It is well documented that MDD occurs more commonly in females than in males, and it is possible that mood instability may be of greater relevance as a cross-cutting phenotype for women compared to men. We therefore carried out a GWAS of mood instability for males and females separately (Supplementary Fig. S1 and Fig. S2). These stratified analyses found no genome-wide significant loci for females and only one genome-wide significant locus for males (the previously identified locus on chromosome 18). Furthermore, there was perfect genetic correlation between mood instability in males and females. Although these analyses had reduced power, they suggest that there was no evidence for a large number of sex-specific loci for mood instability. Similarly, we carried out GWAS stratified by age, for those in the sample at or below the median age of 58 and for those above age 58 (Supplementary Figs. S3 and S4). As with stratification by sex, these age-stratified analyses did not identify any genome-wide significant loci, and there was perfect correlation between mood instability in the younger and older subgroups.

It is not possible to be certain which of the genes within associated loci are likely to be most relevant to the pathophysiology of mood instability but several genes of interest were identified. For example, the lead SNP within the associated region on chromosome 18 lies in intron 9 of the *DCC netrin 1 receptor* (originally named *deleted in colorectal cancer; DCC)* gene, with no other protein-coding genes for >500 kb on either side (Fig. 3d). *DCC* is the receptor for the guidance cue *netrin 1*, which has a central role in the development of the nervous system, including (but not limited to) the organisation and function of mesocorticolimbic dopamine systems[28]. Recent studies have shown a range of human phenotypes associated with loss-of-function mutations in *DCC*, including agenesis of the corpus callosum, learning disabilities and mirror movements, all associated with a large-scale disruption of the development of commissural connectivity and lateralisation[29, 30]. Manitt et al. have identified that *DCC* has a role in regulating the connectivity of the medial prefrontal cortex during adolescence and found that *DCC* expression was elevated in the brain tissue of antidepressant-free subjects who committed suicide[31]. This suggests a possible role for *DCC* variants in increasing predisposition to mood instability and mood disorders, as well as related psychopathological phenotypes.

The associated region on chromosome 14 contains at least 10 candidate genes (Table 2 and Fig. 3c). One of these is *translation initiation factor 2B subunit beta*

(EIF2B2), mutations in which are known to cause a range of clinically heterogeneous leukodystrophies[32]. Reduced white matter integrity has been consistently associated with negative emotionality traits (such as harm avoidance, neuroticism and trait anxiety)[33], as well as with MDD and BD[34]. It is therefore possible that variation in *EIF2B2* may have a role in mood instability.

Another gene within the associated region on chromosome 14 is *placental growth factor (PGF)*, a member of the angiogenic vascular endothelial growth factor *(VEGF)* family[35, 36], which is expressed at high levels in the placenta and thyroid[37]. *PGF* has a wide range of functions, including embryonic thyroid development[38] and immune system function[39, 40], as well as a role in atherosclerosis, angiogenesis in cancer, cutaneous delayed-type hypersensitivity, obesity, rheumatoid arthritis and pre-eclampsia[39, 41–44]. *PGF* may be of interest because of the long-established association between thyroid dysfunction and both MDD and BD[45], along with the recent observation that pre-eclampsia may be a marker for the subsequent development of mood disorders[46].

Also of interest is the finding that the gene for *protein tyrosine phosphatase, receptor type D (PTPRD)* lies within 1 Mb of the associated region on chromosome 9 (Fig. 3b). *PTPRD* encodes a receptor type protein tyrosine phosphatase known to be expressed in the brain and with an organising role at a variety of synapses, including those that play a role in synaptic plasticity[47]. As such, it may have a role in a broad range of psychopathology.

Two of the genomic loci associated with mood instability (on chromosomes eight and nine) overlap with loci found to be associated with neuroticism in a recent GWAS and meta-analysis, which combined data from the UK Biobank cohort, the Generation Scotland cohort and a cohort from the Queensland Institute of Medical Research[48]. The neuroticism study made use of scores on the 12-item EPQ-R-S questionnaire, of which one of the questions was the mood instability question used in the present study. This overlap in findings suggests that mood instability is a key component of neuroticism as defined by the EPQ-R-S and that at least some of the gene variants implicated in mood instability are likely to contribute to the broader phenotype of neuroticism. We did not assess for genetic correlation between mood instability and neuroticism using LDSR because both GWAS outputs were predominantly from the same UK Biobank sample.

Strengths and limitations

To the best of our knowledge, this is the first reported GWAS of mood instability. It has enabled objective estimates of heritability and genetic correlation with important psychiatric disorders to be made for the first time. In the future, genotyping data for the full UK Biobank sample (502,000 participants) will be available. This increased sample size may identify larger estimates of shared variance between mood instability and psychiatric disorders.

Some important limitations of this work are acknowledged. The mood instability phenotype used was based on response to a single-item question (*'Does your mood often goes up and down?'*), which may be an imperfect measure of mood instability. Approximately 44% of the whole UK Biobank cohort answered 'yes' to this question, a much larger proportion than the 13% of participants classified as having mood instability within the UK APMS[2]. This may be because the assessment of mood instability in the APMS was based on a slightly different question *('Do you have a lot of sudden mood changes'*) and because respondents had to additionally report that they *'suffered this symptom over the last several years'*. Clearly, a potential limitation of self-report is the possibility of responder bias and, further, a more complete and objectively assessed measure of mood instability would have been preferable. However, this was not available to us in the UK Biobank phenotype data set and is unlikely to be feasible to collect within a population cohort of this size.

Conclusions

Despite a recognition that mood instability is likely to be an important phenotype underpinning a range of psychiatric disorders—particularly mood disorders[4]—there has to date been very little work on its neural correlates. Early investigations tentatively suggest a role for altered function and/or connectivity of the amygdala[49], but this is an area that is currently underdeveloped. It is hoped that our findings will stimulate new research on mood instability, which may be a clinically useful and biologically valid trait that cuts across traditional diagnostic categories[50].

Acknowledgements

This research was conducted using the UK Biobank resource. UK Biobank was established by the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government and Northwest Regional Development Agency. UK Biobank has also had funding from the Welsh Assembly Government and the British Heart Foundation. Data collection was funded by UK Biobank. J.W. is supported by the JMAS Sim Fellowship for depression research from the Royal College of Physicians of Edinburgh (173558). D.J.S. is supported by an Independent Investigator Award from the Brain and Behaviour Research Foundation (21930) and a Lister Prize Fellowship (173096). A.F. is supported by an MRC Doctoral Training Programme Studentship at the University of Glasgow (MR/K501335/1). The work at Cardiff University was funded by Medical Research Council (MRC) Centre (G0800509) and Programme Grants (G0801418). The funders had no role in the design or analysis of this study, decision to publish or preparation of the manuscript.

Author details

[1]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK. [2]Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden. [3]School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. [4]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK

Competing interests

The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information
The online version of this article (doi:10.1038/s41398-017-0012-7) contains supplementary material.

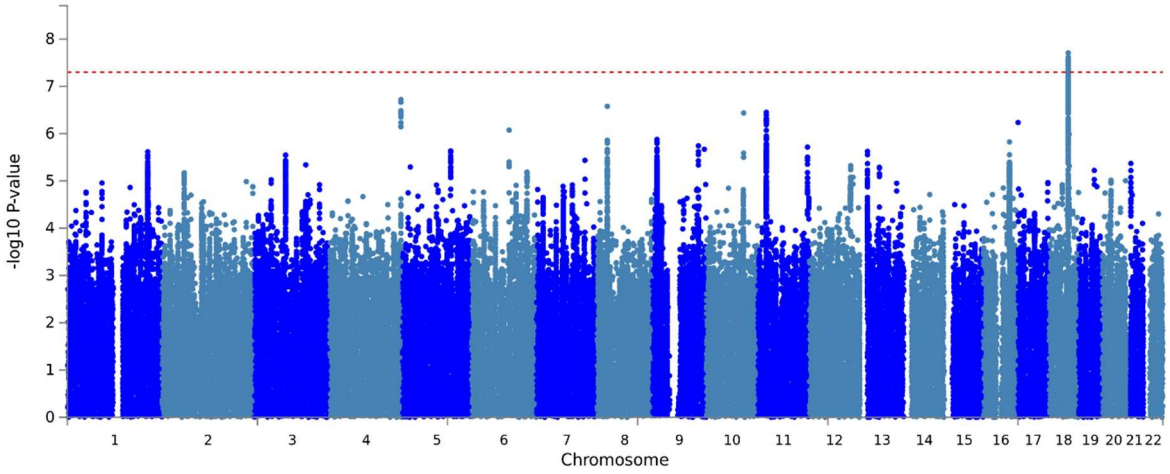
Received: 7 April 2017 Revised: 9 June 2017 Accepted: 9 August 2017
Published online: 30 November 2017

References

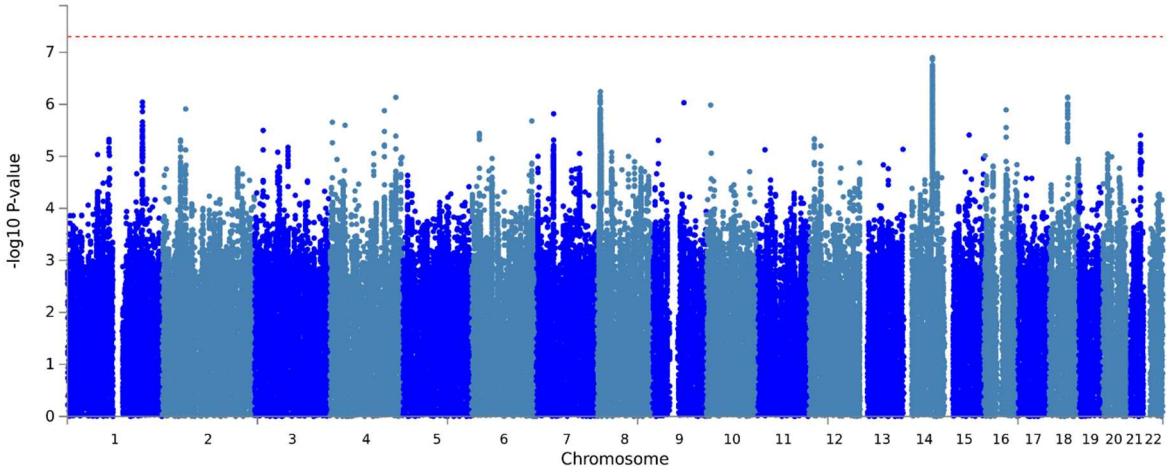
1. Balbuena, L., Bowen, R., Baetz, M. & Marwaha, S. Mood instability and irritability as core symptoms of major depression: an exploration using rasch analysis. *Front. Psychiatry* **7**, 174 (2016).
2. Marwaha, S., Parsons, N., Flanagan, S. & Broome, M. The prevalence and clinical associations of mood instability in adults living in England: results from the adult psychiatric morbidity survey 2007. *Psychiatr. Res.* **205**, 262–268 (2013).
3. Cuthbert, B. & Insel, T. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* **11**, 126 (2013).
4. Broome, M. R., Saunders, K. E. A., Harrison, P. J. & Marwaha, S. Mood instability: significance, definition and measurement. *Br. J. Psychiatry* **207**, 283–285 (2015).
5. Marwaha, S. et al. Affective instability, childhood trauma and major affective disorders. *J. Affect. Disord.* **190**, 764–771 (2016).
6. Howes, O. D. et al. A comprehensive review and model of putative prodromal features of bipolar affective disorder. *Psychol. Med.* **41**, 1567–1577 (2011).
7. Strejilevich, S. A. et al. Mood instability and functional recovery in bipolar disorders. *Acta Psychiatr. Scand.* **128**, 194–202 (2013).
8. Coccaro, E. F., Ong, A. D., Seroczynski, A. D. & Bergeman, C. S. Affective intensity and lability: heritability in adult male twins. *J. Affect. Disord.* **136**, 1011–1016 (2012).
9. Marwaha, S. et al. How is affective instability defined and measured? A systematic review. *Psychol. Med.* **44**, 1793–1808 (2014).
10. Sudlow, C. 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.* **12**, e1001779 (2015).
11. Eysenck, S. B. G., Eysenck, H. J. & Barrett, P. A revised version of the psychoticism scale. *Pers. Individ. Dif.* **6**, 21–29 (1985).
12. UK Biobank. Genotype imputation and genetic association studies of UK Biobank, Interim Data Release, 11 September 2015; http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015.pdf.
13. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Method* **10**, 5–6 (2013).
14. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457 (2011).
15. Huang, J. et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6**, 8111 (2015).
16. UK Biobank. Genotyping of 500,000 UK Biobank participants. Description of sample processing workflow and preparation of DNA for genotyping, 11 September 2015; https://biobank.ctsu.ox.ac.uk/crystal/docs/genotyping_sample_workflow.pdf.
17. Purcell S. et al. PLINK: A Tool Set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
18. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
19. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
20. Bipolar Disorder Working Group of the Psychiatric Genomics Consortium. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43**, 977–983 (2011).
21. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
22. Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
23. Neale, B. M. et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **49**, 884–897 (2010).
24. Otowa, T. et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* **21**, 1391–1399 (2016).
25. Duncan, L. E. et al. Largest GWAS of PTSD (N = 20[thinsp]070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol. Psychiatry* (2017).
26. Smith, D. J. et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK Biobank: cross-sectional study of 172,751 participants. *PLoS ONE* **8**, e75362 (2013).
27. Mennin, D. S. & Fresco, D. M. What, me worry and ruminate about DSM-5 and RDoC? The importance of targeting negative self-referential processing. *Clin. Psychol.* **20**, 258–267 (2013).
28. Manitt, C. et al. The Netrin receptor DCC is required in the pubertal organization of mesocortical dopamine circuitry. *J. Neurosci.* **31**, 8381–8394 (2011).
29. Jamuar, S. S. et al. Biallelic mutations in human DCC cause developmental split-brain syndrome. *Nat. Genet.* **49**, 606–612 (2017).
30. Marsh, A. P. L. et al. Mutations in DCC cause isolated agenesis of the corpus callosum with incomplete penetrance. *Nat. Genet.* **49**, 511–514 (2017).
31. Manitt, C. et al. dcc orchestrates the development of the prefrontal cortex during adolescence and is altered in psychiatric patients. *Transl. Psychiatry* **3**, e338 (2013).
32. Horzinski, L. et al. Eukaryotic initiation factor 2B (eIF2B) GEF activity as a diagnostic tool for EIF2B-related disorders. *PLoS ONE* **4**, e8318 (2009).
33. Mincic, A. M. Neuroanatomical correlates of negative emotionality-related traits: a systematic review and meta-analysis. *Neuropsychologia* **77**, 97–118 (2015).
34. Sexton, C. E., Mackay, C. E. & Ebmeier, K. P. A systematic review of diffusion tensor imaging studies in affective disorders. *Biol. Psychiatry* **66**, 814–823 (2009).
35. De Falco, S. The discovery of placenta growth factor and its biological activity. *Exp. Mol. Med.* **44**, 1–9 (2012).
36. Xie, T. et al. VEGF-related polymorphisms identified by GWAS and risk for major depression. *Transl. Psychiatry* **7**, e1055 (2017).
37. Viglietto, G. et al. Upregulation of vascular endothelial growth factor (VEGF) and downregulation of placenta growth factor (PlGF) associated with malignancy in human thyroid tumors and cell lines. *Oncogene* **11**, 1569–1579 (1995).
38. Korevaar, T. I. et al. Soluble Flt1 and placental growth factor are novel determinants of newborn thyroid (dys)function: the generation R study. *J. Clin. Endocrinol. Metab.* **99**, E1627–E1634 (2014).
39. Oura, H. et al. A critical role of placental growth factor in the induction of inflammation and edema formation. *Blood* **101**, 560–567 (2003).
40. Luttun, A. et al. Revascularization of ischemic tissues by PlGF treatment, and inhibition of tumor angiogenesis, arthritis and atherosclerosis by anti-Flt1. *Nat. Med.* **8**, 831–840 (2002).
41. Carmeliet, P. et al. Synergism between vascular endothelial growth factor and placental growth factor contributes to angiogenesis and plasma extravasation in pathological conditions. *Nat. Med.* **7**, 575–583 (2001).
42. Yoo, S. A. et al. Role of placenta growth factor and its receptor flt-1 in rheumatoid inflammation: a link between angiogenesis and inflammation. *Arthritis Rheum.* **60**, 345–354 (2009).
43. Lijnen, H. R. et al. Impaired adipose tissue development in mice with inactivation of placental growth factor function. *Diabetes* **55**, 2698–2704 (2006).
44. Chappell, L. C. et al. Diagnostic accuracy of placental growth factor in women with suspected preeclampsia: a prospective multicenter study. *Circulation* **128**, 2121–2131 (2013).
45. Bauer, M., Goetz, T., Glenn, T. & Whybrow, P. C. The thyroid-brain interaction in thyroid disorders and mood disorders. *J. Neuroendocrinol.* **20**, 1101–1114 (2008).
46. Bergink, V. et al. Pre-eclampsia and first-onset postpartum psychiatric episodes: a Danish population-based cohort study. *Psychol. Med.* **45**, 3481–3489 (2015).
47. Takahashi, H. C. A. Protein tyrosine phosphatases PTPδ, PTPσ, and LAR: presynaptic hubs for synapse organization. *Trends Neurosci.* **36**, 522–534 (2013).
48. Smith, D. J. et al. Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. *Mol. Psychiatry* **21**, 749–757 (2016).
49. Broome, M. R., He, Z., Iftikhar, M., Eyden, J. & Marwaha, S. Neurobiological and behavioural studies of affective instability in clinical populations: a systematic review. *Neurosci. Biobehav. Rev.* **51**, 243–254 (2015).
50. Insel, T. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *Am. J. Psychiatry* **171**, 395–397 (2014).

4.2.1 Appendix B

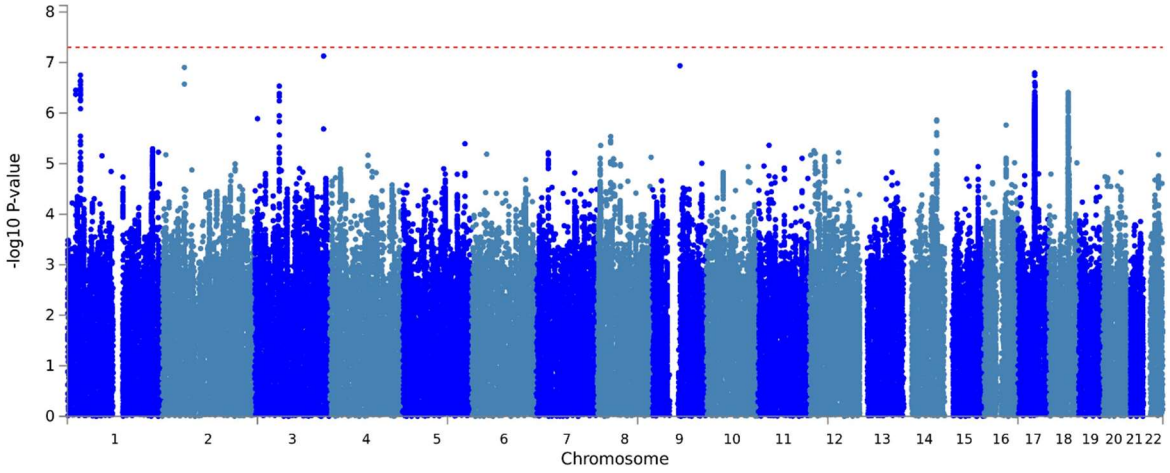
4.2.1.1 Figure S1. Manhattan plot of GWAS of mood instability in UK Biobank (males only).



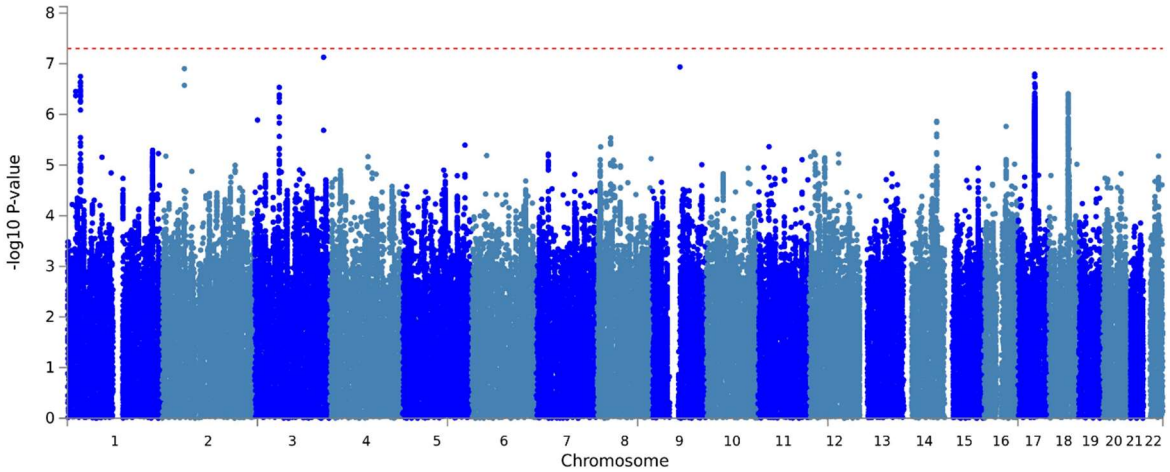
4.2.1.2 Figure S2. Manhattan plot of GWAS of mood instability in UK Biobank (females only).



4.2.1.3 Figure S3. Manhattan plot of GWAS of mood instability in UK Biobank (age 58 and below).



4.2.1.4 Figure S4. Manhattan plot of GWAS of mood instability in UK Biobank (age 59 and above).



4.2.1.5 Table S1. Genome-wide significant loci associated with mood instability in UK Biobank (excluding 9,865 participants with psychiatric disorder)

Index SNP	Chr	Position	Risk Allele/Other Allele	RAF	Beta (SE)	P-value
rs7829975	8	8,548,117	A/T	0.52	0.052 (0.009)	5.32×10^{-9}
rs10959826	9	11,459,410	G/A	0.79	0.055 (0.01)	4.77×10^{-7}
rs397852991	14	75,268,920	C/CA	0.67	0.045 (0.009)	1.25×10^{-6}
rs8084280	18	50,726,749	T/A	0.51	0.047 (0.008)	1.35×10^{-7}

4.2.1.6 Table S2. Psychiatric polygenic risk score analysis of mood instability

Adjusted for age, sex, genotyping chip and PGCs 1-8; n_{total} = 104,103, n_{cas} = 43,660, n_{con} = 60,443

Predictor	P	Beta	SE	OR	conf lower	conf upper	Nagelkerke r^2
MDD_0.01	$1.88*10^{-11}$	0.0148	0.00221	1.0149	1.011	1.0193	0.0216
MDD_0.05	$5.47*10^{-22}$	0.0216	0.00224	1.0218	1.02	1.0263	0.0221
MDD_0.1	$1.11*10^{-26}$	0.0242	0.00226	1.0244	1.02	1.029	0.0224
MDD_0.5	$1.00*10^{-34}$	0.0281	0.00229	1.0285	1.02	1.0332	0.0229
bipolar_gws	$3.64*10^{-01}$	0.00201	0.00221	1.002	0.99768	1.0064	0.021
bipolar_0.01	$1.67*10^{-01}$	0.00305	0.00221	1.0031	0.99873	1.0074	0.021
bipolar_0.05	$1.28*10^{-01}$	0.00338	0.00222	1.0034	0.99903	1.0078	0.021
bipolar_0.1	$1.65*10^{-01}$	0.00311	0.00224	1.0031	0.99872	1.0075	0.021
bipolar_0.5	$1.00*10^{-01}$	0.0037	0.00225	1.0037	0.99929	1.0081	0.021
SCZ_gws	$1.08*10^{-01}$	0.0035	0.00218	1.0035	0.99924	1.0078	0.021
SCZ_0.01	$1.55*10^{-03}$	0.00718	0.00227	1.0072	1.0027	1.0117	0.0211
SCZ_0.05	$1.19*10^{-04}$	0.00885	0.0023	1.0089	1.0044	1.0134	0.0211
SCZ_0.1	$6.71*10^{-05}$	0.0092	0.00231	1.0092	1.0047	1.0138	0.0212
SCZ_0.5	$1.24*10^{-04}$	0.00893	0.00233	1.009	1.0044	1.0136	0.0212

4.2.1.7 Table S3. Psychiatric polygenic risk score analysis of mood instability in females

Adjusted for age, genotyping chip and PGCs 1-8; n_{total} = 53,279, n_{cas} = 23,308, n_{con} = 29,971

Predictor	P	Beta	SE	OR	conf lower	conf upper	Nagelkerke R2
MDD_0.01	$8.80*10^{-06}$	0.0137	0.00308	1.0138	1.0077	1.0199	0.024668
MDD_0.05	$1.17*10^{-10}$	0.0201	0.00312	1.0203	1.0141	1.0266	0.025205
MDD_0.1	$2.96*10^{-12}$	0.022	0.00315	1.0222	1.0159	1.0286	0.025383
MDD_0.5	$2.89*10^{-17}$	0.027	0.0032	1.0274	1.021	1.0338	0.025945
bipolar_gws	$6.20*10^{-01}$	0.00153	0.00309	1.0015	0.99549	1.0076	0.024186
bipolar_0.01	$5.35*10^{-02}$	0.00593	0.00307	1.0059	0.99991	1.012	0.024272
bipolar_0.05	$1.01*10^{-01}$	0.00509	0.0031	1.0051	0.99901	1.0112	0.024246
bipolar_0.1	$1.27*10^{-01}$	0.00475	0.00312	1.0048	0.99864	1.0109	0.024237
bipolar_0.5	$6.92*10^{-02}$	0.00569	0.00313	1.0057	0.99955	1.0119	0.024261
SCZ_gws	$3.43*10^{-01}$	0.00287	0.00303	1.0029	0.99694	1.0088	0.024202
SCZ_0.01	$1.62*10^{-02}$	0.00759	0.00316	1.0076	1.0014	1.0139	0.024323
SCZ_0.05	$7.32*10^{-03}$	0.00857	0.0032	1.0086	1.0023	1.015	0.024357
SCZ_0.1	$5.90*10^{-03}$	0.00885	0.00321	1.0089	1.0026	1.0153	0.024367
SCZ_0.5	$1.47*10^{-02}$	0.00791	0.00324	1.0079	1.0016	1.0144	0.024327

4.2.1.8 Table S4. Psychiatric polygenic risk score analysis of mood instability in males only

Adjusted for age, genotyping chip and PGCs 1-8, n_{total} = 50,824, n_{cas} = 24,804, n_{con} = 27,939

Predictor	P	Beta	SE	OR	conf lower	conf upper	Nagelkerke R2
MDD_0.01	$4.42*10^{-07}$	0.016	0.00317	1.0162	1.0099	1.0225	0.01567
MDD_0.05	$4.99*10^{-13}$	0.023	0.00322	1.0235	1.0171	1.03	0.016372
MDD_0.1	$2.69*10^{-16}$	0.027	0.00324	1.0269	1.0204	1.0334	0.016762
MDD_0.5	$3.01*10^{-19}$	0.03	0.00329	1.0299	1.0233	1.0366	0.017115
bipolar_gws	$4.30*10^{-01}$	0.003	0.00317	1.0025	0.99629	1.0088	0.015015
bipolar_0.01	$9.96*10^{-01}$	-0.00001	0.00317	0.99999	0.99379	1.0062	0.014999
bipolar_0.05	$6.48*10^{-01}$	0.001	0.00319	1.0015	0.99521	1.0077	0.015004
bipolar_0.1	$6.98*10^{-01}$	0.001	0.00321	1.0012	0.99496	1.0076	0.015003
bipolar_0.5	$6.45*10^{-01}$	0.001	0.00323	1.0015	0.99517	1.0079	0.015004
SCZ_gws	$1.82*10^{-01}$	0.004	0.00314	1.0042	0.99805	1.0104	0.015046
SCZ_0.01	$4.17*10^{-02}$	0.007	0.00326	1.0067	1.0003	1.0131	0.015108
SCZ_0.05	$6.18*10^{-03}$	0.009	0.00331	1.0091	1.0026	1.0157	0.015196
SCZ_0.1	$4.26*10^{-03}$	0.01	0.00332	1.0095	1.003	1.0161	0.015214
SCZ_0.5	$2.80*10^{-03}$	0.01	0.00335	1.0101	1.0035	1.0167	0.015234

4.2.1.9 Table S5. Psychiatric polygenic risk score analysis of mood instability in those equal to or younger than the median age of 58

Adjusted for age, sex genotyping chip and PGCs 1-8; n_{total} = 52,743, n_{cas} = 24,804, n_{con} = 27,939

Predictor	P	Beta	SE	OR	conf lower	conf upper	Nagelkerke R2
MDD_0.01	$4.42*10^{-07}$	0.016	0.00317	1.0162	1.0099	1.0225	0.01567
MDD_0.05	$4.99*10^{-13}$	0.023	0.00322	1.0235	1.0171	1.03	0.016372
MDD_0.1	$2.69*10^{-16}$	0.027	0.00324	1.0269	1.0204	1.0334	0.016762
MDD_0.5	$3.01*10^{-19}$	0.03	0.00329	1.0299	1.0233	1.0366	0.017115
bipolar_gws	$4.30*10^{-01}$	0.003	0.00317	1.0025	0.99629	1.0088	0.015015
bipolar_0.01	$9.96*10^{-01}$	-0.00001	0.00317	0.99999	0.99379	1.0062	0.014999
bipolar_0.05	$6.48*10^{-01}$	0.001	0.00319	1.0015	0.99521	1.0077	0.015004
bipolar_0.1	$6.98*10^{-01}$	0.001	0.00321	1.0012	0.99496	1.0076	0.015003
bipolar_0.5	$6.45*10^{-01}$	0.001	0.00323	1.0015	0.99517	1.0079	0.015004
SCZ_gws	$1.82*10^{-01}$	0.004	0.00314	1.0042	0.99805	1.0104	0.015046
SCZ_0.01	$4.17*10^{-02}$	0.007	0.00326	1.0067	1.0003	1.0131	0.015108
SCZ_0.05	$6.18*10^{-03}$	0.009	0.00331	1.0091	1.0026	1.0157	0.015196
SCZ_0.1	$4.26*10^{-03}$	0.01	0.00332	1.0095	1.003	1.0161	0.015214
SCZ_0.5	$2.80*10^{-03}$	0.01	0.00335	1.0101	1.0035	1.0167	0.015234

4.2.1.10 able S6. Psychiatric polygenic risk score analysis of mood instability in those older than the median age of 58

Adjusted for age, sex genotyping chip and PGCs 1-8, n_{total} = 51,360, n_{cas} = 18,856, n_{con} = 23,504

Predictor	P	Beta	SE	OR	conf lower	conf upper	Nagelkerke R2
MDD_0.01	$2.60*10^{-06}$	0.0151	0.0032	1.0152	1.0088	1.0216	0.0028399
MDD_0.05	$7.71*10^{-11}$	0.0211	0.00325	1.0214	1.0149	1.0279	0.0033778
MDD_0.1	$1.41*10^{-12}$	0.0232	0.00327	1.0234	1.0169	1.03	0.0035861
MDD_0.5	$3.58*10^{-14}$	0.0251	0.00331	1.0254	1.0188	1.0321	0.0037779
bipolar_gws	$2.54*10^{-01}$	0.00365	0.0032	1.0037	0.99738	1.01	0.0022874
bipolar_0.01	$4.03*10^{-01}$	0.00267	0.0032	1.0027	0.99641	1.009	0.0022715
bipolar_0.05	$4.95*10^{-01}$	0.0022	0.00322	1.0022	0.9959	1.0085	0.0022653
bipolar_0.1	$3.70*10^{-01}$	0.00291	0.00324	1.0029	0.99655	1.0093	0.0022742
bipolar_0.5	$3.55*10^{-01}$	0.00301	0.00326	1.003	0.99664	1.0094	0.0022756
SCZ_gws	$9.12*10^{-01}$	-0.000347	0.00314	0.99965	0.99351	1.0058	0.0022532
SCZ_0.01	$7.72*10^{-02}$	0.00581	0.00329	1.0058	0.99937	1.0123	0.0023359
SCZ_0.05	$5.71*10^{-02}$	0.00633	0.00333	1.0063	0.99981	1.0129	0.0023491
SCZ_0.1	$3.71*10^{-02}$	0.00696	0.00334	1.007	1.0004	1.0136	0.0023684
SCZ_0.5	4.85E-02	0.00664	0.00337	1.0067	1	1.0133	0.0023563

4.3 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.

Molecular Psychiatry
https://doi.org/10.1038/s41380-019-0439-8

ARTICLE

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[1] · Elizabeth M. Tunbridge[2,3] · Cynthia Sandor[4] · Laura M. Lyall[1] · Amy Ferguson[1] · Rona J. Strawbridge[1,5] · Donald M. Lyall[1] · Breda Cullen[1] · Nicholas Graham[1] · Keira J. A. Johnston[1] · Caleb Webber[4,6] · Valentina Escott-Price[7] · Michael O'Donovan[7] · Jill P. Pell[1] · Mark E. S. Bailey[8] · Paul J. Harrison[2,3] · Daniel J. Smith[1]

Received: 8 February 2019 / Revised: 3 April 2019 / Accepted: 29 April 2019
© The Author(s) 2019. This article is published with open access

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 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 mood instability were identified with a SNP heritability estimate of 9%. Linkage Disequilibrium Score Regression (LDSR) analyses identified genetic correlations with Major Depressive Disorder (MDD), Bipolar Disorder (BD), Schizophrenia, anxiety, and Post Traumatic Stress Disorder (PTSD). Gene-level and gene set analyses identified 244 significant genes and 6 enriched gene sets. Tissue expression analysis of the SNP-level data found enrichment in multiple brain regions, and eQTL analyses highlighted an inversion on chromosome 17 plus two brain-specific eQTLs. In addition, 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 mood instability. These findings may be relevant for the identification of novel trans-diagnostic drug targets and could help to inform future stratified medicine innovations in mental health.

Introduction

Mood instability 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, mood instability may

Supplementary information The online version of this article (https://doi.org/10.1038/s41380-019-0439-8) contains supplementary material, which is available to authorised users.

✉ Daniel J. Smith
Daniel.Smith@glasgow.ac.uk

[1] Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK

[2] Department of Psychiatry, University of Oxford, Oxford, UK

[3] Oxford Health NHS Foundation Trust, Oxford, UK

[4] UK Dementia Research Institute, Cardiff University, Cardiff, UK

[5] Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden

[6] Department of Physiology, Anatomy and Genetics, Oxford, UK

[7] University of Cardiff, Cardiff, UK

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

Published online: 05 June 2019

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 mood instability 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 mood instability in the European ancestry subset of UK Biobank dataset ($n = 363{,}705$), using a BOLT-LMM approach to maximise 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 contextualise 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 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 mood instability cases and those who responded 'no' were defined as controls. To minimise any impact of psychiatric disorders on observed genetic associations with mood instability, 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; UK 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 the subset classified as British participants of European ancestry.

Genetic association and heritability

Genetic association analysis was performed using BOLT-LMM [7, 8]. This approach makes use of a genetic relationship matrix (GRM) to control as robustly as possible for population structure without the need to adjust the model for principal components (PCs), while maximising power by avoiding the need to exclude related individuals. In addition, BOLT-LMM builds an infinitesimal model including all directly genotyped SNPs simultaneously, thereby further increasing power compared with 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-off criterion to be substantially reduced and increases the number of SNPs available to test for association with the outcome. BOLT-LMM treats binary variables as a linear trait but is able to handle binary outcomes well when the sample size is large and when the number of cases and controls are evenly balanced, as is the case here.

Models were adjusted for age, sex and genotyping array. SNPs were filtered to remove those with MAF < 0.01, Hardy-Weinberg Equilibrium $p < 1 \times 10^{-6}$, or imputation quality score <0.3. BOLT-LMM was also used to provide a heritability estimate and λ_{GC} estimate. A secondary analysis was also performed on a subsample of the cohort which excluded those used in the previous GWAS and anyone who was related to another participant.

Regional plots were made via LocusZoom v1.4 [9] as FUMA lacks SNPs from the HRC reference panel which 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 5 MB 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.

The summary statistics were processed by FUMA [10] (See URLs) for visualisation, MAGMA Gene Analysis, Gene-set Analysis and Tissue Expression Analysis [11]. The Gene-level Analysis operates by grouping p values for individual SNPs into a gene test statistic using the mean χ^2 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 [12], 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.

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 LIBD 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 False Discovery Rate (FDR) corrected threshold of $q \leq 0.05$ 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 q values quoted in the text are FDR corrected.

Genetic principal component generation

Genetic principal components were created using Plink 2 [19] using pca approx (with default settings) within the region between base positions 40,850,001 and 41,850,000 on chromosome 17 for the analysis of the inversion polymorphism.

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 conducted a further analysis in which we re-weighted our generic PLN to be more sensitive to functional genomics data most informative to mood instability by considering only phenotypes within the over-arching mouse phenotype ontology (MPO) category *Nervous System* (MP:0003631). The PLN and nervous-PLN (NS-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. [21], 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 GWA and subGWA intervals were defined by considering SNPs attaining an association p-value of 5×10^{-8} and 1×10^{-6}. This identified 6375 (GWA) and 9358 (subGWA) SNPs respectively. We then identified the haplotypic block containing each of these of these SNPs using genotypes in the 1000 Genome Project and the python pipeline developed by Brent Pedersen (See URLs). We defined GWA/sub GWA intervals by identifying the most distant block on a chromosome within a region of 500 kb 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 this GWAS sample of 363,705 individuals without a history of psychiatric disorder, 43.2% reported mood instability ($n = 157{,}039$) and the rest did not ($n = 206{,}666$). There was a higher proportion of females amongst the mood instability cases than in controls (55.4% vs. 51.2%, respectively), and the mean age of cases was lower than for controls (55.8 years vs. 57.7 years).

GWAS findings

We detected 46 loci across the genome with $p < 5 \times 10^{-8}$ (Fig. 1 and Table S1) and an estimated SNP heritability (h^2) of 0.09 (SE = 0.02). The heritability estimate has increased from the previous GWAS by ~2% (Previous $h^2 = 0.07$, SE = 0.007). We attribute the increase in SE to the differing methodologies. The distribution of test statistics was consistent with a polygenic contribution to risk ($\lambda_{GC} = 1.21$; $\lambda_{1000} = 1.001$; LDSR intercept = 1.041; SE = 0.006). In addition, to help validate the four loci identified in the previous mood instability GWAS, we tested the top SNP from each locus in a sub-sample that excluded individuals in the previous smaller GWAS and those individuals related to another Biobank participant ($n = 169{,}857$). All four SNPs were associated with mood instability after Bonferroni correction ($\alpha < 0.0125$, Table S2). We also note that the directions of effect were the same as for the previous GWAS findings.

Gene-level and gene set analysis

A total of 244 significant genes were detected by MAGMA (Supplementary Table S3) and FUMA gene analysis. The Gene Set Analysis returned 6 enriched gene sets that met the threshold for significance after Bonferroni correction (Supplementary Table S4). 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 (Fig. 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 mood instability and five psychiatric phenotypes of interest (Table 1). All genetic correlations remained significant after FDR correction ($q < 0.05$). The largest correlations were with MDD ($r_g = 0.74$, $q = 8.50 \times 10^{-157}$) and anxiety ($r_g = 0.64$, $q = 8.08 \times 10^{-6}$). PTSD had a moderate correlation with mood instability ($r_g = 0.32$, $q = 1.23 \times 10^{-2}$) and both schizophrenia and bipolar disorder had weak but significant correlations (schizophrenia $r_g = 0.14$, $q = 1.60 \times 10^{-5}$, BD $r_g = 0.09$, $q = 0.0012$).

eQTL analysis

Nine of the GWAS loci showed significant eQTLs, many potentially with specific isoforms (summarised in Table S5 and presented in full in Supplementary Table S6). The strongest evidence of association with expression levels was for rs669915, an eQTL located within a region of strong linkage disequilibrium (LD) in chromosome 17q21 resulting from the existence of a 900 kb 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 $q = 1.96 \times 10^{-99}$; CMC dataset $q = 3.99 \times 10^{-65}$), 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 S5).

The chromosome 17q21 inversion polymorphism has itself been reported to affect the expression of genes in this region [23]. We therefore investigated whether rs669915

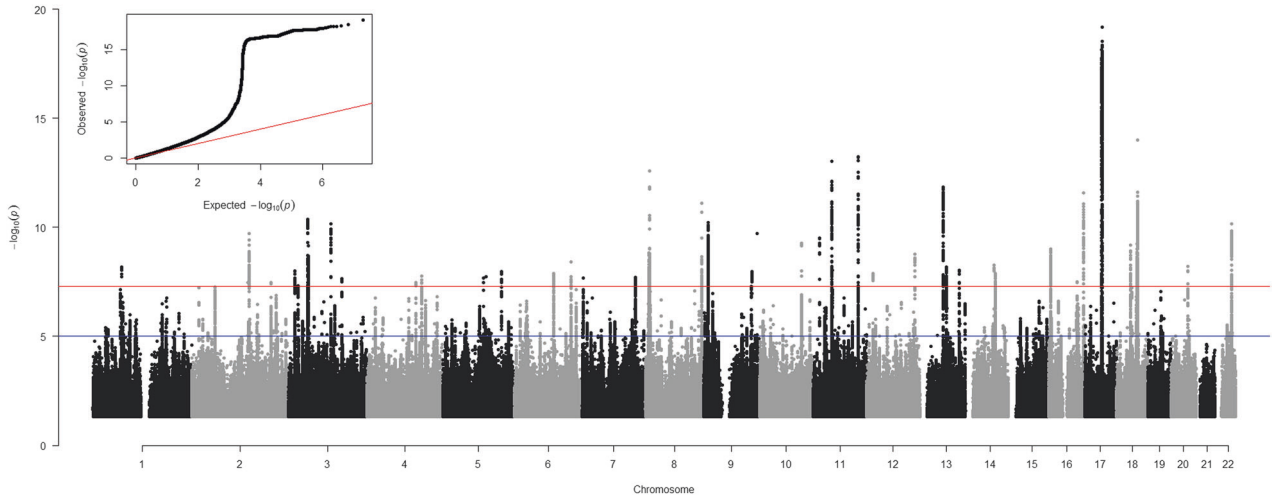


Fig. 1 Manhattan and QQ plot of mood instability GWAS

Table 1 Genetic correlations of mood instability with psychiatric phenotypes

Trait	r_g	S.E.	Z	p	q
MDD	0.74	0.03	26.7	1.70×10^{-157}	8.50×10^{-157}
Anxiety	0.64	0.14	4.7	3.23×10^{-6}	8.08×10^{-6}
PTSD	0.32	0.13	2.5	1.12×10^{-2}	1.23×10^{-2}
Schizophrenia	0.14	0.03	4.4	9.57×10^{-6}	1.60×10^{-5}
Bipolar disorder	0.09	0.037	2.5	1.23×10^{-2}	1.23×10^{-2}

r_g = genetic correlation with mood instability, S.E. = 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

might 'tag' the expression effects mediated by the inversion polymorphism in our sample. Using the method of de Jong and colleagues [23], 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 GPCs is shown in Fig. 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 mood instability phenotype and inversion genotype using a model of additive allelic effects (no. of H2 alleles) and adjusting for age, sex and genotyping array ($p = 0.835$).

Nervous system PLN analyses (NS-PLN)

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. 2a). 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 mood instability 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 (Fig. 2b).

Discussion

Main findings

These analyses represent the largest genetic study of mood instability to date. Forty six associated loci 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 five psychiatric traits assessed had a significant genetic correlation with mood instability. Some of these correlations were strong (particularly for

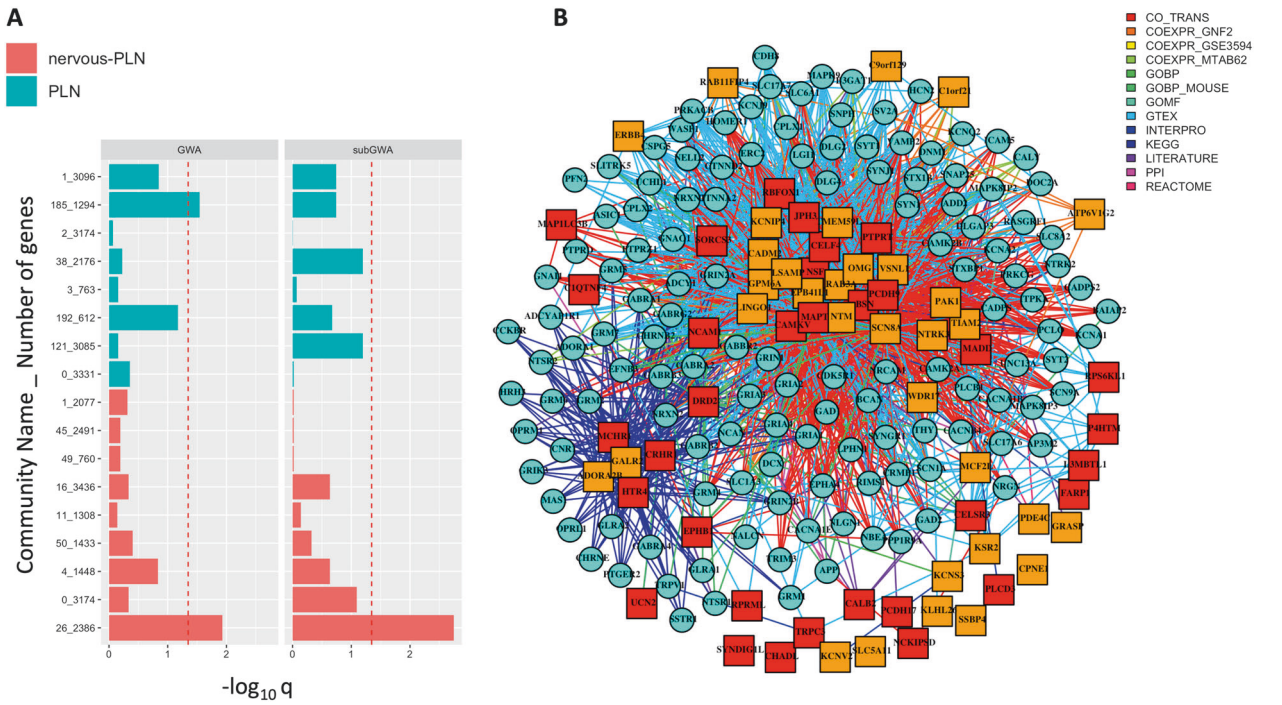


Fig. 2 Different Mood associated genetic risk variants converge in a nervous specific gene network. **a** Enrichments of gene functional communities from a generic PLN and from a Nervous-System (NS) PLN within Mood-GWA and subGWA loci (see Methods). The Community ID is given first in the descriptor followed 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

MDD and anxiety) but others were weaker than expected. For example, the genetic correlation between mood instability and BD was only 9%, perhaps suggesting that the mood instability phenotype in this study differs from the affective instability that is a core feature of BD.

Biology of mood instability

Loci associated with mood instability 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]. In addition, we identified *CALB2* which has many biological functions, including a role in modulating neuronal excitability [28]. Both *DCC* (identified in the previous mood instability 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 psychiatric disorders, it has been shown to regulate dendritic complexity [32] (reduced dendritic complexity is a recognised 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 mood instability 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*. In addition, 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 mood instability.

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 (LIBD data was generated using poly-A RNA and so targets messenger RNA). However, our findings are consistent with a recent report implicating dopamine neuron-enriched enhancer activity in this region in several dopamine-related psychiatric and neurological conditions [45].

Genes within regions associated with mood instability 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 and implicated in depression and its treatment [46]. *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 [47] and the locus showed association with schizophrenia in a Danish sample [48].

Strengths and limitations

As noted above, this is the largest GWAS of a mood instability phenotype to date and has successfully identified new loci, eQTLs, tissues, genetic correlations and gene network enrichments. However, there are several limitations, most notably the use of a single question to define mood instability, and the lack of objective verification of this phenotype. There are more detailed suggested measurement scales for mood instability, such as that developed by Chaturvedi and colleagues [49]. In the future, the use of these more comprehensive assessments in large samples may provide some clarification of our findings, for example the lack of strong genetic correlation between mood instability and BD. Nevertheless, the single question approach to mood instability has been widely used, and shown to identify robust associations with a range of health outcomes and disorders [1, 3]. Similarly, exclusions for psychiatric disorder were based on self-report.

We validated the top hits of the previous GWAS, however the cohort used was not truly independent (it was also part of the UK Biobank cohort). It would be of interest to replicate the 46 loci identified here, in sufficiently large independent replication cohorts, when they become available in the future. As well as replication of the loci, further analysis of sex specific differences would be of interest because mood instability was more common in females than males. Although this difference was relatively small, our reported analyses were adjusted for sex and these differences are similar to those reported elsewhere [50].

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 do not 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 mood instability genes will also reveal this phenotype when more completely annotated databases become available.

Finally, we 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 over representation in those who joined UK Biobank. It is however important to note that the inversion itself would be likely to contribute only a small proportion of the mood instability phenotype, such that even larger sample sizes than were used here would be needed to detect a correlation where one exists.

Conclusion

In summary, with a tripling in sample size from the previous GWAS, we identified substantially more associations with mood instability in the UK Biobank cohort [4]. Future analyses of the precise roles that these associations play in the clinical expression of mood instability will be relevant for the wide range of psychiatric phenotypes in which mood instability occurs. We have also been able to more confidently place these GWAS findings within relevant biological contexts and some of the loci and pathways identified may represent candidates for future novel drug development.

Our findings are also of interest in the context of precision medicine innovations for mental health. It is possible that polygenic risk scores derived from this work could be

applied to clinical populations to conduct pharmacogenomics studies and to inform patient stratification approaches. Overall, we hope that our findings will stimulate further research on the biology and treatment of mood instability across a range of mood and psychotic disorders.

URLs

FUMA–http://fuma.ctglab.nl/
Python pipeline developed by Brent Pedersen–https://gist.github.com/brentp/5050522
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 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). EMT and PJH are supported by the Oxford Health NIHR Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR or the Department of Health. 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). We thank all participants in the UK Biobank study. UK Biobank was established by the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government and Northwest Regional Development Agency. UK Biobank has also had funding from the Welsh Assembly Government and the British Heart Foundation. Data collection was funded by UK Biobank. The summary statistics of the GWAS are available upon request by contacting the corresponding author.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

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;44:1793–808.
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. Br J Psychiatry. 2015;207:283–5.
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. Transl Psychiatry. 2017;7:1264.
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:e1001779.
6. Biobank U. Genotyping of 500,000 UK Biobank participants. Description of sample processing workflow and preparation of DNA for genotyping. 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;47:284–90.
8. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nat Genet. 2018;50:906–8.
9. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26:2336–7
10. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.
11. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLOS Comput Biol. 2015;11:e1004219.
12. 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. Proc Natl Acad Sci. 2005;102:15545.
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;47:291–5.

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;50:668–81.
15. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. Cell 2018;173:1705–15. 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;23:666–73.
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. Mol Psychiatry. 2016;21:1391–9.
18. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5.
19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira Manuel AR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Human Genet. 2007;81:559–75.
20. Honti F, Meader S, Webber C. Unbiased functional clustering of gene variants with a phenotypic-linkage network. PLoS Comput Biol. 2014;10: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:e1005816.
22. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, et al. A common inversion under selection in Europeans. Nat Genet. 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. EMBO J. 2002;21:1004–11.
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;67:1159–64.
26. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–7.
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 Mol Genet. 2006;15:1949–62.
28. Camp AJ, Wijesinghe R. Calretinin: modulator of neuronal excitability. Int J Biochem Cell Biol. 2009;41:2118–21.
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;57:389–97.
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;195:23–29.
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;61:359–72.
32. Cheadle L, Biederer T. Activity-dependent regulation of dendritic complexity by semaphorin 3A through Farp1. J Neurosci. 2014;34:7999.
33. Broadbelt K, Byne W, Jones LB. Evidence for a decrease in basilar dendrites of pyramidal cells in schizophrenic medial prefrontal cortex. Schizophrenia Res. 2002;58:75–81.
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;102:1645–50.
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;125:545–56.
36. Wilkinson DG. Multiple roles of EPH receptors and ephrins in neural development. Nat Rev Neurosci. 2001;2:155–64.
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. Eur J Human Genet. 2016;24:556–61.
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:781–4.
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. Nat Genet. 2016;48:1031–6.
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. J Psychiat Res. 2013;47:1572–80.
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. Mol Neuropsychiatry. 2017;3:181–91.
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 Biol. 2014;15:R56–R56.
44. Su L, Ling W, Jiang J, Hu J, Fan J, Guo X, et al. Association of EPHB1rs11918092 and EFNB2 rs9520087 with psychopathological symptoms of schizophrenia in Chinese Zhuang and Han populations. Asia PacPsychiatry. 2016;8:306–8.
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. Nat Neurosci. 2018;21:1482–92.
46. Guo J, Zhang W, Zhang L, Ding H, Zhang J, Song C, et al. Probable involvement of p11 with interferon alpha induced depression. Sci Rep. 2016;6.
47. 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;99:384–91.
48. 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;22:62–9.
49. Chaturvedi M, Chander R. Development of emotional stability scale. Ind Psychiatry J. 2010;19:37.
50. Marwaha S, Parsons N, Flanagan S, Broome M. The prevalence and clinical associations of mood instability in adults living in England: Results from the Adult Psychiatric Morbidity Survey 2007. Psychiatry Res. 2013;205:262–8.

4.3.1 Appendix C

4.3.1.1 Supplementary Table S1 Genomic loci associated with mood instability.

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, S.E. = Standard Error for beta, p = p value, start_BP = start of associated region, stop_bp = end of associated region, info = imputation score, Candidate Genes = genes that may be the cause of the association.

SNP	CHR	BP	A1	A2	BETA	S.E.	P	start_BP	stop_BP	INFO	Candidate Genes
rs2000228	1	72723863	T	C	-0.007	0.0012	$6.5*10^{-9}$	72512988	73048522	1.00	NEGR1
rs377073464	2	141087768	CAATA/ATAAT	C	0.007	0.0012	$3.3*10^{-8}$	140932485	141266032	0.87	LRP1B
2:144243977	2	144243977	CA	C	0.007	0.0012	$1.9*10^{-10}$	144057444	144364876	0.94	ARHGAP15
rs12477961	2	199495674	A	G	0.006	0.0012	$3.3*10^{-8}$	199317934	199653885	0.99	PLCL1
rs9857592	3	16845144	T	A	-0.007	0.0013	$1.0*10^{-8}$	16692878	17151126	0.99	PLCL2
rs12486699	3	25125530	C	A	-0.007	0.0013	$4.8*10^{-8}$	25065129	25351513	1.00	RARB
rs73082357	3	49083566	G	A	0.012	0.0018	$4.2*10^{-11}$	46946029	52052904	1.00	QARS, LAMB2, BSN, AMIGO3, CCDC36,GPX1, RHOA,NICN1, DAG1,NDUFAF3
rs150221903	3	107295550	T	TATTG	-0.008	0.0013	$6.8*10^{-11}$	107150826	107560269	0.98	
rs4082244	3	134729537	C	G	0.007	0.0012	$2.2*10^{-8}$	134649207	134755990	0.99	EPHB1
rs45510500	4	123179900	C	T	0.014	0.0025	$3.3*10^{-8}$	122870537	123626249	1.00	
4:139091457	4	139091457	CAT	C	0.007	0.0012	$1.7*10^{-8}$	138733515	139175327	0.99	SLC7A11
rs6882578	5	103745824	G	C	-0.006	0.0011	2.1*10-8	103670314	104119000	0.99	*RP11-6N13.1*
rs6861117	5	109226417	A	G	0.012	0.0022	$1.8*10^{-8}$	108664929	109233763	0.99	
rs58746316	5	147837672	A	T	-0.007	0.0012	$1.1*10^{-8}$	147682118	147856522	0.99	HTR4
rs9490443	6	98512158	A	G	-0.010	0.0017	$1.3*10^{-8}$	98228178	98896216	0.99	*RP11-436d23.1*
rs67447472	6	143062216	G	T	-0.011	0.0019	$3.8*10^{-9}$	142677214	143083063	0.98	*HIVEP2*
7:2060397	7	2060397	CT	C	0.008	0.0014	$2.1*10^{-8}$	1812267	2260034	0.95	
rs2483509	7	133341162	G	A	-0.006	0.0012	$2.0*10^{-8}$	132821236	133803962	1.00	EXOC4
rs7818437	8	10209623	T	C	-0.010	0.0014	$2.6*10^{-13}$	8088230	11896335	0.98	
rs1962104	8	141635329	T	C	-0.008	0.0012	$7.7*10^{-12}$	141537406	142031024	0.98	PTK2
rs56116032	9	11453149	A	G	0.009	0.0014	$6.1*10^{-11}$	10961747	11995517	1.00	
rs9775606	9	122663012	G	C	-0.007	0.0012	$1.1*10^{-8}$	122624675	122676561	1.00	
rs999483	9	135301389	T	G	-0.008	0.0013	$1.9*10^{-10}$	135298675	135301389	1.00	

rs11599236	10	106454672	T	C	0.007	0.0012	$5.4*10^{-10}$	106385271	106768514	0.98	SORCS3
rs16932966	11	16382882	G	A	-0.008	0.0013	$3.1*10^{-10}$	16049786	16779416	0.95	
rs2868996	11	47575370	A	T	-0.009	0.0012	$9.4*10^{-14}$	46331362	51579831	0.97	LRP4,NR1H3, RAPSN,NDUFS3, PTPMT1,KBTBD4, MTCH2,AC019330.1 , AC133680.1
rs72995548	11	113138921	C	T	0.014	0.0023	$3.2*10^{-9}$	112834877	113305505	1.00	NCAM1, DRD2
rs4309187	11	113412443	A	C	-0.009	0.0012	$5.7*10^{-14}$	113267064	113451229	0.99	
rs61915924	12	16363975	A	C	0.008	0.0013	$1.3*10^{-8}$	16313384	16423287	1.00	
rs28655666	12	122186317	G	A	0.007	0.0011	$1.7*10^{-9}$	121925104	122315537	1.00	
rs3843954	13	58548511	G	C	-0.009	0.0013	$1.4*10^{-12}$	58318099	58710258	1.00	PCDH17,FARP1
rs3012850	13	58713360	C	G	-0.009	0.0014	$8.3*10^{-12}$	58215142	59208681	0.97	PCDH17,FARP1
rs61361413	13	66542715	C	G	-0.009	0.0016	$6.8*10^{-9}$	66346805	66908302	0.99	
rs144511092	13	99101057	C	CATTTT	-0.007	0.0012	$9.4*10^{-9}$	99090040	99262567	0.99	STK24
14:71702336	14	71702336	GTA	G	-0.007	0.0013	$5.3*10^{-9}$	71352456	72219146	0.97	
14:75225369	14	75225369	CAAA/TAAAT	C	-0.007	0.0012	$1.3*10^{-8}$	75056894	75672107	0.99	
rs8045174	16	5843110	C	T	-0.009	0.0014	$1.0*10^{-9}$	5785432	5854206	0.99	
rs1559422	16	71359660	T	C	-0.007	0.0012	$3.2*10^{-8}$	71016597	71986073	0.98	*CALB2*
rs1050863	16	87440895	G	A	0.008	0.0012	$2.6*10^{-12}$	87270478	87610218	0.99	
rs2950706	17	44343004	C	T	-0.014	0.0015	$6.6*10^{-20}$	43399058	44874453	0.91	CRHR1, MAPT, WNT3
rs4799949	18	35155910	C	T	0.008	0.0012	$6.5*10^{-10}$	35086406	35408205	1.00	CELF4
rs766498304	18	50871256	G	GA	-0.009	0.0014	$5.6*10^{-10}$	50358109	51055069	0.80	DCC
rs56403421	18	52765283	A	C	-0.008	0.0012	$3.1*10^{-11}$	52717101	52834988	0.99	
rs613872	18	53210302	G	T	-0.012	0.0015	$9.9*10^{-5}$	52835378	53725929	1.00	TCF4
rs6103271	20	42015848	A	G	0.009	0.0015	$6.4*10^{-9}$	41862171	42177113	0.98	
rs765051254	22	41461514	C	CA	-0.009	0.0014	$6.9*10^{-11}$	40937725	42234937	0.89	RANGAP1

4.3.1.2 Supplementary Table S2 Validation of the four top SNPs identified in the previous mood instability GWAS.

SNP = rs tag for the SNP being validated. CHR = chromosome number, BP = base position of the SNP, A1 = the tested allele, A2 = the other allele, A1FREQ = frequency of the A1 allele , info = imputation score, BETA = the coefficient for the SNP, S.E.= the standard error of the beta, CHISQ = the test statistic of the SNP, p = the p value.

SNP	CHR	BP	A1	A2	A1FREQ	INFO	BETA	S.E.	CHISQ	p
rs7829975	8	8548117	A	T	0.52	0.99	0.0049	0.0017	8.40291	0.0037
rs10959826	9	11459410	G	A	0.78	0.99	0.0065	0.0020	10.4055	0.0013
14:75268920_CA_C	14	75268920	C	CA	0.60	0.98	0.0041	0.0017	5.74312	0.017
rs8084280	18	50726749	T	A	0.51	1.00	0.0051	0.0017	9.13489	0.0025

4.3.1.3 Supplementary Table S3 Genes identified.

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.

geneName	txStart	txEnd	geneName	txStart	txEnd	geneName	txStart	txEnd
PTGER3	71318035	71513491	GALNT15	16215828	16273499	TDGF1	46616044	46623952
ZRANB2-AS1	71512188	71532865	DPH3	16298567	16306496	LOC100132146	46653924	46668033
ZRANB2	71528973	71546972	OXNAD1	16306666	16347594	ALS2CL	46710484	46735146
MIR186	71533313	71533399	RFTN1	16357351	16555222	TMIE	46742822	46752413
ZRANB2-AS2	71547006	71703406	LINC00690	16577824	16582880	PRSS50	46753605	46759373
NEGR1	71868624	72748277	DAZL	16628300	16646297	PRSS46	46761072	46777921
NEGR1-IT1	72259914	72302695	PLCL2	16926451	17132098	PRSS45	46783580	46786245
LINC01360	73771852	73773255	MIR3714	16974687	16974752	PRSS42	46871893	46875585
LRP1B	140988995	142889270	TBC1D5	17198653	17741512	MYL3	46899356	46904973
MIR7157	141344194	141344254	LOC105376975	17784435	17787154	PTH1R	46919235	46945289
KYNU	143635194	143747106	LOC339862	18004063	18310410	CCDC12	46963219	47017991
ARHGAP15	143886898	144525921	LINC00691	24141464	24144738	NBEAL2	47021172	47051194
LOC101928386	144694633	144721722	THRB	24158644	24536313	NRADDP	47053031	47054957
GTDC1	144703580	144969146	LOC101927854	24192814	24231447	SETD2	47057897	47205467
ZEB2	145141941	145277958	THRB-AS1	24535577	24541502	KIF9-AS1	47205859	47285606
ZEB2-AS1	145277180	145278465	MIR4792	24562852	24562926	KIF9	47269515	47324337
LINC01412	145279434	145337001	RARB	24870813	25639422	SNORD13J	47292008	47292099
LOC105373656	145281116	145282011	LOC105376997	25424595	25426418	KLHL18	47324329	47388306

COQ10B	198318146	198340033	TOP2B	25639395	25705863	PTPN23	47422471	47454931
HSPD1	198351307	198364640	MIR4442	25706363	25706430	SCAP	47455174	47517449
SNORA105A	198351511	198351627	NGLY1	25760434	25824989	ELP6	47537129	47555199
SNORA105B	198351511	198351627	OXSM	25831562	25836025	CSPG5	47603727	47620359
HSPE1	198364720	198368187	LINC00692	25900022	25915186	SMARCC1	47627377	47823405
HSPE1-MOB4	198364720	198418423	LZTFL1	45864809	45957216	SNORD146	47703054	47703184
MOB4	198380294	198418423	FYCO1	45959390	46037316	DHX30	47844398	47858462
RFTN2	198435526	198540584	CXCR6	45984972	45989845	MIR1226	47891044	47891119
MARS2	198570027	198573114	XCR1	46062290	46068979	MAP4	47892179	48130769
BOLL	198591602	198649734	CCR1	46243199	46249832	CDC25A	48198667	48229801
PLCL1	198669425	199014608	CCR3	46283871	46308197	MIR4443	48238053	48238106
LINC01923	199164086	199239821	CCR2	46395234	46400868	CAMP	48264836	48266981
SATB2	200134222	200322819	LOC102724297	46406150	46448550	ZNF589	48282595	48312479
SATB2-AS1	200332820	200337481	CCR5	46411632	46417697	FCF1P2	48331768	48333168
LINC01877	200472790	200523855	CCRL2	46448720	46451014	NME6	48333806	48342901
FTCDNL1	200625258	200715896	LTF	46477495	46505161	SPINK8	48348335	48369831
ANKRD28	15708743	15798058	RTP3	46539484	46542439	MIR2115	48357849	48357949
MIR3134	15738804	15738878	LRRC2	46556877	46608040	FBXW12	48413708	48436190
MIR563	15915277	15915356	LRRC2-AS1	46598887	46601178	PLXNB1	48445260	48470872
geneName	txStart	txEnd	geneName	txStart	txEnd	geneName	txStart	txEnd
CCDC51	48473579	48481529	MIR4271	49311552	49311619	TUSC2	50362339	50365669
TMA7	48481647	48485616	USP4	49314576	49377536	RASSF1	50367216	50374895
ATRIP	48488113	48507054	GPX1	49394603	49396033	RASSF1-AS1	50374941	50375727
TREX1	48506918	48509044	RHOA	49396568	49449530	ZMYND10	50378536	50383177
SHISA5	48509196	48514742	TCTA	49449638	49453909	NPRL2	50384918	50388486
PFKFB4	48555116	48594356	AMT	49454210	49460111	CYB561D2	50388125	50391496
MIR6823	48587393	48587454	NICN1	49459765	49466757	TMEM115	50392179	50396939
UCN2	48599150	48601201	DAG1	49506135	49573051	CACNA2D2	50400043	50541028
COL7A1	48601505	48632593	BSN-AS2	49586738	49591799	C3orf18	50595455	50605223
MIR711	48616334	48616410	BSN	49591921	49708982	HEMK1	50606582	50622421
UQCRC1	48636431	48647098	APEH	49711434	49720934	CISH	50643884	50649262
SNORA94	48642352	48642583	MST1	49721379	49726196	MAPKAPK3	50649292	50686728
TMEM89	48658274	48659189	RNF123	49726931	49758962	MIR4787	50712510	50712594
SLC26A6	48663155	48671279	AMIGO3	49754266	49757238	DOCK3	50712671	51421629
MIR6824	48671068	48671131	GMPPB	49758908	49761407	MANF	51422667	51426828
CELSR3	48673895	48700348	IP6K1	49761727	49823627	RBM15B	51428698	51435336
MIR4793	48681626	48681713	CDHR4	49828166	49837268	DCAF1	51433297	51534018
CELSR3-AS1	48701200	48706603	FAM212A	49840686	49842463	RAD54L2	51572698	51702676
NCKIPSD	48711271	48723366	UBA7	49842637	49851391	TEX264	51705190	51738339
IP6K2	48725435	48754711	MIR5193	49843569	49843678	GRM2	51741048	51752625
PRKAR2A	48784011	48885283	TRAIP	49866027	49893992	IQCF6	51812576	51813203
PRKAR2A-AS1	48885369	48889415	CAMKV	49895413	49907655	IQCF4	51851618	51852355
SLC25A20	48894355	48936426	MST1R	49924435	49941070	IQCF3	51860898	51864874
ARIH2OS	48955220	48956818	MON1A	49946301	49967445	IQCF2	51895644	51897440
ARIH2	48956252	48966107	RBM6	49977476	50114685	IQCF5-AS1	51907611	51909783

P4HTM	49027340	49044581	RBM5	50126340	50156397	IQCF5	51907736	51909600
WDR6	49044494	49053386	RBM5-AS1	50137035	50138421	IQCF1	51928891	51937386
DALRD3	49052920	49056041	SEMA3F-AS1	50153454	50193518	RRP9	51967441	51975957
MIR425	49057580	49057667	SEMA3F	50192477	50226508	PARP3	51976320	51982883
NDUFAF3	49057907	49060926	MIR566	50210758	50210852	GPR62	51989329	51991520
MIR191	49058050	49058142	GNAT1	50229042	50235129	PCBP4	51991469	51996908
IMPDH2	49061761	49066875	SLC38A3	50242678	50258411	ABHD14B	52002525	52008646
QRICH1	49067139	49131065	GNAI2	50264119	50296786	ABHD14A	52009041	52015216
QARS	49133364	49142562	MIR5787	50264867	50264922	ABHD14A-ACY1	52009041	52023218
MIR6890	49137286	49137347	SEMA3B-AS1	50304072	50304803	ACY1	52017299	52023218
USP19	49145478	49158371	SEMA3B	50304989	50314602	RPL29	52027643	52029958
LAMB2	49158546	49170599	MIR6872	50310666	50310728	DUSP7	52082936	52090461
LAMB2P1	49190291	49191834	LSMEM2	50316457	50325545	LINC00696	52096109	52099128
CCDC71	49199967	49203785	IFRD2	50325162	50330026	POC1A	52109248	52188706
KLHDC8B	49209017	49213919	HYAL3	50330258	50336293	ALAS1	52232098	52248343
C3orf84	49215068	49229291	NAT6	50333832	50336320	TLR9	52255095	52260179
CCDC36	49235860	49295537	HYAL1	50337319	50341032	TWF2	52262625	52273183
C3orf62	49306029	49314508	HYAL2	50355220	50360281	LOC101929054	52273253	52275113

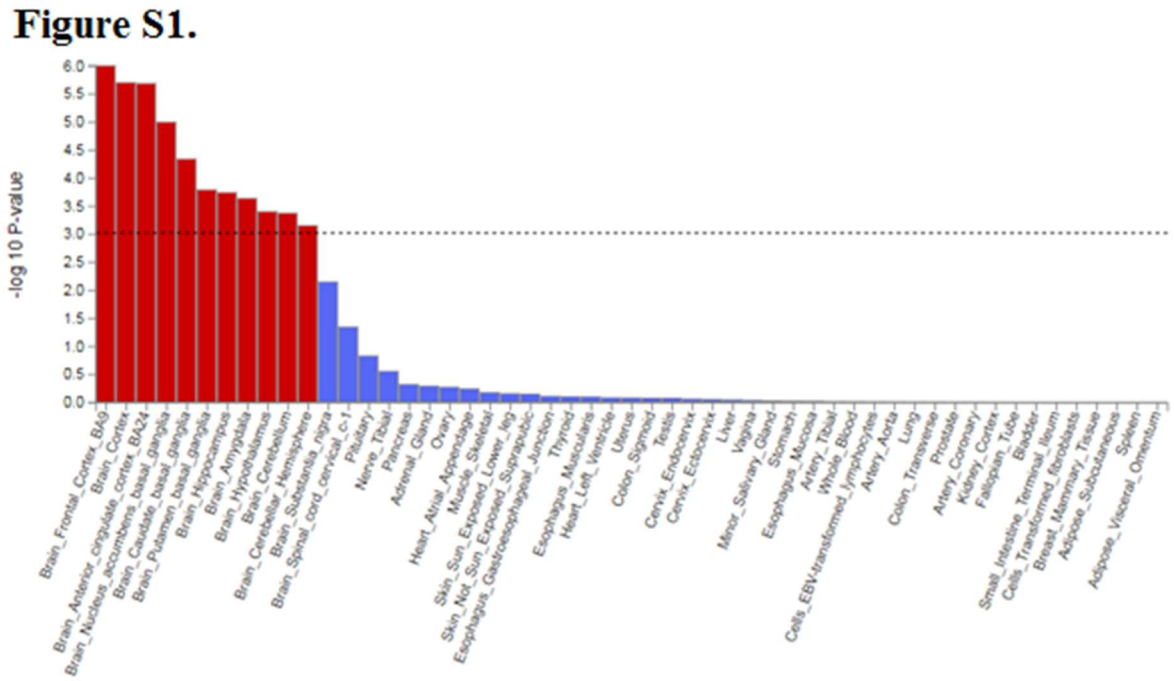
geneName	txStart	txEnd
PPM1M	52279808	52284615
WDR82	52288437	52312659
MIRLET7G	52302293	52302377

4.3.1.4 Supplementary Table S4, 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, S.E. = the standard error of the genes in the gene set, p = p value, pbon = the Bonferroni corrected p value

Gene Set	N genes	Beta	Beta STD	S.E.	p	pbon
go astrocyte differentiation	38	0.85	0.04	0.16	$2.57*10^{-8}$	0.0003
nikolsky breast cancer 16q24 amplicon	53	1.18	0.06	0.22	$3.63*10^{-8}$	0.0004
go neurogenesis	1346	0.14	0.04	0.03	$1.87*10^{-7}$	0.002
go neuron differentiation	835	0.16	0.03	0.03	$2.07*10^{-6}$	0.02
go glial cell differentiation	133	0.38	0.03	0.08	$3.03*10^{-6}$	0.03
go prepulse inhibition	12	1.43	0.04	0.32	$4.21*10^{-6}$	0.04

4.3.1.5 Supplementary Figure S1 MAGMA tissue expression analysis.



4.3.1.6 Supplementary Table S5 eQTL analysis summary results.

[a]Putative isoform specificity was defined as a SNP showing association to a specific transcript or exon junction. [b]Qualitative assessment of relative expression in brain vs. peripherally based on GTEx dataset. [c]SNP also associated with expression of an unannotated region or regions.

SNP	Implicated gene(s)	Evidence of potential isoform specificity?[a]	Expression in brain[b]
rs73082357	IP6K2	Yes	Present. Particularly prominent in cerebellum
rs6889822	SPINK9	No	Enriched in brain vs. periphery
	RP11-373N22.3	Yes	Present at similar levels to periphery.
rs763646	EXOC4	Yes	Present at similar levels to periphery.
rs1962104	PTK2	Yes	Present at similar levels to periphery.
rs7818437	RP11-981G7.1	No	Enriched in brain vs. periphery. Particularly prominent in cerebellum.
rs2898260	XKR6	No	Generally present at slightly lower levels than periphery but prominent in cerebellum.
	TDH	Yes	Present at similar levels to periphery.
	PINX1	Yes	Generally present at slightly lower levels than periphery but prominent in cerebellum.
	SOX7	Yes	Present, but at lower levels than periphery.
	FAM167A	No	Enriched in brain vs. periphery

rs4398922	FAM86B3P	Yes	Generally present at slightly lower levels than periphery but prominent in cerebellum.
	FAM85B	No	Present at similar levels to periphery.
	ALG1L13P	No	Generally present at slightly lower levels than periphery but prominent in cerebellum.
rs11039182[c]	SLC39A13	Yes	Generally present at slightly lower levels than periphery but prominent in cerebellum.
	FNBP4	No	Generally present at slightly lower levels than periphery but prominent in cerebellum.
rs669915[c]	LRRC37A4P	Yes	Present. Particularly prominent in cerebellum
	DND1P1	Yes	Generally present at slightly lower levels than periphery but prominent in cerebellum.
	CRHR1-IT1	Yes	Present. Particularly prominent in cerebellum
	RP11-707O23.5	No	Enriched in brain vs. periphery
	RP11-798G7.4	No	No information in GTEx
	RP11-798G7.5	No	Generally present at slightly lower levels than periphery but prominent in cerebellum.
	RP11-707O23.5	No	Enriched in brain vs. periphery
	RPS26P8	No	Present, but at lower levels than periphery.
	CRHR1	Yes	Enriched in brain vs. periphery. Particularly prominent in cerebellum
	MAPT	Yes	Enriched in brain vs. periphery
	MAPT-AS1	No	Enriched in brain vs. periphery
	PLEKHM1	Yes	Generally present at slightly lower levels than periphery but prominent in cerebellum.
	KANSL1	No	Generally present at slightly lower levels than periphery but prominent in cerebellum.
	ARHGAP27	No	Present, but at lower levels than periphery.
	RP11-798G7.6	No	Generally present at slightly lower levels than periphery but prominent in cerebellum.

4.3.1.7 Supplementary Table S6. eQTL analysis full results.

num loci	SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
3	rs73082357	chr3:48732258-48732522(*)	Junction	-0.44	3.77E-06	8.72E-04	N/A	chr3:48732258-48732522(*)	IP6K2	["ENST00000416707", "ENST00000412795", "ENST00000436134"]	28	G	A	3
3	rs73082357	e222801	Exon	-0.23	4.42E-06	5.53E-04	N/A	chr3:48731605-48731673(-)	IP6K2	["ENST00000455545"]	28	G	A	3
3	rs73082357	e222806	Exon	-0.23	4.82E-06	5.97E-04	N/A	chr3:48731608-48731673(-)	IP6K2	["ENST00000412795"]	28	G	A	3
3	rs73082357	e222805	Exon	-0.22	9.01E-06	1.04E-03	N/A	chr3:48732135-48732257(-)	IP6K2	["ENST00000412795"]	28	G	A	3
3	rs73082357	e222815	Exon	-0.25	9.36E-06	1.07E-03	N/A	chr3:48732204-48732257(-)	IP6K2	["ENST00000436134"]	28	G	A	3
3	rs73082357	e222795	Exon	-0.21	1.21E-05	1.34E-03	N/A	chr3:48731417-48731673(-)	IP6K2	["ENST00000450045"]	28	G	A	3
3	rs73082357	er73944	ER	-0.12	1.82E-05	1.92E-03	N/A	chr3:48730387-48732924(*)	IP6K2	[]	28	G	A	3
3	rs73082357	e222792	Exon	-0.20	2.19E-05	2.24E-03	N/A	chr3:48731386-48731673(-)	IP6K2	["ENST00000413298"]	28	G	A	3
3	rs73082357	e222782	Exon	-0.21	2.90E-05	2.86E-03	N/A	chr3:48732127-48732257(-)	IP6K2	["ENST00000416707"]	28	G	A	3
3	rs73082357	e222786	Exon	-0.13	3.61E-05	3.45E-03	N/A	chr3:48730954-48732854(-)	IP6K2	["ENST00000431721"]	28	G	A	3
3	rs73082357	e222766	Exon	-0.17	3.98E-05	3.75E-03	N/A	chr3:48730881-48731673(-)	IP6K2	["ENST00000340879"]	28	G	A	3
3	rs73082357	e222770	Exon	-0.17	4.11E-05	3.85E-03	N/A	chr3:48730886-48731673(-)	IP6K2	["ENST00000432678"]	28	G	A	3
3	rs73082357	e222779	Exon	-0.17	4.13E-05	3.87E-03	N/A	chr3:48730889-48731673(-)	IP6K2	["ENST00000433104"]	28	G	A	3
3	rs73082357	e222783	Exon	-0.17	4.26E-05	3.97E-03	N/A	chr3:48730921-48731673(-)	IP6K2	["ENST00000416707"]	28	G	A	3
3	rs73082357	e222788	Exon	-0.13	5.16E-05	4.67E-03	N/A	chr3:48731218-48732854(-)	IP6K2	["ENST00000449610"]	28	G	A	3
2	rs73082357	chr3:48731674-48732522(*)	Junction	-0.24	5.43E-05	8.73E-03	N/A	chr3:48731674-48732522(*)	IP6K2	["ENST00000340879", "ENST00000413298", "ENST00000450045", "ENST00000455545"]	28	G	A	3
2	rs73082357	e222798	Exon	-0.12	9.32E-05	7.70E-03	N/A	chr3:48731484-48732851(-)	IP6K2	["ENST00000446860"]	28	G	A	3
3	rs6889822	ENSG00000204909	Gene	0.10	2.12E-06	2.01E-04	8.99E-07	chr5:147700766-147719412(+)	SPINK9	["ENST00000511717", "ENST00000377906"]	2	A	G	3
3	rs6889822	TCONS_00538277	Transcript	0.04	8.21E-06	3.27E-03	3.74E-03	chr5:147703438-147704872(+)	RP11-373N22.3	["ENST00000501695"]	1	A	G	3

SNP	Feature	Type	beta	Discovery p-value	FDR p-value	CMC Replication	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt
rs6889822	e360799	Exon	0.14	1.20E-05	1.33E-03	1.35E-07	chr5:147703242-147703431(+)	SPINK9	["ENST00000511717"]	2	A	G
rs6889822	e360797	Exon	0.17	1.82E-05	1.91E-03	4.34E-06	chr5:147701843-147701891(+)		["ENST00000597116"]	1	A	G
rs6889822	ENSG00000268126	Gene	0.09	2.55E-05	1.84E-03	1.71E-06	chr5:147698550-147701891(+)		["ENST00000597116"]	1	A	G
rs6889822	e360798	Exon	0.11	5.85E-05	5.20E-03	5.69E-03	chr5:147700766-147700963(+)	SPINK9	["ENST00000511717"]	2	A	G
rs763646	chr7:133246671 -133314797(*)	Junction	0.31	6.28E-06	1.36E-03	3.04E-02	chr7:133246671-133314797(*)	EXOC4	[]	27	C	T
rs1962104	e558630	Exon	0.09	2.43E-09	6.02E-07	7.03E-04	chr8:141993941-141994138(-)	PTK2	["ENST00000523803"]	59	T	C
rs1962104	e558635	Exon	0.09	9.47E-08	1.74E-05	1.03E-03	chr8:141993941-141994093(-)	PTK2	["ENST00000521907"]	59	T	C
rs1962104	e558273	Exon	0.06	5.21E-06	6.39E-04	2.23E-06	chr8:141678368-141680363(-)	PTK2	["ENST00000517712"]	59	T	C
rs1962104	chr8:141678524 -141684396(*)	Junction	0.12	6.99E-06	1.50E-03	9.52E-09	chr8:141678524-141684396(*)	PTK2	["ENST00000522684", "ENST00000519654", "ENST00000535192", "ENST00000519465", "ENST00000517887", "ENST00000521059", "ENST00000523539", "ENST00000538769", "ENST00000519419", "ENST00000524202",	59	T	C
rs1962104	er176119	ER	0.06	2.98E-05	2.94E-03	4.09E-07	chr8:141678368-141679382(*)	PTK2	[]	59	T	C
rs1962104	er176120	ER	0.10	4.81E-05	4.44E-03	8.79E-06	chr8:141679456-141680431(*)	PTK2	[]	59	T	C
rs7818437	ENSG00000261451	Gene	0.07	1.38E-09	2.38E-07	4.16E-03	chr8:10291182-10295822(+)		["ENST00000562242"]	1	T	C
rs7818437	e522810	Exon	0.06	3.28E-08	6.62E-06	2.46E-02	chr8:10291182-10295822(+)		["ENST00000562242"]	1	T	C
rs2898260	er164726	ER	0.17	2.56E-07	4.25E-05	1.65E-05	chr8:10968617-10969402(*)	XKR6	[]	5	T	G
rs2898260	chr8:11219045 -11219143(*)	Junction	-0.41	3.50E-07	1.04E-04	8.89E-04	chr8:11219045-11219143(*)	TDH	["ENST00000525246", "ENST00000326605",	5	T	G
rs2898260	e522948	Exon	0.10	7.91E-07	1.19E-04	1.90E-03	chr8:10690401-10690494(-)	PINX1	["ENST00000520018"]	9	T	G

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs2898260	e522890	Exon	0.10	7.91E-07	1.19E-04	1.94E-03	chr8:10690402-10690494(-)	SOX7	["ENST00000553390", "ENST00000554914", "ENST00000314787", "ENST00000426190", "ENST00000519088", "ENST00000524114", "ENST00000523559",	2	T	G	3
rs2898260	chr8:10690495 -10692175(*)	Junction	0.12	4.13E-06	9.44E-04	1.61E-04	chr8:10690495-10692175(*)	SOX7	["ENST00000553390", "ENST00000554914", "ENST00000314787", "ENST00000426190", "ENST00000519088", "ENST00000524114",	2	T	G	3
rs2898260	er164725	ER	0.08	4.32E-06	5.41E-04	2.10E-03	chr8:10964926-10968482(*)	XKR6	[]	1	T	G	3
rs2898260	er164690	ER	0.11	7.04E-06	8.33E-04	4.34E-04	chr8:10690402-10690494(*)	SOX7	[]	9	T	G	3
rs2898260	ENSG00000154319	Gene	-0.06	6.51E-05	4.14E-03	1.68E-05	chr8:11278972-11332224(-)	FAM167A	["ENST00000284486", "ENST00000531564", "ENST00000534308", "ENST00000527445", "ENST00000528897", "ENST00000531804",	7	T	G	3
rs4398922	e522393	Exon	-0.30	1.49E-13	6.79E-11	8.02E-07	chr8:8088365-8088427(+)	FAM86B3P	["ENST00000522601", "ENST00000588728", "ENST00000523992"]	8	G	C	3
rs4398922	ENSG00000173295	Gene	-0.09	4.90E-12	1.20E-09	6.24E-08	chr8:8086117-8102387(+)	FAM86B3P	["ENST00000522601", "ENST00000588728", "ENST00000523992", "ENST00000591069", "ENST00000590591", "ENST00000310542", "ENST00000588159",	8	G	C	3
rs4398922	e522386	Exon	-0.21	2.25E-11	7.65E-09	8.77E-05	chr8:8083918-8084136(-)	FAM85B	["ENST00000519726"]	1	G	C	3
rs4398922	e522420	Exon	-0.12	8.94E-11	2.79E-08	3.91E-04	chr8:8097891-8099954(+)	FAM86B3P	["ENST00000310542"]	8	G	C	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs4398922	e522401	Exon	-0.17	2.12E-10	6.23E-08	3.12E-07	chr8:8086122-8086319(+)	FAM86B3P	["ENST00000588728"]	8	G	C	3
rs4398922	e522417	Exon	-0.22	1.63E-09	4.16E-07	9.44E-03	chr8:8101621-8101818(+)	FAM86B3P	["ENST00000591069"]	8	G	C	3
rs4398922	e522399	Exon	-0.21	2.07E-09	5.19E-07	1.28E-02	chr8:8101621-8101844(+)	FAM86B3P	["ENST00000522601"]	8	G	C	3
rs4398922	er164399	ER	-0.31	4.08E-09	9.47E-07	1.02E-03	chr8:8098658-8099462(*)	FAM86B3P	[]	4	G	C	3
rs4398922	e522424	Exon	-0.17	4.56E-09	1.08E-06	5.10E-04	chr8:8099150-8099484(+)	FAM86B3P	["ENST00000587202"]	8	G	C	3
rs4398922	chr8:8086320 -8088364(*)	Junction	-0.44	5.74E-09	2.45E-06	1.31E-03	chr8:8086320-8088364(*)	FAM86B3P	["ENST00000588728"]	8	G	C	3
rs4398922	e522387	Exon	-0.20	8.21E-09	1.85E-06	8.48E-03	chr8:8081348-8081409(-)	FAM85B	["ENST00000519726"]	1	G	C	3
rs4398922	e522422	Exon	-0.17	2.07E-08	4.33E-06	1.01E-03	chr8:8098797-8099202(+)	FAM86B3P	["ENST00000588159"]	8	G	C	3
rs4398922	e522406	Exon	-0.09	2.88E-08	5.88E-06	4.56E-03	chr8:8097149-8097552(+)	FAM86B3P	["ENST00000588728"]	8	G	C	3
rs4398922	e522416	Exon	-0.21	4.00E-08	7.94E-06	3.62E-03	chr8:8099063-8099202(+)	FAM86B3P	["ENST00000591069"]	8	G	C	3
rs4398922	e522421	Exon	-0.14	4.03E-08	7.98E-06	6.75E-03	chr8:8101621-8102385(+)	FAM86B3P	["ENST00000310542"]	8	G	C	3
rs4398922	er164400	ER	-0.28	1.11E-07	1.99E-05	3.51E-02	chr8:8101621-8102254(*)	FAM86B3P	[]	4	G	C	3
rs4398922	e522431	Exon	-0.20	1.27E-07	2.28E-05	4.32E-03	chr8:8097951-8098071(-)	ALG1L13P	["ENST00000523017", "ENST00000522393"]	4	G	C	3
rs4398922	e522407	Exon	-0.17	2.23E-07	3.80E-05	4.15E-06	chr8:8086127-8086222(+)	FAM86B3P	["ENST00000523992"]	8	G	C	3
rs4398922	e522436	Exon	-0.20	2.42E-07	4.09E-05	5.80E-03	chr8:8097960-8098070(-)	ALG1L13P	["ENST00000519320"]	4	G	C	3
rs4398922	e522396	Exon	-0.16	2.42E-07	4.09E-05	5.13E-05	chr8:8093338-8093471(+)	FAM86B3P	["ENST00000522601", "ENST00000588728"]	8	G	C	3
rs4398922	e522392	Exon	-0.16	2.51E-07	4.23E-05	4.08E-06	chr8:8086117-8086222(+)	FAM86B3P	["ENST00000522601"]	8	G	C	3
rs4398922	e522411	Exon	-0.16	3.10E-07	5.12E-05	5.70E-05	chr8:8093338-8093469(+)	FAM86B3P	["ENST00000523992"]	8	G	C	3
rs4398922	ENSG00000253893	Gene	-0.09	4.47E-07	4.94E-05	1.72E-02	chr8:8025341-8084136(-)	FAM85B	["ENST00000519726"]	1	G	C	3
rs4398922	e522394	Exon	-0.21	1.06E-06	1.55E-04	3.30E-04	chr8:8090316-8090396(+)	FAM86B3P	["ENST00000522601", "ENST00000588728", "ENST00000523992"]	8	G	C	3
rs4398922	e522435	Exon	-0.18	1.67E-06	2.33E-04	1.57E-02	chr8:8098818-8098933(-)	ALG1L13P	["ENST00000519320"]	4	G	C	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs4398922	chr8:8093472 -8094647(*)	Junction	-0.36	2.34E-06	5.72E-04	2.86E-03	chr8:8093472-8094647(*)	FAM86B3P	["ENST00000588728"]	8	G	C	3
rs4398922	chr8:8088428 -8090315(*)	Junction	-0.40	3.64E-06	8.44E-04	9.39E-03	chr8:8088428-8090315(*)	FAM86B3P	["ENST00000522601", "ENST00000588728",	8	G	C	3
rs4398922	e522445	Exon	-0.16	6.58E-06	7.85E-04	1.12E-04	chr8:8097746-8097873(-)	ALG1L13P	["ENST00000522393"]	4	G	C	3
rs4398922	e522430	Exon	-0.19	1.46E-05	1.58E-03	1.23E-02	chr8:8098844-8098922(-)	ALG1L13P	["ENST00000523017", "ENST00000522393"]	4	G	C	3
rs4398922	e522437	Exon	-0.16	5.28E-05	4.77E-03	3.20E-04	chr8:8097728-8097835(-)	ALG1L13P	["ENST00000519320"]	4	G	C	3
rs4398922	e522388	Exon	-0.13	7.52E-05	6.43E-03	1.55E-02	chr8:8077566-8077704(-)	FAM85B	["ENST00000519726"]	1	G	C	3
rs4398922	er164391	ER	0.20	1.19E-04	9.59E-03	3.77E-02	chr8:8095751-8095863(*)	FAM86B3P	[]	4	G	C	3
rs2729940	ENSG00000255098	Gene	0.13	3.01E-06	2.76E-04	2.96E-05	chr8:11870545-11873043(-)		["ENST00000527396"]	1	G	A	3
rs2729940	er164894	ER	0.30	4.51E-06	5.61E-04	1.23E-04	chr8:11872785-11875054(*)		[]	1	G	A	3
rs2729940	er164886	ER	0.23	5.19E-06	6.35E-04	3.92E-07	chr8:11863372-11864269(*)		[]	0	G	A	3
rs2729940	e523586	Exon	0.12	2.81E-05	2.78E-03	7.11E-04	chr8:11872340-11872575(-)		["ENST00000527396"]	1	G	A	3
rs2729940	TCONS_00646265	Transcript	0.14	2.89E-05	9.50E-03	8.06E-04	chr8:11860909-11861591(-)	RP11-481A20.10	["ENST00000602581"]	2	G	A	3
rs2729940	e523583	Exon	0.11	4.78E-05	4.38E-03	4.22E-06	chr8:11860909-11861591(-)		["ENST00000602581"]	2	G	A	3
rs2729940	chr8:11872582 -11872970(*)	Junction	-0.33	5.18E-05	8.39E-03	1.77E-03	chr8:11872582-11872970(*)	NA	[]	1	G	A	3
rs2729940	er164726	ER	0.12	5.80E-05	5.22E-03	1.04E-02	chr8:10968617-10969402(*)	XKR6	[]	5	G	A	3
rs2729940	er164882	ER	0.16	1.07E-04	8.77E-03	1.50E-09	chr8:11859938-11863296(*)		[]	2	G	A	3
rs2729940	er164755	ER	0.14	1.18E-04	9.54E-03	1.63E-02	chr8:11203483-11204149(*)	TDH	[]	1	G	A	3
rs2729940	ENSG00000255020	Gene	0.07	1.29E-04	7.41E-03	1.70E-02	chr8:11203257-11205011(-)		["ENST00000525867"]	1	G	A	3
rs2729940	ENSG00000254507	Gene	0.10	1.36E-04	7.74E-03	2.17E-06	chr8:11860909-11861591(-)		["ENST00000602581", "ENST00000532329"]	2	G	A	3
rs11039182	chr11:47434019 -47435147(*)	Junction	-0.63	9.64E-14	8.20E-11	4.62E-12	chr11:47434019-47435147(*)	SLC39A13	["ENST00000531419", "ENST00000531865"]	13	T	C	3
rs11039182	er219046	ER	-0.38	6.65E-10	1.75E-07	1.23E-07	chr11:47434073-47434254(*)	SLC39A13	[]	13	T	C	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs11039182	er219120	ER	0.36	2.68E-07	4.43E-05	5.85E-07	chr11:47626773-47628326(*)		[]	0	T	C	3
rs11039182	er219127	ER	0.37	3.11E-06	4.04E-04	5.48E-07	chr11:47628412-47629025(*)		[]	0	T	C	3
rs11039182	er219119	ER	0.33	1.64E-05	1.75E-03	4.95E-06	chr11:47625941-47626529(*)	C1QTNF4	[]	0	T	C	3
rs11039182	er219132	ER	0.34	2.46E-05	2.50E-03	9.90E-05	chr11:47629101-47629354(*)		[]	0	T	C	3
rs11039182	er219134	ER	0.32	3.34E-05	3.25E-03	4.52E-05	chr11:47629374-47629622(*)		[]	0	T	C	3
rs11039182	er219139	ER	0.33	4.95E-05	4.56E-03	2.53E-05	chr11:47629687-47630246(*)		[]	0	T	C	3
rs11039182	er219147	ER	0.33	6.42E-05	5.69E-03	1.40E-05	chr11:47631432-47631975(*)		[]	0	T	C	3
rs11039182	er219136	ER	0.15	7.99E-05	6.86E-03	3.32E-03	chr11:47629627-47629643(*)		[]	0	T	C	3
rs11039182	er219153	ER	0.26	8.56E-05	7.27E-03	2.13E-02	chr11:47634370-47634420(*)		[]	0	T	C	3
rs11039182	er219148	ER	0.30	1.05E-04	8.65E-03	5.25E-06	chr11:47632290-47632721(*)		[]	0	T	C	3
rs11039182	er219195	ER	-0.16	1.19E-04	9.59E-03	3.61E-02	chr11:47788621-47788922(*)	FNBP4	[]	16	T	C	3
rs669915	ENSG00000214425	Gene	-0.26	6.60E-104	1.96E-99	3.99E-65	chr17:43578685-43627701(-)	LRRC37A4P	["ENST00000398305", "ENST00000579913", "ENST00000581296"]	3	G	A	3
rs669915	er315985	ER	1.87	1.09E-93	6.98E-89	2.88E-45	chr17:43680193-43680220(*)		[]	0	G	A	3
rs669915	e999882	Exon	-0.45	1.21E-89	4.73E-85	8.15E-26	chr17:43584107-43585905(-)	LRRC37A4P	["ENST00000579913"]	3	G	A	3
rs669915	er315907	ER	-1.53	1.29E-85	4.59E-81	5.47E-17	chr17:43583985-43585304(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er315906	ER	-1.72	5.90E-83	1.82E-78	1.22E-53	chr17:43583301-43583907(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er315901	ER	-1.66	3.21E-79	7.95E-75	6.52E-42	chr17:43582584-43583026(*)	LRRC37A4P	[]	2	G	A	3
rs669915	e999874	Exon	-1.07	1.53E-77	2.59E-73	3.26E-63	chr17:43594544-43594648(-)	LRRC37A4P	["ENST00000398305"]	3	G	A	3
rs669915	chr17:43594649 -43595226(*)	Junction	-1.61	2.15E-74	7.57E-70	3.01E-59	chr17:43594649-43595226(*)	LRRC37A4P	["ENST00000398305"]	3	G	A	3
rs669915	er315929	ER	-1.27	4.59E-74	9.22E-70	4.66E-55	chr17:43594504-43594648(*)	LRRC37A4P	[]	2	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	er315908	ER	-1.34	1.35E-71	2.34E-67	2.17E-35	chr17:43585362-43585905(*)	LRRC37A4P	[]	2	G	A	3
rs669915	e999873	Exon	-1.26	2.50E-70	3.10E-66	8.57E-39	chr17:43595227-43595264(-)	LRRC37A4P	["ENST00000398305"]	3	G	A	3
rs669915	er315924	ER	-1.47	6.58E-57	5.46E-53	9.08E-56	chr17:43591946-43592237(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er316010	ER	0.37	6.21E-56	4.91E-52	1.68E-38	chr17:43722379-43723568(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	e999893	Exon	-0.64	1.07E-51	6.53E-48	3.27E-29	chr17:43585633-43585772(-)	LRRC37A4P	["ENST00000581296"]	3	G	A	3
rs669915	e999877	Exon	-0.62	6.20E-51	3.67E-47	1.41E-29	chr17:43585588-43585752(-)	LRRC37A4P	["ENST00000398305"]	3	G	A	3
rs669915	e999920	Exon	0.37	3.71E-49	2.01E-45	1.27E-50	chr17:43678235-43679706(-)		["ENST00000580257"]	1	G	A	3
rs669915	TCONS_00285480	Transcript	0.20	9.09E-49	2.01E-44	1.51E-15	chr17:43663237-43664295(+)	DND1P1	["ENST00000580842"]	1	G	A	3
rs669915	e999961	Exon	0.37	1.03E-47	5.06E-44	7.21E-42	chr17:43723267-43723598(+)	CRHR1-IT1	["ENST00000592428"]	14	G	A	3
rs669915	e999930	Exon	0.38	1.14E-47	5.63E-44	3.97E-42	chr17:43723267-43723557(+)	CRHR1-IT1	["ENST00000591271"]	14	G	A	3
rs669915	e999954	Exon	0.37	1.25E-47	6.14E-44	7.64E-42	chr17:43723267-43723596(+)	CRHR1-IT1	["ENST00000586362"]	14	G	A	3
rs669915	e999933	Exon	0.32	2.11E-47	1.02E-43	1.76E-23	chr17:43707270-43707822(+)	CRHR1-IT1	["ENST00000582491"]	14	G	A	3
rs669915	ENSG00000263503	Gene	0.36	2.87E-47	7.20E-44	7.73E-50	chr17:43678235-43679706(-)		["ENST00000580257"]	1	G	A	3
rs669915	er315991	ER	0.38	3.45E-46	1.66E-42	5.01E-24	chr17:43707270-43707524(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	TCONS_00304400	Transcript	0.42	4.94E-46	8.18E-42	1.71E-42	chr17:43678235-43679706(-)	RP11-707O23.5	["ENST00000580257"]	1	G	A	3
rs669915	e999925	Exon	0.34	8.57E-46	3.80E-42	1.44E-24	chr17:43707270-43707524(+)	CRHR1-IT1	["ENST00000585118", "ENST00000591271", "ENST00000583740", "ENST00000585122", "ENST00000455565", "ENST00000293493", "ENST00000580220", "ENST00000582044",	14	G	A	3
rs669915	er315910	ER	-0.91	1.12E-45	5.14E-42	9.20E-43	chr17:43587610-43587659(*)	LRRC37A4P	[]	2	G	A	3
rs669915	e999875	Exon	-0.80	1.15E-44	4.79E-41	7.86E-33	chr17:43587610-43587659(-)	LRRC37A4P	["ENST00000398305", "ENST00000579913", "ENST00000581296"]	3	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	chr17:43592892 -43594543(*)	Junction	-1.34	4.26E-44	3.49E-40	7.16E-56	chr17:43592892-43594543(*)	LRRC37A4P	[]	3	G	A	3
rs669915	e999876	Exon	-0.78	7.38E-44	2.97E-40	2.16E-30	chr17:43585851-43585905(-)	LRRC37A4P	["ENST00000398305", "ENST00000581296"]	3	G	A	3
rs669915	e999958	Exon	0.41	1.55E-43	6.10E-40	2.41E-40	chr17:43723267-43723374(+)	CRHR1-IT1	["ENST00000585677"]	14	G	A	3
rs669915	chr17:43587660 -43594543(*)	Junction	-1.13	6.43E-38	3.77E-34	9.95E-46	chr17:43587660-43594543(*)	LRRC37A4P	["ENST00000398305"]	3	G	A	3
rs669915	e999894	Exon	-0.78	1.51E-37	4.42E-34	1.37E-07	chr17:43585475-43585526(-)	LRRC37A4P	["ENST00000581296"]	3	G	A	3
rs669915	e999963	Exon	0.18	1.83E-37	5.35E-34	1.58E-27	chr17:43723037-43725582(+)	CRHR1-IT1	["ENST00000589868"]	14	G	A	3
rs669915	er315982	ER	1.27	9.52E-36	2.62E-32	4.63E-37	chr17:43665675-43665906(*)	DND1P1	[]	0	G	A	3
rs669915	chr17:43699408 -43707269(*)	Junction	0.39	1.69E-34	8.14E-31	9.16E-20	chr17:43699408-43707269(*)	CRHR1-IT1	["ENST00000585118", "ENST00000591271", "ENST00000582491", "ENST00000583740", "ENST00000585122", "ENST00000455565", "ENST00000293493", "ENST00000582044",	14	G	A	3
rs669915	e999889	Exon	-0.17	3.64E-31	7.37E-28	5.76E-68	chr17:43591234-43592745(-)	LRRC37A4P	["ENST00000581296"]	3	G	A	3
rs669915	TCONS_00285483	Transcript	1.35	5.44E-31	4.20E-27	7.08E-55	chr17:43685909-43686349(+)	RPS26P8	["ENST00000536258"]	2	G	A	3
rs669915	er315933	ER	-0.90	6.10E-31	1.24E-27	2.45E-02	chr17:43595393-43595548(*)	LRRC37A4P	[]	2	G	A	3
rs669915	e999878	Exon	-0.22	6.16E-31	1.23E-27	1.99E-18	chr17:43578685-43579396(-)	LRRC37A4P	["ENST00000398305"]	3	G	A	3
rs669915	e999919	Exon	0.19	8.96E-31	1.76E-27	1.73E-12	chr17:43663237-43664295(+)	DND1P1	["ENST00000580842"]	1	G	A	3
rs669915	TCONS_00285481	Transcript	0.12	5.99E-30	4.27E-26	6.44E-33	chr17:43663237-43673326(+)	DND1P1	["ENST00000580842"]	1	G	A	3
rs669915	ENSG00000264070	Gene	0.19	7.89E-30	8.27E-27	4.19E-12	chr17:43663237-43664295(+)	DND1P1	["ENST00000580842"]	1	G	A	3
rs669915	chr17:43585906 -43587609(*)	Junction	-1.00	1.25E-28	4.19E-25	2.04E-28	chr17:43585906-43587609(*)	LRRC37A4P	["ENST00000398305", "ENST00000579913", "ENST00000581296"]	3	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	er315904	ER	-0.75	1.31E-28	2.26E-25	7.54E-05	chr17:43583124-43583197(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er316091	ER	0.58	2.14E-28	3.65E-25	2.51E-15	chr17:43916170-43916706(*)	CRHR1	[]	0	G	A	3
rs669915	er316009	ER	0.36	3.59E-28	5.99E-25	3.86E-03	chr17:43721595-43722262(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	er316180	ER	-0.56	8.35E-28	1.36E-24	1.77E-25	chr17:44086728-44086754(*)	MAPT	[]	19	G	A	3
rs669915	er316086	ER	0.50	4.66E-27	7.14E-24	1.10E-11	chr17:43914830-43915785(*)	CRHR1	[]	0	G	A	3
rs669915	er316107	ER	-0.74	5.75E-27	8.70E-24	2.35E-02	chr17:44038987-44039053(*)	MAPT	[]	19	G	A	3
rs669915	chr17:43707525 -43723266(*)	Junction	1.03	9.71E-27	2.78E-23	8.77E-25	chr17:43707525-43723266(*)	CRHR1-IT1	["ENST00000591271"]	14	G	A	3
rs669915	ENSG00000204650	Gene	0.10	1.25E-26	1.05E-23	2.38E-13	chr17:43697694-43725582(+)	CRHR1-IT1	["ENST00000585118", "ENST00000591271", "ENST00000582491", "ENST00000583740", "ENST00000585122", "ENST00000455565", "ENST00000580220", "ENST00000444561", "ENST00000586362", "ENST00000578000", "ENST00000585677", "ENST00000592428", "ENST00000580655", "ENST00000589868"]	14	G	A	3
rs669915	e1000108	Exon	0.41	6.90E-26	9.82E-23	7.59E-15	chr17:44051751-44051837(+)	MAPT	["ENST00000262410", "ENST00000344290", "ENST00000351559", "ENST00000535772", "ENST00000415613", "ENST00000431008", "ENST00000571987", "ENST00000574436"]	19	G	A	3
rs669915	e999879	Exon	-0.17	2.96E-25	4.03E-22	1.91E-68	chr17:43591234-43592522(-)	LRRC37A4P	["ENST00000579913"]	3	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	er315932	ER	-0.45	5.90E-25	7.72E-22	1.91E-68	chr17:43595045-43595351(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er315923	ER	-0.82	4.05E-24	4.97E-21	3.80E-67	chr17:43591419-43591555(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er316095	ER	0.33	1.50E-23	1.76E-20	4.10E-10	chr17:43917973-43919511(*)	MAPT-AS1	[]	0	G	A	3
rs669915	TCONS_00285510	Transcript	0.24	1.52E-23	6.63E-20	7.18E-17	chr17:43697976-43727752(+)	CRHR1	["ENST00000293493"]	13	G	A	3
rs669915	er315983	ER	0.82	9.41E-22	9.57E-19	1.51E-66	chr17:43677491-43677588(*)		[]	0	G	A	3
rs669915	er315902	ER	-0.35	1.73E-21	1.71E-18	2.23E-05	chr17:43583095-43583113(*)	LRRC37A4P	[]	2	G	A	3
rs669915	e999978	Exon	0.25	5.98E-21	5.80E-18	1.16E-06	chr17:43699292-43699407(+)	CRHR1	["ENST00000587305", "ENST00000339069"]	13	G	A	3
rs669915	e999964	Exon	0.24	6.35E-21	6.15E-18	1.15E-06	chr17:43699267-43699407(+)	CRHR1	["ENST00000293493"]	13	G	A	3
rs669915	e999924	Exon	0.25	7.01E-21	6.76E-18	1.17E-06	chr17:43699274-43699407(+)	CRHR1-IT1	["ENST00000585118", "ENST00000591271", "ENST00000582491", "ENST00000583740", "ENST00000585122", "ENST00000455565", "ENST00000580220",	14	G	A	3
rs669915	chr17:44049312 -44051750(*)	Junction	0.48	1.73E-20	2.92E-17	5.62E-09	chr17:44049312-44051750(*)	MAPT	["ENST00000262410", "ENST00000344290", "ENST00000351559", "ENST00000535772", "ENST00000415613", "ENST00000431008", "ENST00000571987",	19	G	A	3
rs669915	er316094	ER	0.55	3.03E-20	2.70E-17	1.58E-05	chr17:43917448-43917674(*)	MAPT-AS1	[]	0	G	A	3
rs669915	chr17:44039124 -44039686(*)	Junction	-0.83	4.89E-20	7.91E-17	5.08E-07	chr17:44039124-44039686(*)	MAPT	[]	19	G	A	3
rs669915	er316115	ER	0.45	7.91E-20	6.77E-17	1.18E-10	chr17:44051751-44051837(*)	MAPT	[]	19	G	A	3
rs669915	TCONS_00304391	Transcript	-0.24	1.19E-19	3.81E-16	1.40E-16	chr17:43590829-43595661(-)	LRRC37A4P	["ENST00000398305"]	3	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	er315891	ER	-0.26	7.99E-19	6.18E-16	1.03E-06	chr17:43574138-43574149(*)	LRRC37A4P	[]	0	G	A	3
rs669915	er315905	ER	-0.24	1.09E-18	8.33E-16	6.06E-04	chr17:43583204-43583215(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er316007	ER	0.26	1.27E-18	9.63E-16	1.47E-06	chr17:43720049-43720864(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	er316085	ER	0.49	1.75E-18	1.30E-15	1.24E-06	chr17:43914181-43914732(*)	CRHR1	[]	0	G	A	3
rs669915	er316088	ER	0.27	1.23E-17	8.43E-15	1.37E-06	chr17:43916132-43916145(*)	CRHR1	[]	0	G	A	3
rs669915	er316089	ER	0.25	2.53E-17	1.68E-14	2.28E-14	chr17:43916149-43916161(*)	CRHR1	[]	0	G	A	3
rs669915	er316072	ER	0.58	4.94E-17	3.17E-14	1.29E-03	chr17:43910304-43910433(*)	CRHR1	[]	13	G	A	3
rs669915	er316092	ER	0.52	2.00E-16	1.21E-13	8.32E-10	chr17:43916723-43916907(*)	CRHR1	[]	0	G	A	3
rs669915	er315989	ER	0.45	2.84E-16	1.68E-13	2.28E-14	chr17:43698332-43698518(*)	CRHR1-IT1	[]	14	G	A	3
rs669915	er315892	ER	-0.31	1.59E-15	8.66E-13	3.78E-05	chr17:43574190-43574206(*)	LRRC37A4P	[]	0	G	A	3
rs669915	er315886	ER	-0.28	3.31E-15	1.74E-12	4.09E-10	chr17:43572907-43572920(*)	PLEKHM1	[]	0	G	A	3
rs669915	er315885	ER	-0.35	3.61E-15	1.90E-12	9.79E-11	chr17:43572907-43572920(*)	PLEKHM1	[]	0	G	A	3
rs669915	er316087	ER	0.52	5.40E-15	2.77E-12	1.76E-08	chr17:43572907-43572920(*)	CRHR1	[]	0	G	A	3
rs669915	e1000103	Exon	-0.25	9.16E-15	4.81E-12	4.37E-06	chr17:43572907-43572920(*)	MAPT-AS1	["ENST00000581125"]	3	G	A	3
rs669915	er315888	ER	-0.28	1.23E-14	6.04E-12	1.64E-07	chr17:43572907-43572920(*)	LRRC37A4P	[]	0	G	A	3
rs669915	TCONS_00312713	Transcript	0.08	1.75E-14	3.31E-11	1.14E-08	chr17:43572907-43572920(*)	NA	["NA"]	0	G	A	3
rs669915	er315911	ER	-0.61	3.73E-14	1.74E-11	6.96E-07	chr17:43572907-43572920(*)	LRRC37A4P	[]	2	G	A	3
rs669915	e999943	Exon	0.14	4.82E-14	2.33E-11	1.49E-09	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000455565"]	14	G	A	3
rs669915	er316073	ER	0.37	5.63E-14	2.57E-11	8.39E-08	chr17:43572907-43572920(*)	CRHR1	[]	13	G	A	3
rs669915	e999956	Exon	0.13	8.72E-14	4.09E-11	3.11E-05	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000578000"]	14	G	A	3
rs669915	er315884	ER	-0.36	1.08E-13	4.77E-11	3.27E-09	chr17:43572907-43572920(*)	PLEKHM1	[]	0	G	A	3
rs669915	chr17:44051838 -44055740(*)	Junction	0.44	1.11E-13	9.36E-11	5.01E-06	chr17:43572907-43572920(*)	MAPT	["ENST00000262410", "ENST00000344290", "ENST00000351559", "ENST00000535772", "ENST00000415613",	19	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	er315891	ER	-0.26	7.99E-19	6.18E-16	1.03E-06	chr17:43574138-43574149(*)	LRRC37A4P	[]	0	G	A	3
rs669915	er315905	ER	-0.24	1.09E-18	8.33E-16	6.06E-04	chr17:43583204-43583215(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er316007	ER	0.26	1.27E-18	9.63E-16	1.47E-06	chr17:43720049-43720864(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	er316085	ER	0.49	1.75E-18	1.30E-15	1.24E-06	chr17:43914181-43914732(*)	CRHR1	[]	0	G	A	3
rs669915	er316088	ER	0.27	1.23E-17	8.43E-15	1.37E-06	chr17:43916132-43916145(*)	CRHR1	[]	0	G	A	3
rs669915	er316089	ER	0.25	2.53E-17	1.68E-14	2.28E-14	chr17:43916149-43916161(*)	CRHR1	[]	0	G	A	3
rs669915	er316072	ER	0.58	4.94E-17	3.17E-14	1.29E-03	chr17:43910304-43910433(*)	CRHR1	[]	13	G	A	3
rs669915	er316092	ER	0.52	2.00E-16	1.21E-13	8.32E-10	chr17:43916723-43916907(*)	CRHR1	[]	0	G	A	3
rs669915	er315989	ER	0.45	2.84E-16	1.68E-13	2.28E-14	chr17:43698332-43698518(*)	CRHR1-IT1	[]	14	G	A	3
rs669915	er315892	ER	-0.31	1.59E-15	8.66E-13	3.78E-05	chr17:43574190-43574206(*)	LRRC37A4P	[]	0	G	A	3
rs669915	er315886	ER	-0.28	3.31E-15	1.74E-12	4.09E-10	chr17:43572907-43572920(*)	PLEKHM1	[]	0	G	A	3
rs669915	er315885	ER	-0.35	3.61E-15	1.90E-12	9.79E-11	chr17:43572907-43572920(*)	PLEKHM1	[]	0	G	A	3
rs669915	er316087	ER	0.52	5.40E-15	2.77E-12	1.76E-08	chr17:43572907-43572920(*)	CRHR1	[]	0	G	A	3
rs669915	e1000103	Exon	-0.25	9.16E-15	4.81E-12	4.37E-06	chr17:43572907-43572920(*)	MAPT-AS1	["ENST00000581125"]	3	G	A	3
rs669915	er315888	ER	-0.28	1.23E-14	6.04E-12	1.64E-07	chr17:43572907-43572920(*)	LRRC37A4P	[]	0	G	A	3
rs669915	TCONS_00312713	Transcript	0.08	1.75E-14	3.31E-11	1.14E-08	chr17:43572907-43572920(*)	NA	["NA"]	0	G	A	3
rs669915	er315911	ER	-0.61	3.73E-14	1.74E-11	6.96E-07	chr17:43572907-43572920(*)	LRRC37A4P	[]	2	G	A	3
rs669915	e999943	Exon	0.14	4.82E-14	2.33E-11	1.49E-09	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000455565"]	14	G	A	3
rs669915	er316073	ER	0.37	5.63E-14	2.57E-11	8.39E-08	chr17:43572907-43572920(*)	CRHR1	[]	13	G	A	3
rs669915	e999956	Exon	0.13	8.72E-14	4.09E-11	3.11E-05	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000578000"]	14	G	A	3
rs669915	er315884	ER	-0.36	1.08E-13	4.77E-11	3.27E-09	chr17:43572907-43572920(*)	PLEKHM1	[]	0	G	A	3
rs669915	e999962	Exon	0.13	1.14E-13	5.24E-11	3.70E-05	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000580655"]	14	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	chr17:44051838 -44055740(*)	Junction	0.44	1.11E-13	9.36E-11	5.01E-06	chr17:43572907-43572920(*)	MAPT	["ENST00000262410", "ENST00000344290", "ENST00000351559", "ENST00000535772", "ENST00000415613", "ENST00000431008", "ENST00000571987",	19	G	A	3
rs669915	er316045	ER	-0.42	1.16E-13	5.11E-11	1.98E-03	chr17:43572907-43572920(*)	CRHR1	[]	13	G	A	3
rs669915	er316044	ER	-0.48	1.64E-13	7.06E-11	3.73E-03	chr17:43572907-43572920(*)	CRHR1	[]	13	G	A	3
rs669915	er315961	ER	0.52	2.50E-13	1.06E-10	2.38E-23	chr17:43572907-43572920(*)		[]	1	G	A	3
rs669915	e1000101	Exon	-0.20	5.70E-13	2.41E-10	2.39E-06	chr17:43572907-43572920(*)	MAPT-AS1	["ENST00000579599"]	3	G	A	3
rs669915	TCONS_00285503	Transcript	0.31	1.10E-12	1.66E-09	2.39E-09	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000455565"]	14	G	A	3
rs669915	er315917	ER	-0.27	1.11E-12	4.32E-10	4.52E-04	chr17:43572907-43572920(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er315894	ER	0.53	1.11E-12	4.34E-10	7.04E-03	chr17:43572907-43572920(*)	LRRC37A4P	[]	0	G	A	3
rs669915	er316082	ER	0.44	2.46E-12	9.19E-10	7.36E-09	chr17:43572907-43572920(*)	CRHR1	[]	0	G	A	3
rs669915	er316100	ER	-0.31	3.85E-12	1.40E-09	1.59E-06	chr17:43572907-43572920(*)	MAPT-AS1	[]	3	G	A	3
rs669915	er316001	ER	0.25	4.51E-12	1.63E-09	4.95E-02	chr17:43572907-43572920(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	TCONS_00285497	Transcript	-0.46	1.48E-11	1.91E-08	1.23E-13	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000585118"]	14	G	A	3
rs669915	er315999	ER	0.36	3.32E-11	1.06E-08	1.94E-07	chr17:43572907-43572920(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	er315990	ER	0.18	3.34E-11	1.07E-08	1.05E-12	chr17:43572907-43572920(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	er315919	ER	-0.43	4.03E-11	1.28E-08	2.80E-36	chr17:43572907-43572920(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er316084	ER	0.35	9.50E-11	2.85E-08	3.55E-05	chr17:43572907-43572920(*)	CRHR1	[]	0	G	A	3
rs669915	chr17:43585773 -43585850(*)	Junction	-0.72	1.03E-10	5.90E-08	4.11E-09	chr17:43572907-43572920(*)	LRRC37A4P	["ENST00000581296"]	3	G	A	3
rs669915	e1000098	Exon	-0.23	1.50E-10	4.53E-08	3.73E-03	chr17:43572907-43572920(*)	MAPT-AS1	["ENST00000579244"]	3	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	er315998	ER	0.41	1.58E-10	4.58E-08	2.18E-04	chr17:43572907-43572920(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	e999960	Exon	0.23	1.86E-10	5.53E-08	7.21E-03	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000592428"]	14	G	A	3
rs669915	er316005	ER	0.17	2.44E-10	6.89E-08	2.40E-03	chr17:43572907-43572920(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	er315984	ER	0.62	3.80E-10	1.04E-07	1.03E-03	chr17:43572907-43572920(*)		[]	1	G	A	3
rs669915	er315925	ER	-0.19	6.64E-10	1.75E-07	9.64E-67	chr17:43572907-43572920(*)	LRRC37A4P	[]	2	G	A	3
rs669915	chr17:43921528 -43972845(*)	Junction	-0.41	7.90E-10	3.92E-07	1.68E-04	chr17:43572907-43572920(*)	MAPT-AS1	["ENST00000579244"]	3	G	A	3
rs669915	er316079	ER	0.37	7.97E-10	2.08E-07	7.72E-06	chr17:43572907-43572920(*)	CRHR1	[]	0	G	A	3
rs669915	e1000112	Exon	0.20	9.13E-10	2.43E-07	2.21E-05	chr17:43572907-43572920(*)	MAPT	["ENST00000262410", "ENST00000344290", "ENST00000415613", "ENST00000571987"]	19	G	A	3
rs669915	e999773	Exon	0.14	9.24E-10	2.46E-07	3.03E-02	chr17:43572907-43572920(*)	PLEKHM1	["ENST00000430334", "ENST00000421073", "ENST00000579197", "ENST00000446609"]	19	G	A	3
rs669915	er316093	ER	0.43	1.00E-09	2.57E-07	1.48E-07	chr17:43572907-43572920(*)	MAPT-AS1	[]	0	G	A	3
rs669915	er316006	ER	0.13	1.05E-09	2.68E-07	3.34E-05	chr17:43572907-43572920(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	e1000099	Exon	-0.14	1.09E-09	2.87E-07	4.43E-04	chr17:43572907-43572920(*)	MAPT-AS1	["ENST00000579244"]	3	G	A	3
rs669915	er316102	ER	-0.32	1.55E-09	3.85E-07	4.12E-03	chr17:43572907-43572920(*)	MAPT-AS1	[]	19	G	A	3
rs669915	er316052	ER	-0.39	1.63E-09	4.03E-07	2.24E-02	chr17:43572907-43572920(*)	CRHR1	[]	13	G	A	3
rs669915	er315996	ER	0.42	1.73E-09	4.27E-07	1.66E-06	chr17:43572907-43572920(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	er315964	ER	0.41	4.30E-09	9.94E-07	1.43E-11	chr17:43572907-43572920(*)		[]	1	G	A	3
rs669915	er315920	ER	-0.50	5.43E-09	1.23E-06	9.18E-58	chr17:43572907-43572920(*)	LRRC37A4P	[]	2	G	A	3
rs669915	TCONS_00285468	Transcript	-0.21	1.08E-08	8.73E-06	4.12E-10	chr17:43572907-43572920(*)	LRRC37A4P	["ENST00000581296"]	3	G	A	3
rs669915	e999991	Exon	0.16	1.23E-08	2.69E-06	5.94E-07	chr17:43572907-43572920(*)	CRHR1	["ENST00000347197"]	13	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	e1000004	Exon	0.16	1.33E-08	2.88E-06	5.58E-07	chr17:43572907-43572920(*)	CRHR1	["ENST00000314537"]	13	G	A	3
rs669915	chr17:43884464 -43906580(*)	Junction	0.54	1.40E-08	5.55E-06	3.63E-04	chr17:43572907-43572920(*)	CRHR1	[]	13	G	A	3
rs669915	er316142	ER	0.18	2.34E-08	4.76E-06	8.76E-04	chr17:43572907-43572920(*)	MAPT	[]	19	G	A	3
rs669915	er316128	ER	0.26	2.50E-08	5.06E-06	1.54E-05	chr17:43572907-43572920(*)	MAPT	[]	19	G	A	3
rs669915	e999951	Exon	0.21	3.39E-08	6.82E-06	3.39E-04	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000444561"]	14	G	A	3
rs669915	er316081	ER	0.35	6.06E-08	1.14E-05	1.07E-06	chr17:43572907-43572920(*)	CRHR1	[]	0	G	A	3
rs669915	e999895	Exon	-0.07	7.16E-08	1.35E-05	4.32E-05	chr17:43572907-43572920(*)		["ENST00000253803"]	2	G	A	3
rs669915	er315874	ER	0.18	8.17E-08	1.50E-05	3.67E-03	chr17:43572907-43572920(*)	PLEKHM1	[]	19	G	A	3
rs669915	er315992	ER	0.38	9.96E-08	1.80E-05	2.35E-06	chr17:43572907-43572920(*)	CRHR1-IT1	[]	2	G	A	3
rs669915	TCONS_00285486	Transcript	0.02	1.65E-07	1.04E-04	2.25E-10	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000591271"]	14	G	A	3
rs669915	chr17:43717822 -43723266(*)	Junction	0.51	1.80E-07	5.71E-05	6.81E-04	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000592428"]	14	G	A	3
rs669915	e999921	Exon	0.06	2.13E-07	3.64E-05	6.85E-15	chr17:43572907-43572920(*)	RPS26P8	["ENST00000536258"]	2	G	A	3
rs669915	ENSG00000267198	Gene	0.01	3.66E-07	4.11E-05	6.49E-18	chr17:43572907-43572920(*)		["ENST00000590995", "ENST00000587078", "ENST00000591950", "ENST00000586348",	5	G	A	3
rs669915	er316139	ER	0.17	4.96E-07	7.75E-05	2.18E-03	chr17:43572907-43572920(*)	MAPT	[]	19	G	A	3
rs669915	er315962	ER	0.24	6.47E-07	9.85E-05	3.49E-07	chr17:43572907-43572920(*)		[]	1	G	A	3
rs669915	chr17:43535806 -43552465(*)	Junction	-0.47	1.06E-06	2.82E-04	1.76E-05	chr17:43572907-43572920(*)	PLEKHM1	["ENST00000581448"]	19	G	A	3
rs669915	TCONS_00304381	Transcript	-0.06	1.33E-06	6.67E-04	4.27E-03	chr17:43572907-43572920(*)	LRRC37A4P	["ENST00000581296"]	3	G	A	3
rs669915	er315937	ER	-0.16	1.51E-06	2.12E-04	6.61E-63	chr17:43572907-43572920(*)	LRRC37A4P	[]	2	G	A	3
rs669915	er316038	ER	0.24	1.53E-06	2.14E-04	1.33E-07	chr17:43572907-43572920(*)	CRHR1	[]	2	G	A	3
rs669915	er315922	ER	-0.25	1.63E-06	2.26E-04	1.31E-65	chr17:43572907-43572920(*)	LRRC37A4P	[]	2	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	ENSG00000204652	Gene	0.05	1.65E-06	1.61E-04	5.25E-15	chr17:43572907-43572920(*)	RPS26P8	["ENST00000536258", "ENST00000393490"]	2	G	A	3
rs669915	ENSG00000264589	Gene	-0.06	1.83E-06	1.77E-04	6.83E-03	chr17:43572907-43572920(*)	MAPT-AS1	["ENST00000579244", "ENST00000579599", "ENST00000581125"]	3	G	A	3
rs669915	e999886	Exon	-0.11	2.54E-06	3.38E-04	1.71E-10	chr17:43572907-43572920(*)	LRRC37A4P	["ENST00000581296"]	3	G	A	3
rs669915	TCONS_00285498	Transcript	0.14	3.23E-06	1.46E-03	3.08E-14	chr17:43572907-43572920(*)	CRHR1-IT1	["ENST00000591271"]	14	G	A	3
rs669915	e1000017	Exon	0.20	3.35E-06	4.33E-04	2.34E-06	chr17:43572907-43572920(*)	CRHR1	["ENST00000352855"]	13	G	A	3
rs669915	e999922	Exon	0.06	3.80E-06	4.83E-04	5.72E-14	chr17:43572907-43572920(*)	RPS26P8	["ENST00000393490"]	2	G	A	3
rs669915	er315967	ER	0.31	4.37E-06	5.46E-04	5.65E-13	chr17:43572907-43572920(*)		[]	1	G	A	3
rs669915	chr17:43906688 -43907459(*)	Junction	0.19	6.14E-06	1.34E-03	6.11E-04	chr17:43572907-43572920(*)	CRHR1	["ENST00000293493", "ENST00000339069", "ENST00000314537", "ENST00000352855", "ENST00000577353", "ENST00000582766", "ENST00000580955", "ENST00000583888"]	13	G	A	3
rs669915	e1000282	Exon	-0.03	1.13E-05	1.26E-03	2.17E-13	chr17:43572907-43572920(*)	KANSL1	["ENST00000572218"]	18	G	A	3
rs669915	TCONS_00304352	Transcript	-0.18	1.33E-05	4.93E-03	8.12E-04	chr17:43572907-43572920(*)	PLEKHM1	["ENST00000581448"]	19	G	A	3

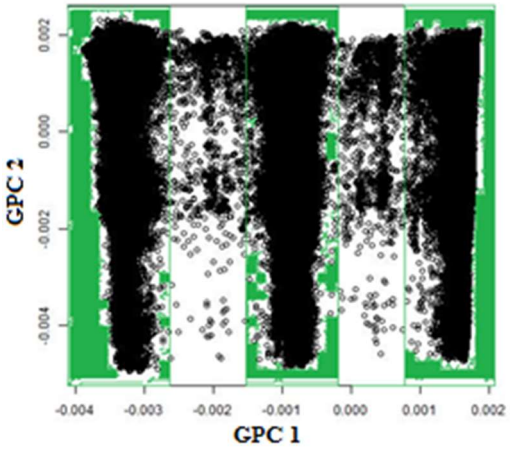
SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	ENSG00000186868	Gene	-0.04	1.84E-05	1.38E-03	5.49E-10	chr17:43572907-43572920(*)	MAPT	["ENST00000262410", "ENST00000340799", "ENST00000344290", "ENST00000347967", "ENST00000351559", "ENST00000535772", "ENST00000571311", "ENST00000570299", "ENST00000334239", "ENST00000446361", "ENST00000415613", "ENST00000420682", "ENST00000431008", "ENST00000571987", "ENST00000574436", "ENST00000576518", "ENST00000572440", "ENST00000576238",	19	G	A	3
rs669915	er316054	ER	-0.25	2.28E-05	2.33E-03	1.65E-02	chr17:43572907-43572920(*)	CRHR1	[]	13	G	A	3
rs669915	e1000100	Exon	-0.07	2.72E-05	2.70E-03	3.91E-03	chr17:43572907-43572920(*)	MAPT-AS1	["ENST00000579599"]	3	G	A	3
rs669915	e999818	Exon	0.07	2.83E-05	2.79E-03	4.62E-02	chr17:43572907-43572920(*)	PLEKHM1	["ENST00000581448"]	19	G	A	3
rs669915	e1000029	Exon	0.20	2.84E-05	2.81E-03	5.65E-03	chr17:43572907-43572920(*)	CRHR1	["ENST00000577353", "ENST00000398285"]	13	G	A	3
rs669915	er316029	ER	-0.15	3.14E-05	3.08E-03	5.65E-03	chr17:43572907-43572920(*)	CRHR1	[]	2	G	A	3
rs669915	er316041	ER	0.17	3.80E-05	3.63E-03	2.36E-03	chr17:43572907-43572920(*)	CRHR1	[]	13	G	A	3
rs669915	e1000105	Exon	-0.07	5.72E-05	5.10E-03	6.54E-03	chr17:43572907-43572920(*)	MAPT	["ENST00000262410", "ENST00000340799", "ENST00000344290", "ENST00000347967", "ENST00000351559", "ENST00000535772"]	19	G	A	3
rs669915	e1000184	Exon	-0.07	7.07E-05	6.11E-03	4.56E-03	chr17:43572907-43572920(*)	MAPT	["ENST00000570299"]	19	G	A	3

SNP	Feature	Type	beta	Discovery p value	FDR p value	CMC Replication p value	Coordinates	Symbol	WhichTx	NumTx Gene	Ref	alt	num Imp
rs669915	e1000179	Exon	-0.07	7.08E-05	6.12E-03	4.60E-03	chr17:43572907-43572920(*)	MAPT	["ENST00000571311"]	19	G	A	3
rs669915	e999966	Exon	0.14	7.16E-05	6.17E-03	3.12E-08	chr17:43572907-43572920(*)	CRHR1	["ENST00000293493", "ENST00000582044", "ENST00000587305", "ENST00000339069", "ENST00000347197", "ENST00000314537", "ENST00000352855", "ENST00000577353", "ENST00000398285", "ENST00000582766"]	13	G	A	3
rs669915	e1000088	Exon	0.13	8.10E-05	6.85E-03	3.14E-06	chr17:43572907-43572920(*)		["ENST00000582044"]	2	G	A	3
rs669915	e1000200	Exon	-0.08	9.69E-05	7.96E-03	7.11E-04	chr17:43572907-43572920(*)	MAPT	["ENST00000446361"]	19	G	A	3
rs669915	er315979	ER	-0.34	9.93E-05	8.24E-03	4.88E-05	chr17:43572907-43572920(*)		[]	5	G	A	3
rs669915	ENSG00000159314	Gene	0.04	0.000139	7.88E-03	1.52E-02	chr17:43572907-43572920(*)	ARHGAP27	["ENST00000532038", "ENST00000376922", "ENST00000531735", "ENST00000528384", "ENST00000590026", "ENST00000581991", "ENST00000455881", "ENST00000428638", "ENST00000442348", "ENST00000532891", "ENST00000581638", "ENST00000582826", "ENST00000529357", "ENST00000526484", "ENST00000524404", "ENST00000532667", "ENST00000528677", "ENST00000527678", "ENST00000579357", "ENST00000528273",	21	G	A	3

4.3.1.8 Supplementary Figure S2 Analysis of chromosome 17q21.31 inversion polymorphism genotype using genetic principal components.

GPCs 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 S2.



4.4 Novel genome-wide associations for suicidality in UK Biobank, genetic correlation with psychiatric disorders and polygenic association with completed suicide.

EBioMedicine 41 (2019) 517–525

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Identification of novel genome-wide associations for suicidality in UK Biobank, genetic correlation with psychiatric disorders and polygenic association with completed suicide

Rona J. Strawbridge [a,b,1], Joey Ward [a,1], Amy Ferguson [a], Nicholas Graham [a], Richard J. Shaw [a], Breda Cullen [a], Robert Pearsall [a], Laura M. Lyall [a], Keira J.A. Johnston [a,c,d], Claire L. Niedzwiedz [a], Jill P. Pell [a], Daniel Mackay [a], Julie Langan Martin [a], Donald M. Lyall [a], Mark E.S. Bailey [d], Daniel J. Smith [a,*]

[a] Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK
[b] Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden
[c] Division of Psychiatry, College of Medicine, University of Edinburgh, UK
[d] School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

ARTICLE INFO

Article history:
Received 5 November 2018
Received in revised form 4 February 2019
Accepted 4 February 2019
Available online 8 February 2019

ABSTRACT

Background: Suicide is a major issue for global public health. Suicidality describes a broad spectrum of thoughts and behaviours, some of which are common in the general population. Although suicide results from a complex interaction of multiple social and psychological factors, predisposition to suicidality is at least partly genetic.
Methods: Ordinal genome-wide association study of suicidality in the UK Biobank cohort comparing: 'no suicidality' controls ($N = 83{,}557$); 'thoughts that life was not worth living' ($N = 21{,}063$); 'ever contemplated self-harm' ($N = 13{,}038$); 'act of deliberate self-harm in the past' ($N = 2498$); and 'previous suicide attempt' ($N = 2666$).
Outcomes: We identified three novel genome-wide significant loci for suicidality (on chromosomes nine, 11 and 13) and moderate-to-strong genetic correlations between suicidality and a range of psychiatric disorders, most notably depression (r_g 0·81).
Interpretation: These findings provide new information about genetic variants relating to increased risk of suicidal thoughts and behaviours. Future work should assess the extent to which polygenic risk scores for suicidality, in combination with non-genetic risk factors, may be useful for stratified approaches to suicide prevention at a population level.
Fund: UKRI Innovation-HDR-UK Fellowship (MR/S003061/1). MRC Mental Health Data Pathfinder Award (MC_PC_17217). MRC Doctoral Training Programme Studentship at the University of Glasgow (MR/K501335/1). MRC Doctoral Training Programme Studentship at the Universities of Glasgow and Edinburgh. UKRI Innovation Fellowship (MR/R024774/1).
© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Suicide is a major and growing issue for global public health. Annually, approximately 800,000 people die by suicide and 20 times this number will attempt suicide during their lifetime [1]. 'Suicidality' encompasses a broad range of experiences and behaviours, from suicidal ideas/thoughts, to acts of deliberate self-harm and suicide attempts, occurring along a spectrum towards completed suicide [2]. Some components of suicidal thoughts and behaviours, such as feeling that life is not worth living or contemplating self-harm, are relatively common in the general population, as well as in patients affected by a variety of separate clinical diagnoses. Suicidality can therefore be considered a complex dimensional trait that fits within a Research Domain Criteria (RDoC) approach because it cuts across traditional psychiatric diagnostic classifications.

Pathways to completed suicide are complex and multifactorial [3]. Suicidal thoughts and actions are a consequence of a dynamic interplay between genetic, other biological, psychiatric, psychological, and a wide range of important social, economic and cultural factors [4]. Clinically, deliberate self-harm (DSH) is a major risk factor for subsequent suicidal behaviour. It is also recognised that substance abuse-related disorders

* Corresponding author at: Institute of Health and Wellbeing, University of Glasgow, Room 111, Public Health, 1 Lilybank Gardens, Glasgow G12 8RZ, UK.
E-mail address: Daniel.Smith@glasgow.ac.uk (D.J. Smith).
[1] Equal contribution.

https://doi.org/10.1016/j.ebiom.2019.02.005
2352-3964/© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Research in context

Evidence before this study

To date genetic studies of suicidal behaviour (ideation, attempt, completion) have mainly been conducted in individuals with severe mental illness and the few findings reported have failed to replicate in subsequent studies.

Added value of this study

This is the first study to explore the genetics of a broad suicidality phenotype within a large population-based cohort (N = 122 k individuals). Mutually exclusive categories of 'no suicidal behaviour', 'contemplated self-harm', 'actual self-harm', 'suicidal ideation' and 'suicide attempt' were assessed in an ordinal genome-wide association study (GWAS). A risk score of suicidality demonstrated associations with an independent subset of completed suicide. Moderate-to-strong genetic correlations were observed with all major psychiatric traits. Separate GWAS analyses of deliberate self-harm or suicide ideation and attempts suggested that the genetic contributions to these traits have distinct components. In a recent independent study of suicide attempts, the lead genetic variant identified in our study demonstrated consistent replication of effect size and direction.

Implications

This is the first report of robust genetic associations with suicidality and has the potential to lead to improved understanding of biological mechanisms underlying suicidal thoughts and behaviour. Further exploration of the impact of genetic risk scores in combination with clinical, social and environmental factors is now warranted.

and mood disorders are particularly associated with suicide risk [5]. Similarly, early adversities such as childhood sexual abuse [6], maladaptive parenting [7] and parental loss [8] all contribute to suicidal thoughts and behaviours, either directly or by increasing the risk of psychiatric disorders [9,10]. Personality-level traits such as neuroticism, impaired decision-making and sensitivity to negative social stimuli also contribute to suicidality [11].

Family, adoption and twin studies suggest a heritability estimate for suicidal behaviour of approximately 38–55% [12], and thus suicidality as a behavioural trait is amenable to genetic investigation. Heritability estimates for less clearly defined phenotypes such as suicidal thoughts are difficult to establish [13]. It is clear, however, that genetic predisposition plays a role in suicide alongside individual and social factors. Genetic studies may offer some insight into the biological basis of suicidality but genome-wide association studies (GWAS) [14–24] or family-based [25] findings to date have been limited, perhaps as a result of subsequent studies being under-powered or because of diagnostic heterogeneity. Recent advances in this area include a GWAS of suicide attempts in approximately 50,000 individuals with and without psychiatric disorders which identified some suggestive loci [26], replication of a single GWAS-significant finding on Chromosome 2 [17], nominal evidence for replication of a SNP on Chromosome 6 [22] and a locus on Chromosome 3 [15] and demonstration that suicide attempt and clinically predicted suicide share significant heritability [27].

Our goals in this study were: to identify genetic variants associated with broadly-defined suicidality in 122,935 participants of the UK Biobank cohort; to assess for genetic correlations between suicidality and a range of psychiatric disorders; and to determine whether increased genetic burden for suicidality was associated with both psychiatric disorders and completed suicide in a non-overlapping sample. Broadly-defined suicidality included thoughts and actions of both suicide and deliberate self-harm, despite current debate over the extent to which deliberate self-harm and suicidal intent overlap. Our primary analyses used dimensions of suicidality ordered to reflect clinical severity: from no suicidal thoughts or behaviours, thoughts that life is not worth living, considered self-harm, actual self-harm and attempted suicide. In secondary analyses, mindful that not all DSH behaviours necessarily carry active suicidal intent, we also conducted additional separate GWAS analyses of DSH and suicidal ideation/attempts (SIA).

2. Materials and methods

2.1. Sample description

UK Biobank is a large general population cohort. Between 2006 and 2010, approximately 502,000 participants (age range 37–73 years) were recruited and attended one of 22 assessment centres across mainland UK [28,29]. Comprehensive baseline assessments included sociodemographic characteristics, cognitive abilities, lifestyle and measures of mental and physical health status (Supplementary Methods and Supplementary Fig. 1). To maximise sample homogeneity, only white British participants were included in the current analysis. Informed consent was obtained by UK Biobank from all participants. This study was carried out under the generic approval from the NHS National Research Ethics Service (approval letter dated 13 May 2016, Ref 16/NW/0274) and under UK Biobank approval for application #6553 "Genome-wide association studies of mental health" (PI Daniel Smith).

2.2. Suicidality phenotypes

Suicidality groups were based on four questions from the self-harm behaviours section of the online mental health (*'Thoughts and Feelings'*) questionnaire administered in 2016/2017: (http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=136 and Supplementary Methods [30]). Non-overlapping categories of increasing severity of suicidality were derived: 'no suicidality' controls; 'thoughts that life was not worth living'; 'ever contemplated self-harm or suicide'; 'acts of deliberate self-harm not including attempted suicide'; 'attempted suicide'. If participants met criteria for more than one category they were assigned to the most severe category. Those in the lowest category were required to have answered "no" to all the questions. Linkage to death certification (until February 2016) identified a separate sub-group of participants classified as 'completed suicide' (defined as primary cause of death by intentional self-harm, ICD codes X60-X84; N = 137). The latter group was not used in the ordinal GWAS but rather was used in a separate analysis to test for association with genetic loading for suicidality. Participants were classified based on the most extreme form of suicidality that they reported and placed within the 'no suicidality' group if they responded negatively to all self-harm and suicidality questions. Those who preferred not to answer any of the questions (0·7%) were excluded from analysis.

2.3. Genotyping, imputation and quality control

In July 2017 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) [29]. These arrays have over 95% content in common. Pre-imputation quality control, imputation and post-imputation cleaning were conducted centrally by UK Biobank (described in the UK Biobank release documentation [28,29]. Fully imputed genetic data released in March 2018 were used for this study, therefore a total of 8,930,390 SNPs were available for analysis.

2.4. Ordinal GWAS of suicidality, DSH and SIA

For each GWAS, we excluded at random one person from each related pair of individuals with a kinship coefficient > 0·042 (second cousins) that have valid phenotypes, therefore the number of controls and participants within each suicidality level is different for each analysis (Supplementary Methods and Supplementary Figs. 1 and 2).

The primary GWAS included 122,935 individuals. Of these, 83,557 were classified as controls (category 0), 21,063 were classified in the 'thoughts that life is not worth living' group (category 1), 13,038 in the 'thoughts of self-harm' group (category 2), 2498 in the 'actual self-harm' group (category 3), and 2666 in the 'attempted suicide' group (category 4).

For the secondary analyses, two further ordinal GWAS were conducted. For DSH, the categories of controls, ($N = 84{,}499$), 'thoughts of self-harm' ($N = 13{,}203$) and 'actual self-harm' ($N = 2532$) were assessed. For SIA, the categories of controls (N = 84,167), 'thoughts that life is not worth living' group ($N = 21{,}234$) and 'attempted suicide' ($N = 2689$) were assessed.

Analyses were performed in R (Version 3.1) using the clm function of the ordinal package [31] treating the multilevel suicidality, DSH or SIA outcome variable as an ordinal variable. Models were adjusted for age, sex, genotyping chip and UK Biobank-derived genetic principal components (GPCs) 1–8. For sensitivity analyses, a variable for psychiatric diagnosis was included as a covariate (where psychiatric diagnosis was defined as likely or self-reported bipolar disorder (BD), Generalized Anxiety Disorder (GAD) and Major Depressive Disorder (MDD) and schizophrenia (SZ)) [30,32]. Further sensitivity analyses also included self-reported childhood sexual abuse as a covariate [33]. Genome-wide significance was set at $P < 5 \times 10^{-8}$ and plots were generated using FUMA [34].

2.5. Polygenic Risk score (PRS) variables for suicidality, mood disorders and related traits

Polygenic risk scores (PRS) were calculated from the primary ordinal suicidality GWAS summary statistics after pruning based on linkage disequilibrium (Supplementary Methods). SNPs were included in the PRS if they met *p*-value thresholds of $p < 5 \times 10^{-8}$, $p < 5 \times 10^{-5}$, $p < 0{\cdot}01$ $p < 0{\cdot}05$, $p < 0{\cdot}1$ or $p \leq 0{\cdot}5$ (Supplementary Methods). PRS deciles were computed using STATA (version 12, STATACorp) and modelling of associations between the PRS and completed suicide was analysed with logistic regression, adjusting for age, sex, chip and GPCs 1–8. In this analysis, cases were individuals classified as 'completed suicide' ($n = 127$), and controls were those recorded as category 0 in the ordinal variable but who had been excluded from the GWAS due to relatedness ($n = 5330$). It should be noted that the category 0 individuals here were related to individuals across the spectrum of suicidality in the GWAS, thus are more representative of the general population distribution of PRS than the true "no suicidality" distribution. Therefore this analysis is a conservative approach, being biased towards the null. Associations between the PRS and risk of mood disorders and related traits were also assessed (Supplementary Methods). The traits tested (BD, MDD, mood instability, and risk-taking behaviour) were selected based on prior evidence of relevance to suicidality, therefore the threshold for significance was set at $p < 0{\cdot}05$.

2.6. SNP heritability and genetic correlation analyses

Linkage Disequilibrium Score Regression (LDSR) [35] was used to estimate the SNP heritability (h^2_{SNP}) of ordinal suicidality, DSH and SIA. LDSR was also used to calculate genetic correlations with suicide attempt, psychiatric disorders and related traits (Supplementary Methods). The resulting genetic correlation *P*-values were false discovery rate (FDR)-corrected to compensate for multiple testing.

2.7. Gene-based analysis

The ordinal GWAS results were also considered under a gene-based approach, using MAGMA [36], as implemented in FUMA [34].

2.8. Exploration of known biology

The Variant Effect Predictor web-based tool [37], GTEx database [38] and BRAINEAC dataset (http://braineac.org/) were interrogated to try to identify genes (based on expression quantitative trait loci, eQTLs) or mechanisms through which associated SNPs might be acting (Supplementary Methods). The GWAS catalogue (https://www.ebi.ac.uk/gwas/) and NCBI Gene https://www.ncbi.nlm.nih.gov/gene/) were queried for each of the suicidality-associated SNPs and genes reported here.

3. Results

3.1. Sociodemographic characteristics

Sociodemographic, clinical and health-related behaviour measures for each of the suicidality categories are shown in Supplementary Table 1. As expected, a gradient of increasing suicidality was found for increasing levels of social deprivation, living alone, current or previous smoking, parental depression and chronic pain. There were also substantial differences by sex: females accounted for 68·3% of those who reported attempted suicide but only 27·6% of completed suicides. A large proportion of those who had attempted suicide (85·1%) had a history of MDD, compared to only 14·9% of controls. Similarly, 75·8% of those with a suicide attempt also reported childhood trauma, compared to 39·0% of controls.

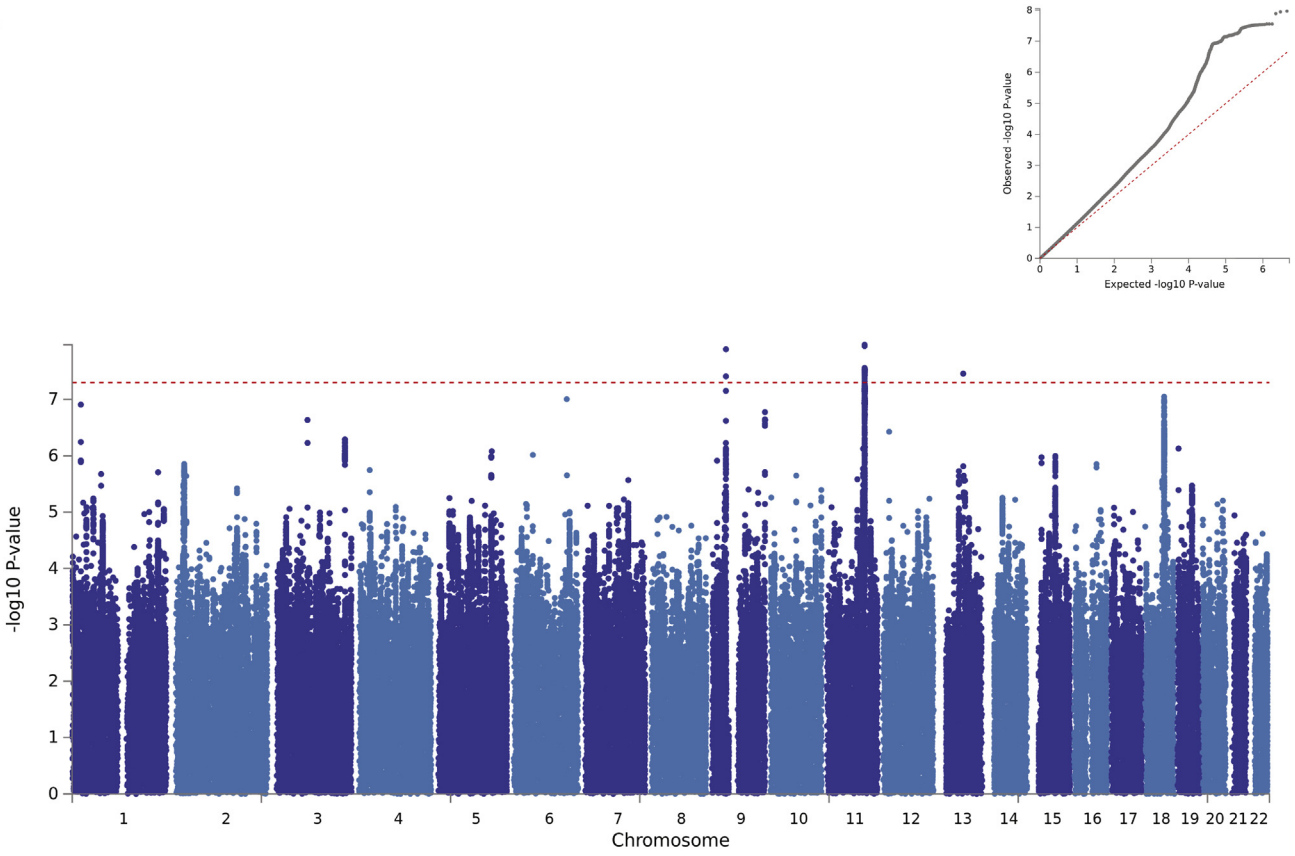
3.2. Primary ordinal GWAS of suicidality

The results of the ordinal GWAS of suicidality are presented in Table 1, Supplementary Table 2 and Fig. 1A. The GWAS results showed some inflation of the test statistics from the null ($\lambda_{GC} = 1{\cdot}16$, Fig. 1A, inset) but this was not significant given the sample size used (λ_{GC} $1000 = 1{\cdot}004$). LDSR demonstrates that polygenic architecture, rather than unconstrained population structure, is the likely reason for this (LDSR intercept = 1·02, SE = 0·0075). SNP heritability was estimated by LDSR as being 7·6% (observed scale $h^2_{SNP} = 0{\cdot}076$; SE = 0·006).

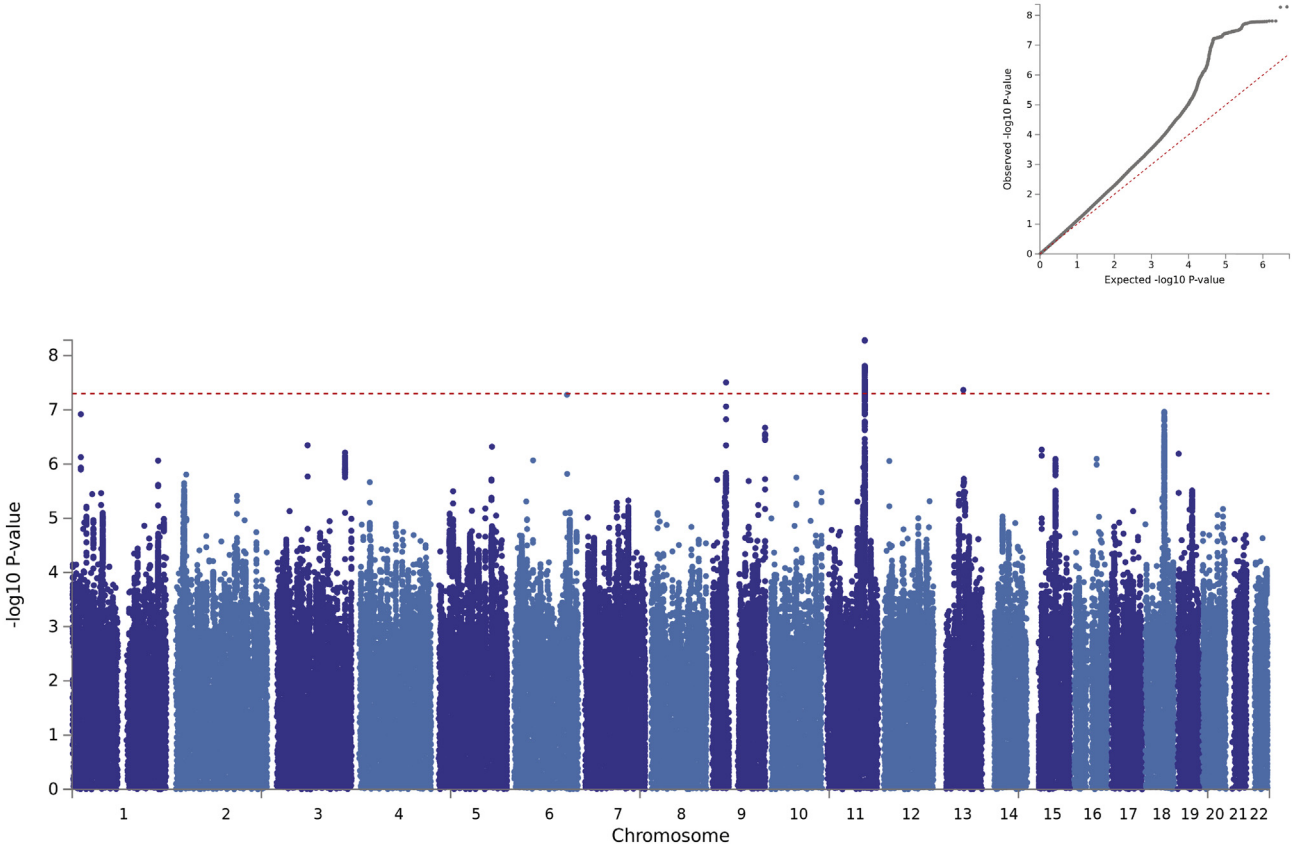
We identified three independent loci associated with suicidality (Table 1, Supplementary Table 2 and Fig. 1A): one on chromosome 9 (index SNP rs62535711, Fig. 2A) within the gene *ZCCHC7*; a second on chromosome 11 (index SNP rs598046, Fig. 2B) located within *CNTN5;* and a third on chromosome 13 (index SNP rs7989250, Fig. 2C). Conditional analyses (Supplementary Methods) in which the lead SNP was included as a covariate demonstrated no significant secondary association signals at these loci (the most significant additional SNP on chromosome 9 was rs999510, $p = 0{\cdot}0008$; that on chromosome 11 was rs608820, $p = 0{\cdot}0005$; and that on chromosome 13 was rs9564176, $p = 0{\cdot}003$). Effect allele frequencies by suicidality category are presented in Supplementary Table 3. Adjustment of the GWAS for psychiatric disorders had little or no effect on the observed associations (Fig. 1B and Supplementary Table 2), whilst adjustment for childhood sexual abuse rendered all associations null (Supplementary Fig. 3).

It is notable that within the recently-reported GWAS of suicide attempt in a Danish sample [26], rs62535711 and rs7989250 were not significant ($p = 0{\cdot}278$ and $p = 0{\cdot}152$ respectively) but rs598046 was reported as borderline significant (G allele, Beta 0·041, SE 0·021 $p = 0{\cdot}051$).

A)


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

B)


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

3.3. Association between genetic loading for suicidality and 'completed suicide'

Demographic characteristics of the controls in this analysis were comparable to those included in the GWAS (Supplementary Table 1). Individuals within the completed suicide group followed a similar pattern of increasing deprivation, more childhood trauma and higher prevalence of mood disorders observed across the categories of increasing suicidality (Supplementary Table 1). We investigated whether greater genetic loading for suicidality, indexed by PRS for suicidality, was associated with completed suicide. Overall, higher values of suicidality PRS were associated with an increased risk of completed suicide at all but one of the PRS significance thresholds assessed (for example, *p*-threshold 0·05: OR 1.23, 95%CI 1·06–1·41, FDR-adjusted $p = 0·04$, Table 2).

3.4. Genetic loading for suicidality and mood disorders

PRS for suicidality also demonstrated consistent significant associations with mood disorders (BD and MDD) and related traits (mood instability, neuroticism, and risk-taking propensity), across most of the significance thresholds assessed (Supplementary Table 4).

3.5. Secondary GWAS analyses

The ordinal GWAS of DSH (comprising controls, contemplated self-harm and actual self-harm) identified no SNPs at genome-wide significance (Supplementary Fig. 4A) and adjustment for psychiatric diagnosis had little impact on these results (Supplementary Fig. 4B): the most significant association was with rs4521702-T, Beta −0·01162, SE 0·0218, $p = 9·56 \times 10^{-8}$; and Beta −0·01159, SE 0·0218, $p = 1·10 \times 10^{-7}$, without and with adjustment for psychiatric diagnosis, respectively.

The ordinal GWAS of SIA (comprising controls, 'suicidal ideation', and 'suicide attempts') also identified no SNPs at genome-wide significance (Supplementary Fig. 5A), however adjustment for psychiatric disorders did identify a genome-wide significant singleton SNP rs116955121 (Supplementary Fig. 5B–C and Supplementary Table 2). In the recent Danish GWAS of suicide attempt [26], rs116955121 was not significant ($p = 0·509$).

3.6. Genetic correlation analyses

When considering the whole genome (rather than SNPs selected for association with suicidality, as is the case for the PRS), we observed significant genetic correlations between suicidality (primary analysis) and attempted suicide, and between suicidality and all of the major psychiatric disorders and traits we assessed (Table 3). The strongest genetic correlations were observed for MDD (r_g 0·81), Anxiety disorder (r_g 0.75), neuroticism (r_g 0·63) and mood instability (r_g 0·50). DSH demonstrated similar genetic correlations with attempted suicide and with psychiatric disorders and related traits as those observed for suicidality (Table 3). In contrast, for SIA, significant genetic correlations were observed only for MDD, schizophrenia, neuroticism and mood instability (Table 3).

3.7. Gene-based analysis

Gene-based analysis was used to identify genes containing potential composite association signals that were not identified by the individual SNP analysis, but which might nevertheless contribute to biological mechanisms underlying suicidality. Gene-based analysis highlighted *CNTN5, ADCK3/COQ8A*, *CEP57* and *FAM76B* and *DCC* for suicidality (primary analysis, Supplementary Fig. 6), *EIF4A1* and *SENP3* and *DCC* for DSH (Supplementary Fig. 7) and *CDKAL1*, *CNTN5*, and *ADCK3/COQ8A* for SIA (Supplementary Fig. 8). Regional plots for these genes (except for *CNTN5*, which was also identified in the SNP-based analysis) are presented in (Supplementary Fig. 9).

3.8. Biology of the suicidality-associated loci

Suggested functions of genes within the suicidality-associated loci are presented in Supplementary Table 5 and the Supplementary Results. Notable findings were the chromosome 11 locus located within *CNTN5*, a very large gene expressed predominantly in brain in adults (Fig. 2D); eQTL analysis supported the possible involvement of several candidate genes on chromosome 9 (Supplementary Fig. 10 and Supplementary Results), and, in combination with the gene-based analysis, additional candidate genes on chromosome 11 (*CEP57*; Supplementary Table 6) and on chromosome 18 (*DCC*). These genes and nearby variants have previously been associated with a variety of relevant traits (Supplementary Table 7). Of the SNPs with suggestive evidence for association (GWAS $p < 1 \times 10^{-5}$) with suicidal behaviour in other studies, 29 were available in our analysis and five of these demonstrated nominal ($p < 0·05$) association with suicidality in this study (Supplementary Table 8), although only one of these, rs72940689, had a direction of effect consistent with that given in the previous report [26].

4. Discussion

4.1. Main findings

Using a very large population-based cohort, we identified multiple genetic loci associated with suicidality. We also found that increased genetic burden for suicidality was associated with increased risk of completed suicide within a non-overlapping sub-sample and there were consistent genetic correlations with a wide range of psychiatric disorders and psychopathological traits, particularly MDD and anxiety disorders. Separate GWAS analyses of DSH and SIA identified one additional signal (for SIA) and suggested that the genetic architecture of DSH is likely to be distinct from that of SIA. Genetic correlations between mental illness and DSH or SIA also differed. More generally, the consistent effect size and direction of the association of rs598046-G (*CNTN5*) with suicide attempts in an independent cohort [26], and our nominal replication of a SNP reported by Erlangsen (rs7862648) is of interest [26].

In line with our evidence that suicidality is polygenic, inclusion of more SNPs (using more relaxed *p*-value thresholds) within a PRS typically demonstrates more significant effects in contrast to the stricter thresholds used because more information and power is provided by including a greater number of SNPs. Currently there is no agreed threshold that should be considered in these analyses, therefore we reported several PRS analyses. Future work on PRSs for suicidality should seek to identify thresholds that optimally facilitate stratification of clinical and non-clinical populations.

4.2. Comparisons with previous studies

Direct comparison with previous studies is hindered by the radically different study design reported here. Nonetheless, we have tried to align our findings with those previously reported.

This is the largest GWAS of suicidality to date and the first to consider a broad spectrum of suicidal behaviours. The loci previously reported for suicidal behaviours do not overlap with those identified here [14–26]. However, consistent (albeit borderline or nominal)

Fig. 1. Manhattan plot of GWAS of ordinal suicidality in UK Biobank (N = 122,935): A) adjusted for age, sex, genotyping chip and population structure, B) adjusted for age, sex, genotyping chip, population structure and psychiatric disorders. Dashed red line = genome wide significance threshold. Inset: QQ plot for genome-wide association with ordinal suicidality. Red line = theoretical distribution under the null hypothesis of no association.

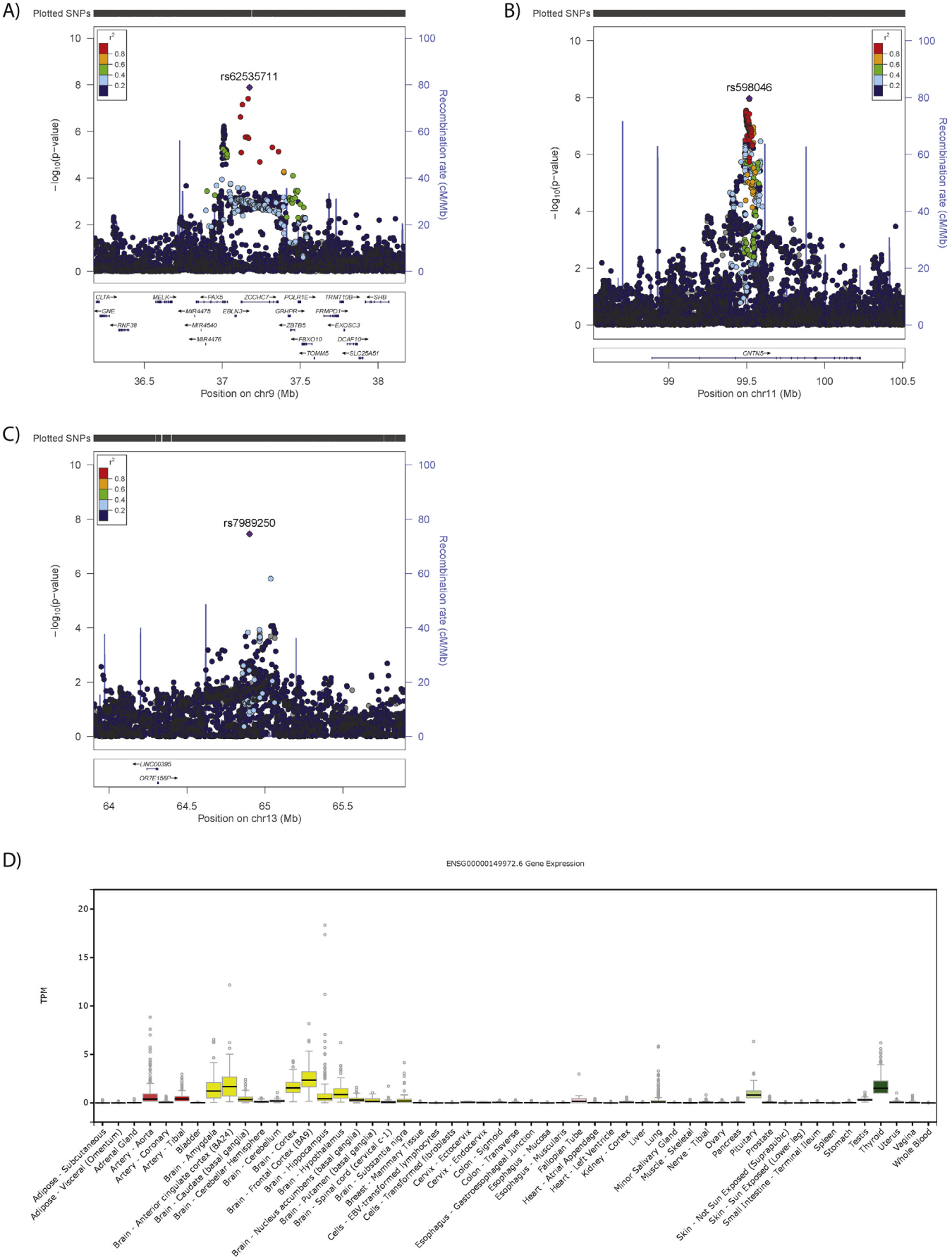

A)
Plotted SNPs
rs62535711
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr9 (Mb)
B)
Plotted SNPs
rs598046
Position on chr11 (Mb)
C)
Plotted SNPs
rs7989250
Position on chr13 (Mb)
D)
ENSG00000149972.6 Gene Expression
TPM

Table 1
Lead SNPs at loci associated with ordinal suicidality at GWAS significance.

Analysis	SNP	CHR	POS	A1	A2	BETA	SE	*P*	A1F[a]	SNPs_gwas[b]	SNPs_sugg[b]
Ordinal suicidality	rs62535711	9	37,174,829	T	C	0.105	0.018	1.29E-08	0.056	15	57
	rs598046	11	99,516,468	T	G	0.053	0.009	1.07E-08	0.319	34	370
	rs7989250	13	64,900,801	A	C	−0.052	0.009	3.49E-08	0.322	1	5

Where: A1, effect allele; A2, other allele; A1F, effect allele frequency; Aligned to Human Genome assembly GRCh37, Chr9 locus, 9:36999369–37,360,767; Chr11 locus, 11:99392678–99,588,751; Chr13 locus, 13:64900801–65,036; Chr6 locus, 6:140442326–140,895,470; SIA, suicidal ideation or attempt.
[a] Calculated in whole cohort.
[b] Within the region defined by suggestive significance ($P < 1 \times 10^{-5}$).

associations of one of our lead SNPs in the recent large Danish study of suicide attempts [26] and one of their lead SNPs in this study is compelling.

Most previous genetic studies of suicidal behaviour and completed suicide (Supplementary Table 8) have been conducted in cohorts with known diagnoses of major mental illness, thereby controlling for mental illness. In this study, sensitivity analysis controlling (by adjustment) for mental illness had negligible effects on the primary suicidality results, however an additional signal was identified for SIA in secondary analyses. It is likely that the various aspects of suicidality have complex relationships with different mental illnesses. Therefore, we cannot exclude the possibility that the results might have been driven by the mixed and different genetic loadings for psychiatric disorders between suicidal and non-suicidal participants. A further sensitivity analysis adjusting for mental illness and childhood sexual abuse rendered all associations null, however as childhood sexual abuse is strongly associated with mental illness, this model could be considered overly conservative.

In addition, the limited overlap between suicidality-associated loci identified across these studies also likely reflects substantial differences in recruitment protocols, participant characteristics (including the mix of psychiatric diagnoses) and variation in the methods of assessment of suicidal behaviour. Recently, a general population study in Denmark of suicide attempts identified genetic loci at genome-wide significance [26]. Our study extends this approach by investigating a broader phenotype within our primary analysis, as well as the specific impact of DSH versus SIA in secondary analyses. In line with a Research Domain Criteria (RDoC) approach, we used the full spectrum of suicidal thoughts and behaviours assessed within a predominantly non-clinical population. The fact that increased genetic burden for suicidality was associated with increased risk of completed suicide in a separate subsample represents an important validation of our suicidality phenotype. The genetic correlation with MDD was strong, but the incomplete overlap, and the fact that the GWAS results were largely unchanged by adjustment for mental health disorder status, supports the hypothesis that at least some of the genetic predisposition to suicidal ideation and behaviours may be distinct from that for MDD [18].

The SNP-based heritability reported here for suicidality (7.6%) is more than that reported for suicide attempts (4.6%) [26], however both of these are lower than heritability estimates from family studies [12], which is consistent with findings from most other complex traits studied to date.

4.3. Biology

The known biology of the suicidality-associated loci (Supplementary Results) highlights three interesting candidate genes: *CNTN5*, *CEP57* and *DCC*. *CNTN5* encodes contactin 5 (also known as NB-2), which is a good functional candidate. CNTN5 is a glycosylphosphatidylinositol (GPI)-anchored extracellular cell adhesion protein of the immunoglobulin superfamily, thought to have a role in the formation and maintenance of brain circuitry [39]. Centrosomal protein of 57 kDa (CEP57, encoded by *CEP57*) is important for cell division, with loss of function variants causing a mosaic variegated aneuploidy syndrome, which can include brain abnormalities and mental retardation (OMIM #607951 and #614114). The netrin 1 receptor (encoded by *DCC*) has been robustly associated with depression [40], schizophrenia [41] and related traits [42]. Speculation as to how variation in these genes act to influence these related traits is difficult because of incomplete understanding of the functions of these genes in the brain during development and aging.

4.4. Strengths and limitations

This is the largest genetic study of suicidality in a population sample reported to date but we acknowledge some limitations to this work. The nature of the data collected did not allow us to distinguish between putative subtypes of suicide, such as stress-responsive and non-stress-responsive suicidality [43]. In addition, both recruitment bias and recall bias are possible within the UK Biobank dataset. Survivor bias might also influence our findings, due to the relatively older age at recruitment, however this would likely lead to more conservative effect estimates. We also recognise that the questions used to create the ordinal suicidality phenotype mean that some individuals with either self-harm ideation/behaviour or suicidal ideation might be included at levels higher or lower than their suicidality predisposition merits. Our SNP heritability estimates were similar to those reported for other complex psychiatric phenotypes such as MDD [44] and the overlap with clinically relevant phenotypes (at the levels of loci, PRS and whole-genome genetic correlations) all suggest that our findings are robust.

Table 2
Increasing burden of suicidal behaviour-associated variants significantly associated with completed suicide.

Threshold	OR	L95	U95	*P*	FDR-adj *P*
5×10^{-8}	1.07	−0.10	0.24	0.410	0.410
5×10^{-5}	1.20	0.01	0.35	0.041	0.049
0.01	1.22	0.02	0.38	0.026	0.390
0.05	1.26	0.05	0.41	0.011	0.034
0.1	1.27	0.06	0.42	0.008	0.034
0.5	1.25	0.03	0.41	0.021	0.039

Where: Threshold, GWAS *P* threshold of the SNPs included in the suicidality PRS; OR, odds ratio; L95, lower 95% confidence interval; U95, upper 95% confidence interval; Z, test statistic, *P*, *P* value for analysis; FDR-adj *P*, false discovery rate adjusted *P*. Signficance was set at $P < 0.05$.

Fig. 2. Regional plots for GWAS significant loci and *CNTN5* tissue expression: A) *ZCCHC7* locus on Chr9, B) *CNTN5* locus on Chr11, C) Chr13 locus, where: SNPs (each point) are aligned according to position (X axis) and strength of association (Y Axis, left); Purple colouring indicates the index SNP, with other colours representing linkage disequilibrium (r^2) with the index SNP, as per the colour key; Recombination rate is presented as a pale blue line graph in the background (Y axis, right); Genes are presented below the association plot by location (X axis) and direction of transcription (arrows). D) Tissue expression profile of *CNTN5*, where tissues are arranged alphabetically along the X-axis and expression level is (TPM; standardised transcripts per million reads) provided on the Y-axis. Box plots represent median and interquartile range, with error bars demonstrating 1.5× the interquartile range and dots representing outliers.

Table 3
Genetic correlations of suicidality with psychiatric disorders and related traits.

Trait	Suicidality					DSH					SIA				
	r_g	se	z	p	FDR-P	r_g	se	z	p	FDR-P	r_g	se	z	p	FDR-P
Attempted suicide	0.57	0.096	5.98	2.23E-09	6.24E-09	0.624	0.106	5.8897	3.87E-09	9.03E-09	−0.012	0.182	−0.068	9.46E-01	9.83E-01
MDD	0.81	0.04	18.66	1.01E-77	1.41E-76	0.79	0.05	14.48	1.68E-47	2.35E-46	0.46	0.11	4.07	4.76E-05	2.22E-04
Neuroticism	0.63	0.04	16.48	5.51E-61	3.86E-60	0.57	0.05	12.09	1.26E-33	8.82E-33	0.56	0.12	4.77	1.80E-06	1.26E-05
Mood Instability	0.50	0.03	16.06	4.53E-58	2.11E-57	0.43	0.04	11.73	8.50E-32	3.97E-31	0.53	0.11	5.01	5.42E-07	7.59E-06
Schizophrenia	0.32	0.04	8.70	3.19E-18	1.12E-17	0.31	0.04	7.02	2.25E-12	7.88E-12	0.28	0.09	2.96	3.06E-03	1.07E-02
Bipolar disorder	0.27	0.05	5.28	1.26E-07	2.94E-07	0.27	0.06	4.34	1.45E-05	2.26E-05	0.19	0.12	1.55	1.21E-01	3.39E-01
Risk-taking behaviour	0.20	0.04	5.05	4.44E-07	7.77E-07	0.25	0.04	6.00	1.93E-09	5.40E-09	0.00	0.08	0.02	9.83E-01	9.83E-01
Anxiety disorder	0.75	0.17	4.35	1.38E-05	2.15E-05	0.87	0.20	4.41	1.03E-05	1.91E-05	0.31	0.30	1.03	3.05E-01	5.34E-01
ADHD	0.21	0.05	4.29	1.79E-05	2.51E-05	0.23	0.05	4.40	1.09E-05	1.91E-05	−0.12	0.11	−1.17	2.44E-01	5.20E-01
PTSD	0.42	0.16	2.70	6.97E-03	8.87E-03	0.52	0.18	2.82	4.82E-03	6.75E-03	0.16	0.28	0.57	5.70E-01	7.98E-01

Where: DSH, deliberate self-harm; SIA, suicidal ideation or attempt; rg, genetic correlation; se, standard error of genetic correlation; z, test statistic, P, P vlaue for analysis; FDR adj P, false discovery rate-adjusted P; Signficicance was set at P ≤ 0.05; MDD, major depressive disorder; PTSD, post-traumatic stress disorder; Relative amplitude, quantitative measure of circadian rythmicity; ADHD, attention deficit hyperactivity disorder.

4.5. Implications for future work

This study highlights a component of suicidal predisposition that is distinct from MDD predisposition and the potential relevance of *CNTN5*, *CEP57* and *DCC* to suicidality, further study of which may provide valuable insight into the underlying biology of suicide. Genetic vulnerability to suicide is of course likely to be only a small part of the overall pathophysiology of what is clearly a highly complex and clinically and psychologically heterogeneous phenotype. A major current challenge for the field of suicide research is to integrate new discoveries on the genetics of suicide with known psychiatric, social, psychological and environmental risk factors (such as poverty, substance misuse and childhood sexual abuse), to develop more sophisticated models of risk, and ultimately to develop genetically-informed social, psychological and public health interventions.

5. Conclusions

In the largest GWAS to date of suicidality to date we identified several new candidate genes that may be relevant to the biology of completed suicide. We also found substantial genetic correlation between suicidality and a range of psychiatric disorders and, by finding an association between genetic loading for suicidality and completed suicide, we provide preliminary evidence for the potential utility of PRSs for patient and population stratification. We hope these discoveries will facilitate new avenues of research on this complex but clinically important phenotype.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.02.005.

Acknowledgements

We thank all participants in the UK Biobank study. UK Biobank was established by the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government and Northwest Regional Development Agency. UK Biobank has also had funding from the Welsh Assembly Government and the British Heart Foundation. Data collection was funded by UK Biobank. RJS is supported by a UKRI Innovation- HDR-UK Fellowship (MR/S003061/1). 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). KJAJ is supported by an MRC Doctoral Training Programme Studentship at the Universities of Glasgow and Edinburgh. CLN is supported by a UKRI Innovation Fellowship (MR/R024774/1). DJS acknowledges the support of the Brain and Behavior Research Foundation (Independent Investigator Award 1930), a Lister Prize Fellowship (173096) and the MRC Mental Health Data Pathfinder Award (MC_PC_17217).

The funders had no role in the design or analysis of this study, decision to publish, or preparation of the manuscript.

Author contributions

Study concept and design: Smith, Strawbridge, Ward, Graham, Shaw, Cullen.

Acquisition, analysis, or interpretation of data: Strawbridge, Ward, L. Lyall, Niedzwiedz, Langan Martin, D.Lyall, Bailey and Smith.

Drafting of the manuscript: Strawbridge, Pearsall, Bailey and Smith.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Strawbridge, Ward, Graham, Shaw and Ferguson.

Administrative, technical, or material support: Pell and Smith.

Study supervision: Smith.

Conflict of interest

The authors have no conflicts of interest.

References

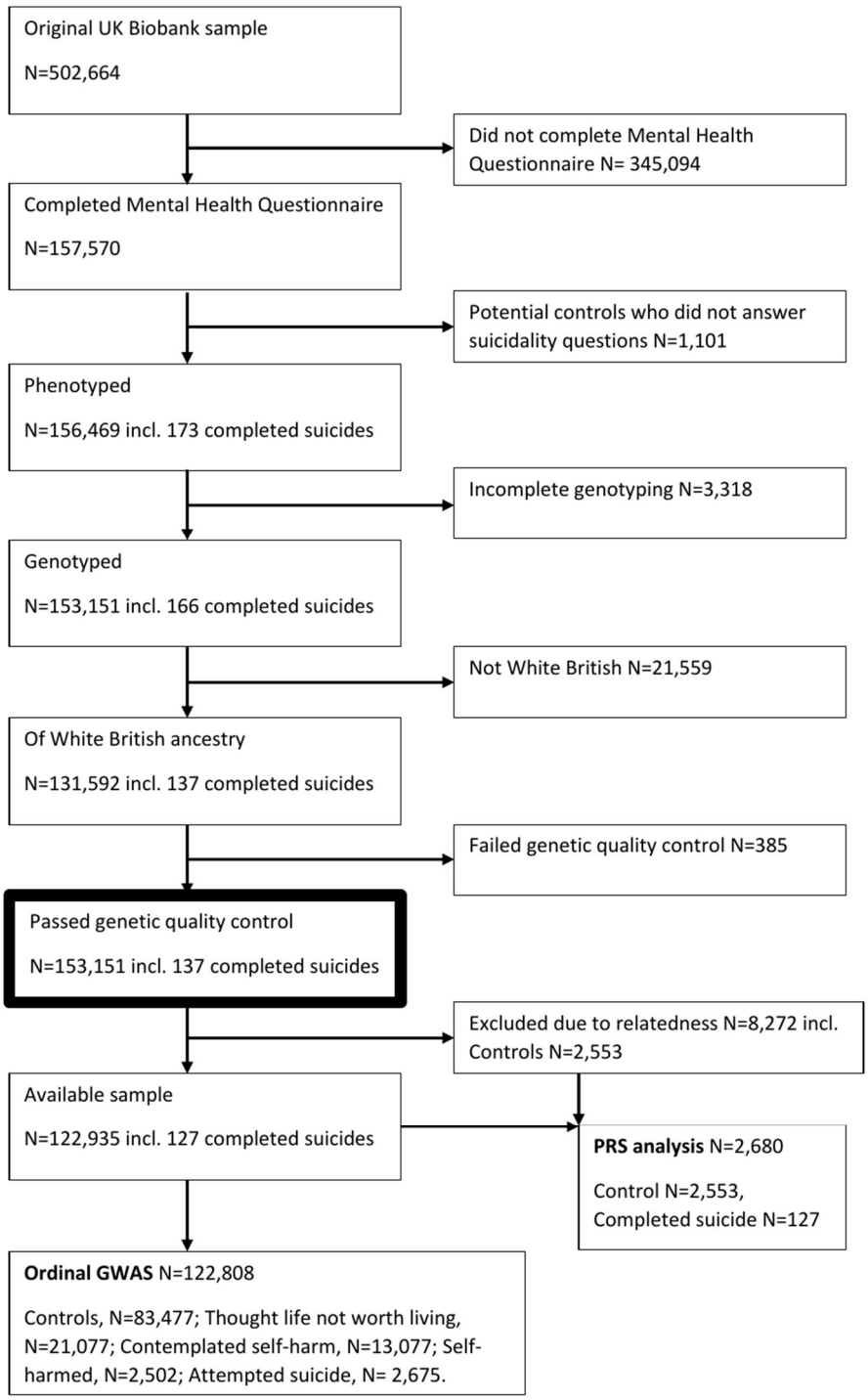
[1] WHO. Summary tables of mortality estimates by cause, age and sex, globally and by region, 2000–2015; 2017.
[2] Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E, Sarchiapone M, et al. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry 2016;3(7):646–59.
[3] Turecki G. The molecular bases of the suicidal brain. Nat Rev Neurosci 2014;15(12):802–16.
[4] Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, et al. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. Br J Psychiatry 2008;192(2):98–105.
[5] Yoshimasu K, Kiyohara C, Miyashita K, Stress Research Group of the Japanese Society for H. Suicidal risk factors and completed suicide: meta-analyses based on psychological autopsy studies. Environ Health Prev Med 2008;13(5):243–56.
[6] Mann JJ, Apter A, Bertolote J, Beautrais A, Currier D, Haas A, et al. Suicide prevention strategies: a systematic review. JAMA 2005;294(16):2064–74.
[7] Hawton K, Saunders KE, O'Connor RC. Self-harm and suicide in adolescents. Lancet 2012;379(9834):2373–82.
[8] Berg L, Rostila M, Hjern A. Parental death during childhood and depression in young adults - a national cohort study. J Child Psychol Psychiatry 2016;57(9):1092–8.
[9] Arsenault-Lapierre G, Kim C, Turecki G. Psychiatric diagnoses in 3275 suicides: a meta-analysis. BMC Psychiatry 2004;4:37.
[10] Kendler KS. Genetic and environmental pathways to suicidal behavior: reflections of a genetic epidemiologist. Eur Psychiatry 2010;25(5):300–3.
[11] Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 2017;143(2):187–232.
[12] Brent DA, Melhem N. Familial transmission of suicidal behavior. Psychiatr Clin North Am 2008;31(2):157–77.
[13] Brent DA, Bridge J, Johnson BA, Connolly J. Suicidal behavior runs in families. A controlled family study of adolescent suicide victims. Arch Gen Psychiatry 1996;53(12):1145–52.
[14] Galfalvy H, Haghighi F, Hodgkinson C, Goldman D, Oquendo MA, Burke A, et al. A genome-wide association study of suicidal behavior. Am J Med Genet B Neuropsychiatr Genet 2015;168(7):557–63.

[15] Kimbrel NA, Garrett ME, Dennis MF, Va Mid-Atlantic Mental Illness Research E, Clinical Center W, Hauser MA, et al. A genome-wide association study of suicide attempts and suicidal ideation in U.S. military veterans. Psychiatry Res 2018;269:64–9.
[16] Levey DF, Polimanti R, Cheng Z, Zhou H, Nunez YZ, Jain S, et al. Genetic associations with suicide attempt severity and genetic overlap with major depression. Transl Psychiatry 2019;9(1):22.
[17] Li J, Yoshikawa A, Meltzer HY. Replication of rs300774, a genetic biomarker near ACP1, associated with suicide attempts in patients with schizophrenia: Relation to brain cholesterol biosynthesis. J Psychiatr Res 2017;94:54–61.
[18] Mullins N, Perroud N, Uher R, Butler AW, Cohen-Woods S, Rivera M, et al. Genetic relationships between suicide attempts, suicidal ideation and major psychiatric disorders: a genome-wide association and polygenic scoring study. Am J Med Genet B Neuropsychiatr Genet 2014;165B(5):428–37.
[19] Perlis RH, Huang J, Purcell S, Fava M, Rush AJ, Sullivan PF, et al. Genome-wide association study of suicide attempts in mood disorder patients. Am J Psychiatry 2010;167(12):1499–507.
[20] Perroud N, Uher R, Ng MY, Guipponi M, Hauser J, Henigsberg N, et al. Genome-wide association study of increasing suicidal ideation during antidepressant treatment in the GENDEP project. Pharmacogenomics J 2012;12(1):68–77.
[21] Ruderfer DM, Fanous AH, Ripke S, McQuillin A, Amdur RL, Schizophrenia Working Group of Psychiatric Genomics C, et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Mol Psychiatry 2014;19(9):1017–24.
[22] Stein MB, Ware EB, Mitchell C, Chen CY, Borja S, Cai T, et al. Genomewide association studies of suicide attempts in US soldiers. Am J Med Genet B Neuropsychiatr Genet 2017;174(8):786–97.
[23] Willour VL, Seifuddin F, Mahon PB, Jancic D, Pirooznia M, Steele J, et al. A genome-wide association study of attempted suicide. Mol Psychiatry 2012;17(4):433–44.
[24] Zai CC, Goncalves VF, Tiwari AK, Gagliano SA, Hosang G, de Luca V, et al. A genome-wide association study of suicide severity scores in bipolar disorder. J Psychiatr Res 2015;65:23–9.
[25] Coon H, Darlington TM, DiBlasi E, Callor WB, Ferris E, Fraser A, et al. Genome-wide significant regions in 43 Utah high-risk families implicate multiple genes involved in risk for completed suicide. Mol Psychiatry 2018. https://doi.org/10.1038/s41380-018-0282-3.
[26] Erlangsen A, Appadurai V, Wang Y, Turecki G, Mors O, Werge T, et al. Genetics of suicide attempts in individuals with and without mental disorders: a population-based genome-wide association study. Mol Psychiatry 2018. https://doi.org/10.1038/s41380-018-0218-y.
[27] Ruderfer DM, Walsh CG, Aguirre MW, Tanigawa Y, Ribeiro JD, Franklin JC, et al. Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. Mol Psychiatry 2019. https://doi.org/10.1038/s41380-018-0326-8.
[28] Biobank UK. Genotype imputation and genetic association studies of UK Biobank, Interim data release. 11 September 2015 http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015.pdf; 2015.
[29] Biobank UK. Genotyping of 500,000 UK Biobank participants. Description of sample processing workflow and preparation of DNA for genotyping. 11 September 2015 https://biobank.ctsu.ox.ac.uk/crystal/docs/genotyping_sample_workflow.pdf; 2015.
[30] Davis KAS, Coleman JRI, Adams M, Allen N, Breen G, Cullen B, et al. Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. BJPsych Open 2018;4(3):83–90.
[31] Christensen RHB. Ordinal - regression models for ordinal data R package version ; 2015 [6-28 ed2015].
[32] Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. PLoS One 2013;8(11):e75362.
[33] Ng QX, Yong BZJ, Ho CYX, Lim DY, Yeo WS. Early life sexual abuse is associated with increased suicide attempts: an update meta-analysis. J Psychiatr Res 2018;99:129–41.
[34] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8(1):1826.
[35] 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):1236–41.
[36] de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11(4):e1004219.
[37] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biol 2016;17(1):122.
[38] Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45(6):580–5.
[39] Oguro-Ando A, Zuko A, Kleijer KTE, Burbach JPH. A current view on contactin-4, −5, and −6: implications in neurodevelopmental disorders. Mol Cell Neurosci 2017;81:72–83.
[40] Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat Genet 2016;48(6):624–33.
[41] Goes FS, McGrath J, Avramopoulos D, Wolyniec P, Pirooznia M, Ruczinski I, et al. Genome-wide association study of schizophrenia in Ashkenazi Jews. Am J Med Genet B Neuropsychiatr Genet 2015;168(8):649–59.
[42] 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. Transl Psychiatry 2017;7(11):1264.
[43] Bernanke JA, Stanley BH, Oquendo MA. Toward fine-grained phenotyping of suicidal behavior: the role of suicidal subtypes. Mol Psychiatry 2017;22(8):1080–1.
[44] 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):668–81.

4.4.1 Appendix D

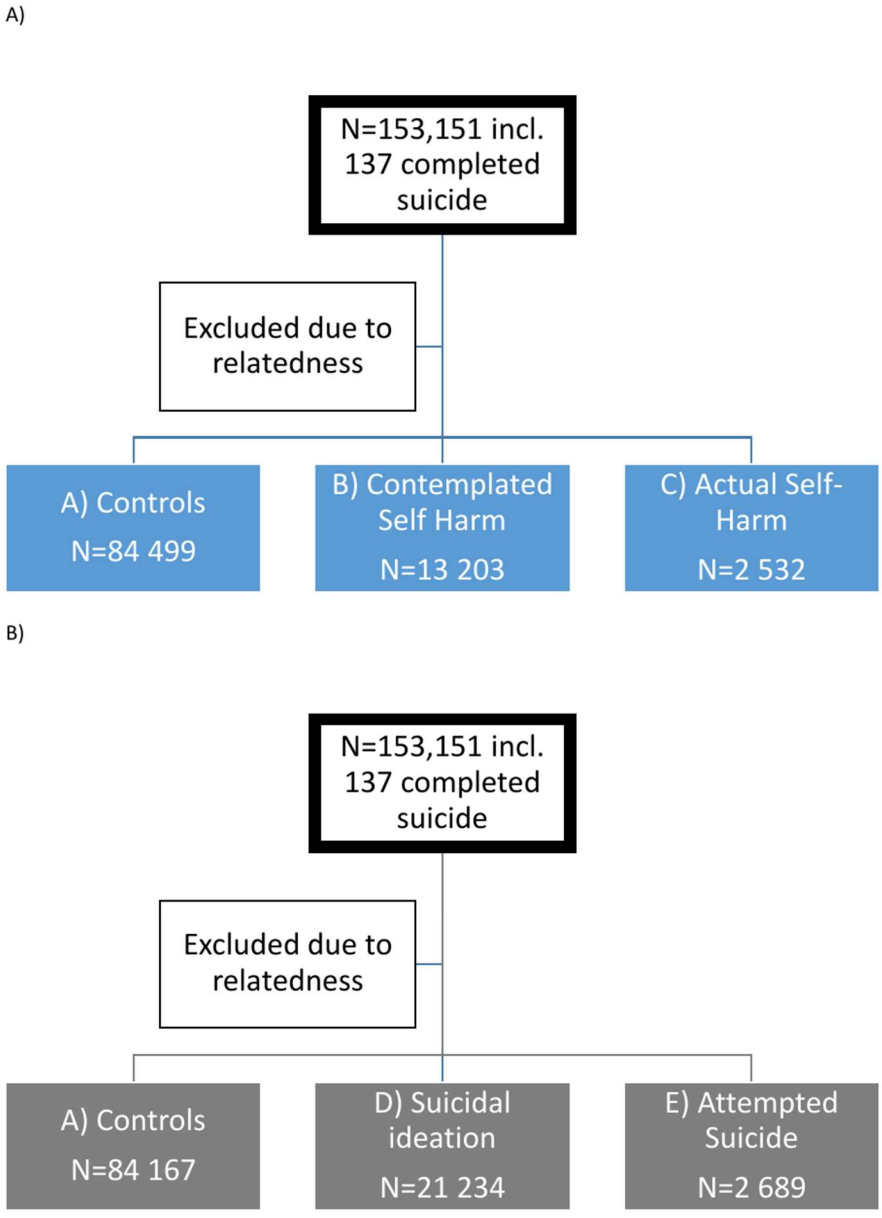
4.4.1.1 Supplementary Figure 1

Flow chart of UK Biobank participants available for primary analyses (Ordinal GWAS and PRS analysis)



4.4.1.2 Supplementary Figure 2

Flow chart of UK Biobank participants available for secondary analyses. The flow chart of participants is the same as Supplementary Fig. 1 up to the highlighted box. Relatedness exclusions were applied for A) the DSH GWAS considering the categories Controls, Contemplated self-harm and Actual self-ham and B) the SIA GWAS considering the categories Controls, Suicidal ideation and attempted suicide.



4.4.1.3 Supplementary Table 1

Cohort demographics of individuals included in the Ordinal suicidality GWAS and the completed suicide PRS analysis

Analysis	Ordinal Suicidality GWAS										Completed Suicide PRS			
Suicidality category	0: No Reported Suicidality		1: Thought life not worth living		2: Contemplated Self harm		4: Deliberately Self harmed		5: Attempted Suicide		0: No Reported Suicidality		6: Completed Suicide[a]	
	(n=83,557)		(n=21,063)		(n=13,038)		(n=2,498)		(n=2,666)		(n=5,330)		(n=137)	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Sex														
Female	44,057	52.7	12,781	60.7	8,299	63.7	1,757	70.3	1,804	67.7	2,926	54.9	37	27
Male	39,500	47.3	8,282	39.3	4,739	36.4	741	29.7	862	32.3	1,174	48.9	100	73
Age range														
35/44	6,887	8.2	2,171	10.3	1,802	13.8	527	21.1	421	15.8	418	7.8	20	14.6
45/54	22,431	26.9	6,900	32.8	5,006	38.4	1,080	43.2	1,039	39	1,400	26.3	52	38
55/64	40,002	47.9	9,326	44.3	5,158	39.6	750	30	1,038	38.9	2,590	48.6	50	36.5
65/74	14,237	17	2,666	12.7	1,072	8.2	141	5.6	168	6.3	922	17.3	15	11
Living arrangements														
Alone	12,456	14.9	4,318	20.6	2,850	21.9	531	21.3	743	28	785	14.8	43	31.6
With partner	66,243	79.5	14,894	70.9	8,865	68.2	1,662	66.8	1,573	59.3	4,213	79.2	76	55.9
Other	4,678	5.6	1,788	8.5	1,282	9.9	297	11.9	338	12.7	321	6	17	12.5
Area Deprivation														
Least deprived	21,116	25.3	4,667	22.2	2,701	20.8	452	18.1	392	14.7	1,293	24.3	20	14.6
4th Quintile	19,608	23.5	4,414	21	2,546	19.6	442	17.7	416	15.6	1,279	24	26	19
3rd Quintile	17,829	21.4	4,464	21.2	2,626	20.2	503	20.2	509	19.1	1,154	21.7	19	13.9
2nd Quintile	14,937	17.9	4,181	19.9	2,785	21.4	585	23.4	636	23.9	982	18.4	34	24.8
Most deprived	9,972	12	3,310	15.7	2,360	18.1	514	20.6	708	26.6	618	11.6	38	27.7
Parental Depression														
Neither parent	69,837	91.5	16,505	87.2	9,754	83	1,769	79.3	1,761	77.3	4,518	92.1	89	77.4
At least one parent	6,486	8.5	2,434	12.9	2,001	17	461	20.7	516	22.7	387	7.9	26	22.6
Chronic Pain														
Free of pain	53,529	64.1	11,658	55.4	6,637	50.9	1,234	49.4	1,157	43.4	3,417	64.2	83	60.6
One or more sites	29,983	35.9	9,389	44.6	6,398	49.1	1,262	50.6	1,508	56.6	1,907	35.8	54	39.4
Smoking														
Never	49,886	59.8	11,868	56.5	6,932	53.3	1,204	48.3	1,103	41.5	3,201	60.2	63	46
Previous	28,650	34.4	7,467	35.5	4,881	37.5	973	39	1,128	42.4	1,839	34.6	42	30.7
Current	4,861	5.8	1,686	8	1,203	9.2	317	12.7	429	16.1	282	5.3	32	23.4
Alcohol use														
Daily/ almost daily	20,319	24.3	5,002	23.8	2,874	22.1	541	21.7	562	21.1	1,271	23.9	37	23.9

Regular	44,393	53.1	10,536	50	6,330	48.6	1,196	47.9	1,133	42.6	2,847	53.4	71	53.4
Occasional	15,005	18	4,397	20.9	3,014	23.1	577	23.1	694	26.1	947	17.8	19	17.7
Former	1,764	2.1	633	3	514	4	133	5.3	207	7.8	110	2.1	8	2.2
Never	2,053	2.5	486	2.3	297	2.3	50	2	65	2.4	154	2.9	2	2.9
Life-time MDD														
No	62,153	85.1	7,952	49.9	3,078	30.3	603	30.4	317	14.9	4,045	86.3	na	na
Yes	10,876	14.9	7,996	50.1	7,088	69.7	1,380	69.6	1,818	85.2	640	13.7	na	na
Childhood trauma														
None	50,494	61	9,474	45.4	4,751	36.7	760	30.6	639	24.1	3,287	62.3	na	na
Any	32,258	39	11,403	54.6	8,186	63.3	1,722	69.4	2,009	75.9	1,992	37.7	na	na

4.4.1.4 Supplementary Table 2

All SNPs associated with ordinal suicidality at GWAS significance

								basic		adjusted for psychiatric disorders	
Analysis	SNP	CHR	BP	A1	A2	A1F[a]	BETA	P		BETA	P
Ordinal suicidality	rs62535709	9	37166189	T	C	0.05	0.101	3.89E-08		0.099	8.67E-08
	rs62535711	9	37174829	T	C	0.06	0.105	1.29E-08		0.103	3.12E-08
	rs635715	11	99495141	G	T	0.33	0.051	3.48E-08		0.052	1.97E-08
	rs639161	11	99495147	G	A	0.33	0.05	4.30E-08		0.051	2.51E-08
	rs637006	11	99495383	A	G	0.33	0.051	3.37E-08		0.052	1.88E-08
	rs637745	11	99495452	C	G	0.33	0.051	2.99E-08		0.052	1.67E-08
	rs637387	11	99495489	T	C	0.33	0.051	2.93E-08		0.052	1.64E-08
	rs1790263	11	99495583	T	G	0.33	0.051	3.10E-08		0.052	1.73E-08
	rs1790262	11	99495651	T	A	0.33	0.051	2.99E-08		0.052	1.67E-08
	rs636504	11	99495693	G	A	0.33	0.05	4.13E-08		0.051	2.32E-08
	rs636073	11	99495789	C	T	0.33	0.051	3.24E-08		0.052	1.80E-08
	rs651048	11	99496234	C	T	0.33	0.051	3.44E-08		0.052	1.92E-08
	rs651446	11	99496269	A	G	0.33	0.051	3.56E-08		0.052	2.00E-08
	rs651447	11	99496271	C	T	0.33	0.051	3.25E-08		0.052	1.87E-08
	rs651496	11	99496308	A	G	0.33	0.051	3.61E-08		0.052	2.02E-08
	rs1790261	11	99496359	C	T	0.33	0.051	3.32E-08		0.052	1.86E-08
	rs1690819	11	99496535	A	G	0.33	0.051	3.01E-08		0.052	1.67E-08
	rs1690818	11	99496554	C	T	0.33	0.051	2.78E-08		0.052	1.55E-08
	rs1690817	11	99496812	A	G	0.33	0.051	2.78E-08		0.052	1.55E-08
	rs1690816	11	99496849	C	T	0.33	0.051	2.78E-08		0.052	1.55E-08
	rs1003122	11	99496892	A	G	0.33	0.051	2.88E-08		0.052	1.60E-08
	rs1690814	11	99497012	A	G	0.33	0.051	2.91E-08		0.052	1.61E-08
	rs1790260	11	99497028	C	A	0.33	0.051	3.05E-08		0.052	1.69E-08
	rs609750	11	99497099	G	T	0.33	0.051	2.98E-08		0.052	1.66E-08
	rs646067	11	99497122	T	C	0.33	0.051	3.18E-08		0.052	1.76E-08
	rs646017	11	99497150	T	G	0.33	0.051	3.09E-08		0.052	1.72E-08
	rs634379	11	99497472	C	T	0.33	0.051	2.94E-08		0.052	1.64E-08
	rs634011	11	99497529	C	T	0.33	0.051	3.27E-08		0.052	1.81E-08
	11:99500407_TTAC_T	11	99500407	TTAC	T	0.33	0.051	3.89E-08		0.052	2.16E-08
	rs112677841	11	99507624	A	G	0.35	0.051	4.76E-08		0.052	2.81E-08
	rs111369127	11	99507634	T	C	0.35	0.051	4.76E-08		0.052	2.81E-08
	rs598046	11	99516468	T	G	0.32	0.053	1.07E-08		0.054	5.18E-09
	rs7122777	11	99516476	T	C	0.32	0.053	1.13E-08		0.054	5.31E-09
	rs61910885	11	99516507	C	G	0.33	0.051	3.72E-08		0.053	1.65E-08
	rs61910886	11	99516510	T	C	0.33	0.051	3.72E-08		0.053	1.65E-08
	rs61910887	11	99516511	G	A	0.33	0.051	3.72E-08		0.053	1.65E-08
	rs7989250	13	64900801	A	C	0.32	-0.052	3.49E-08		-0.051	4.31E-08
SIA	**rs116955121**	6	1.40E+08	A	G	0.02				0.199	1.66E-08

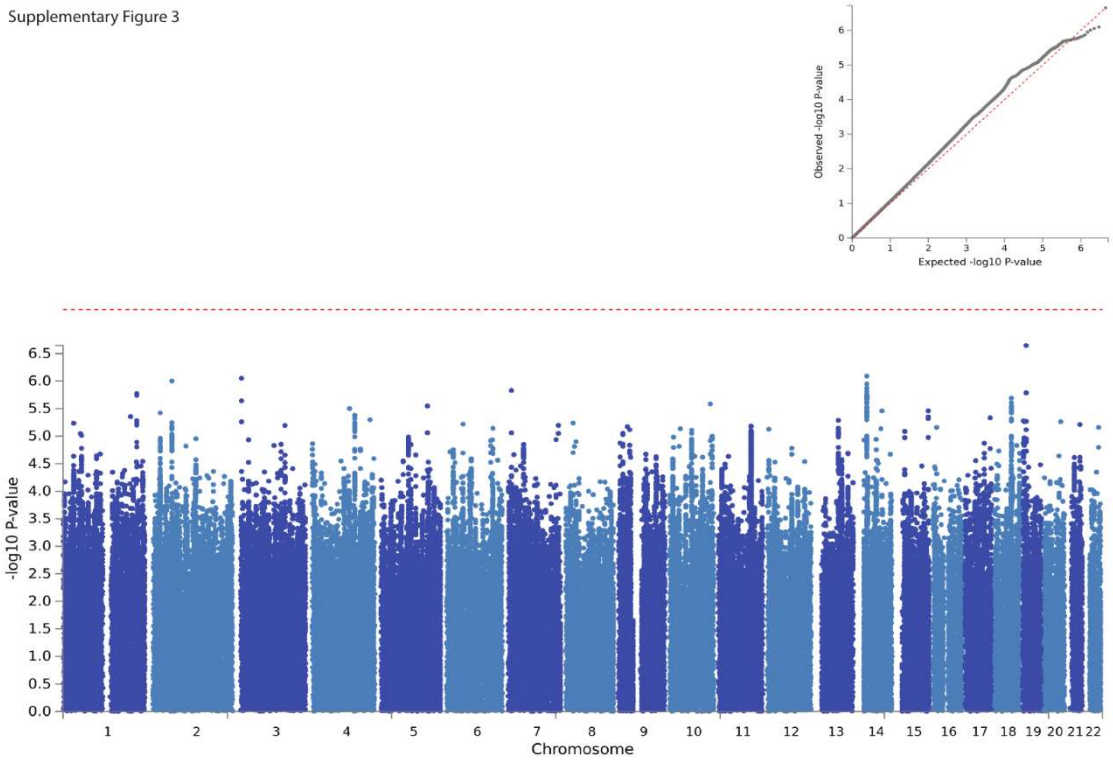
4.4.1.5 Supplementary Table 3

Allele frequencies of lead SNPs by suicidality category

				ordinal GWAS					PRS	
SNP	CHR	POS	A1	0: No Reported Suicidality	1: Thought life not worth living	2: Contemplated Self harm	4: Deliberately Self harmed	5: Attempted Suicide	0: No Reported Suicidality	6: Completed Suicide
rs62535711	9	37174829	T	0.05	0.06	0.06	0.06	0.06	0.06	0.05
rs598046	11	99516468	T	0.32	0.33	0.32	0.32	0.32	0.31	0.34
rs7989250	13	64900801	A	0.33	0.32	0.31	0.31	0.32	0.31	0.29

4.4.1.6 Supplementary Figure 3

Manhattan plot of GWAS of ordinal suicidality in UK Biobank (N = 100,234), adjusted for age, sex, genotyping chip, population structure, psychiatric disorders and childhood sexual abuse. Dashed red line = genome wide significance threshold ($p < 5 \times 10^{-5}$). Inset: QQ plot for genome-wide association with DSH. Red line = theoretical distribution under the null hypothesis of no association.



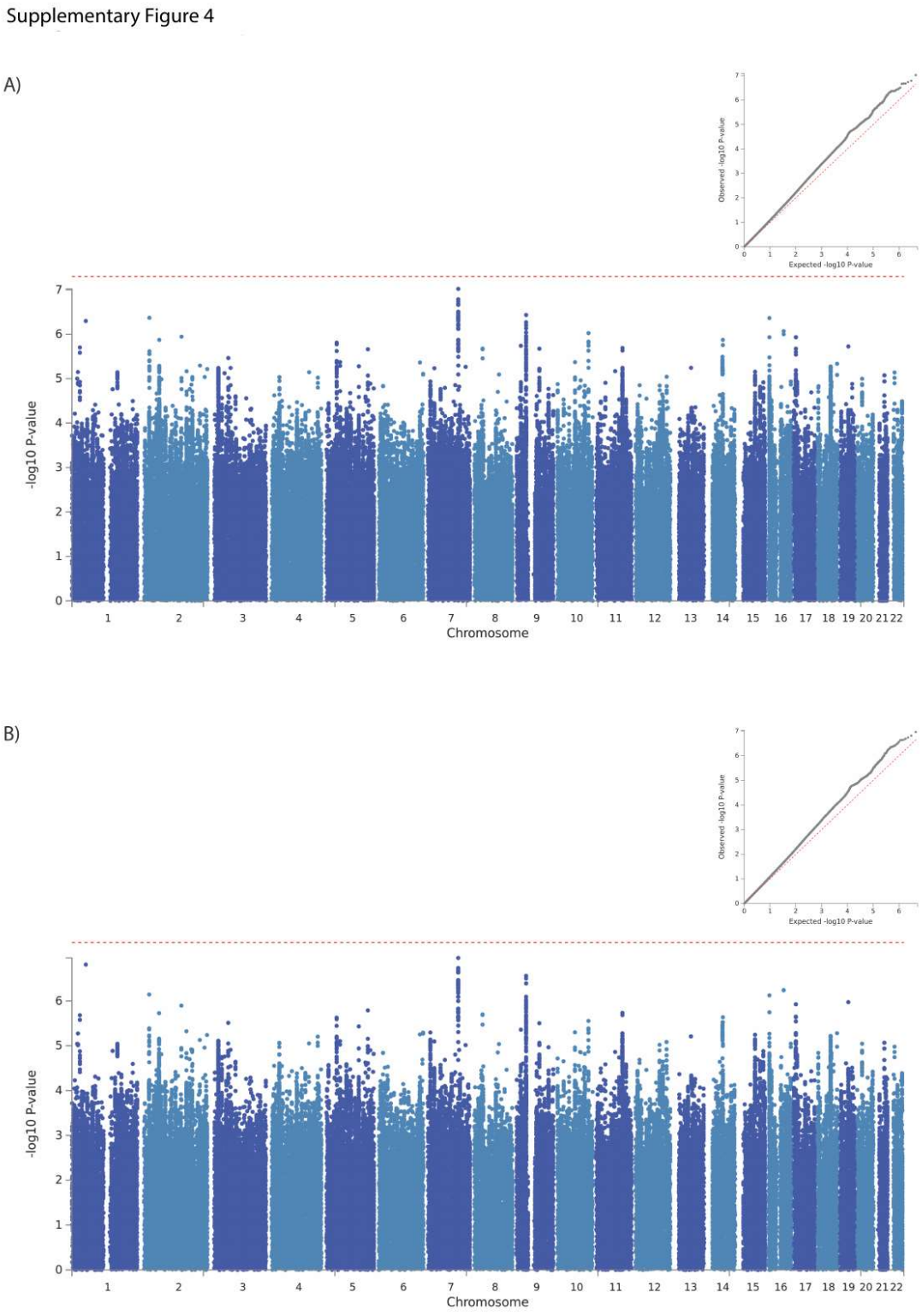
4.4.1.7 Supplementary Table 4

Effect of genetic loading for suicidal behaviour on psychiatric disorders and related traits.

Trait	GWAS threshold		*Unadjusted*							*Adjusted*			
		N	OR	95% CI		P	FDR adj P	N	OR	95% CI		P	FDR P
Self-report BD	5.00E-08	72714	1.03	0.96	1.1	0.43	0.43	72610	1.03	0.96	1.1	0.41	0.41
	5.00E-05	72714	1.11	1.03	1.18	0.0041	0.0049	72610	1.1	1.03	1.18	0.0049	0.0059
	0.01	72714	1.16	1.08	1.25	3.90E-05	5.85E-05	72610	1.15	1.07	1.24	1.10E-04	1.65E-04
	0.05	72714	1.17	1.09	1.26	2.10E-05	4.20E-05	72610	1.16	1.08	1.25	9.10E-05	1.65E-04
	0.1	72714	1.19	1.1	1.28	6.50E-06	1.95E-05	72610	1.17	1.09	1.26	3.20E-05	9.60E-05
	0.5	72714	1.22	1.12	1.32	1.40E-06	8.40E-06	72610	2.2	1.11	1.3	8.70E-06	5.22E-05
Self-report Depression	5.00E-08	82338	1.01	0.99	1.03	0.33	0.33	82217	1.01	0.99	1.03	0.38	0.38
	5.00E-05	82338	1.03	1.01	1.05	0.009	0.0108	82217	1.03	1.01	1.04	0.012	0.0144
	0.01	82338	1.09	1.07	1.11	8.80E-18	1.32E-17	82217	1.09	1.07	1.11	5.40E-17	8.10E-17
	0.05	82338	1.11	1.09	1.14	1.90E-26	3.80E-26	82217	1.11	1.09	1.14	2.50E-25	5.00E-25
	0.1	82338	1.12	1.09	1.14	4.30E-27	1.29E-26	82217	1.12	1.09	1.14	7.20E-26	2.16E-25
	0.5	82338	1.13	1.1	1.15	3.80E-27	1.29E-26	82217	1.12	1.1	1.15	4.40E-26	2.16E-25
Mood Instability	5.00E-08	283088	1.01	1	1.02	0.0037	0.0037	282761	1.01	1	1.02	0.003	0.003
	5.00E-05	283088	1.02	1.02	1.03	2.70E-09	3.24E-09	282761	1.02	1.02	1.03	1.70E-09	2.04E-09
	0.01	283088	1.07	1.06	1.07	2.40E-61	3.60E-61	282761	1.06	1.06	1.07	1.90E-54	2.85E-54
	0.05	283088	1.08	1.07	1.09	6.20E-86	1.24E-85	282761	1.08	1.07	1.09	1.60E-75	3.20E-75
	0.1	283088	1.09	1.08	1.09	2.20E-94	6.60E-94	282761	1.08	1.07	1.09	6.20E-83	1.86E-82
	0.5	283088	1.1	1.09	1.11	6.00E-108	3.60E-107	282761	1.09	1.09	1.1	1.40E-95	8.40E-95
Risk taking	5.00E-08	280508	1	0.99	1.01	0.62	0.62	280183	1	0.99	1.01	0.75	0.75
	5.00E-05	280508	1.01	1.01	1.02	0.0021	0.0025	280183	1.02	1.01	1.02	6.60E-04	7.92E-04
	0.01	280508	1.03	1.03	1.04	1.10E-13	1.65E-13	280183	1.03	1.02	1.04	1.70E-12	2.55E-12
	0.05	280508	1.04	1.03	1.05	2.60E-16	5.20E-16	280183	1.04	1.03	1.05	3.70E-14	7.40E-14
	0.1	280508	1.04	1.03	1.05	5.80E-18	1.80E-17	280183	1.1	1.03	1.05	1.30E-15	7.80E-15
	0.5	280508	1.04	1.03	1.05	6.00E-18	1.80E-17	280183	1.04	1.03	1.05	4.80E-15	1.44E-14
		N	Beta	SE		P	FDR adj P	N	Beta	SE		P	FDR adj P
Neuroticism (score)	5.00E-08	232471	0.03	0.01		2.20E-04	2.20E-04	232205	0.03	0.01		2.20E-04	2.20E-04
	5.00E-05	232471	0.05	0.01		3.20E-13	3.84E-13	232205	0.05	0.01		1.70E-13	2.04E-13
	0.01	232471	0.15	0.01		3.20E-95	4.80E-95	232205	0.14	0.01		5.70E-89	8.55E-89
	0.05	232471	0.18	0.01		2.00E-142	4.00E-142	232205	0.17	0.01		9.00E-133	1.80E-132
	0.1	232471	0.19	0.01		4.00E-158	1.20E-157	232205	0.18	0.01		8.00E-148	2.40E-147
	0.5	232471	0.22	0.01		6.00E-181	3.60E-180	232205	0.21	0.01		9.00E-171	5.40E-170

4.4.1.8 Supplementary Figure 4

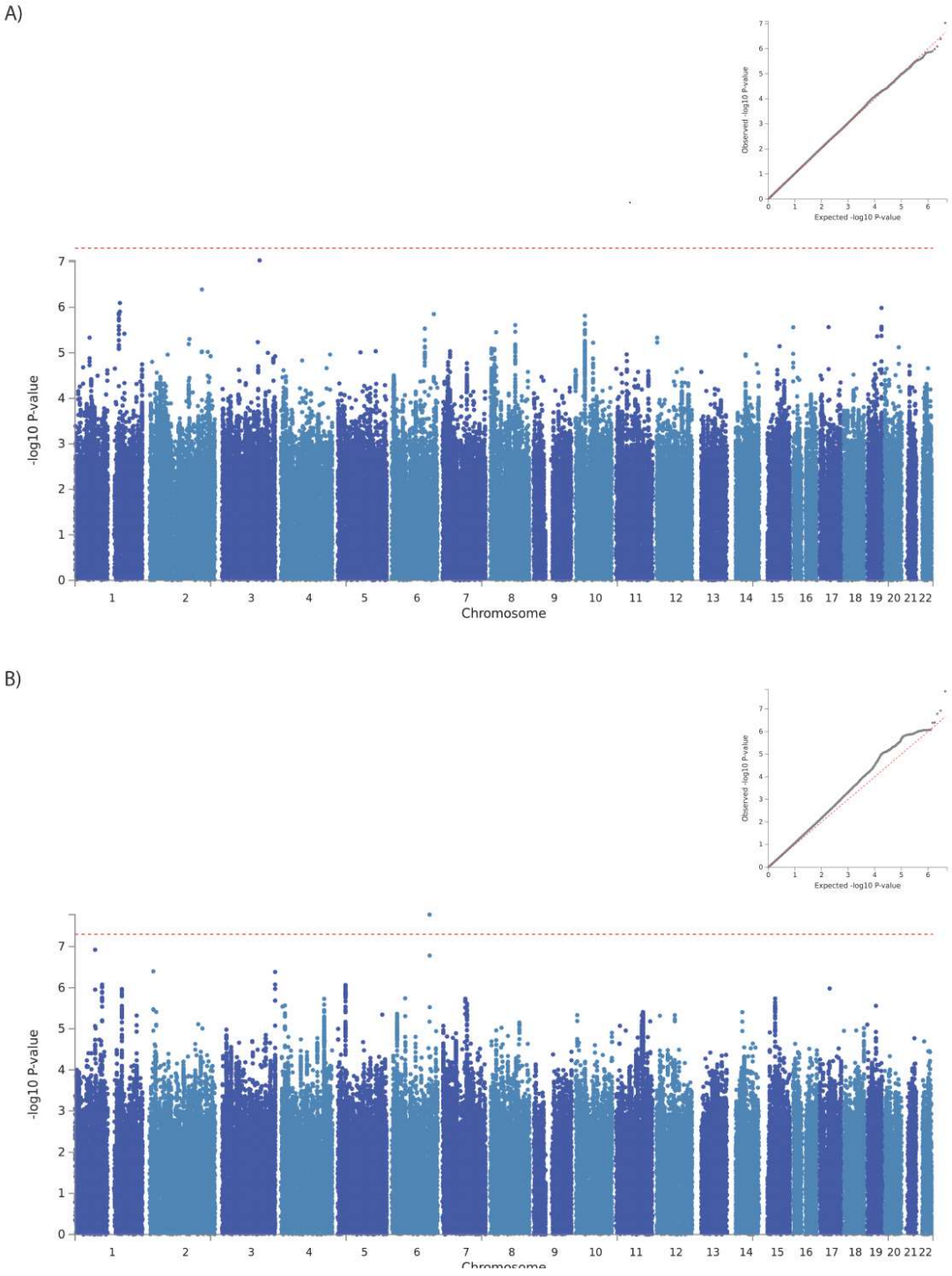
Manhattan plot of GWAS of ordinal DSH in UK Biobank (N = 100,234). Dashed red line = genome wide significance threshold ($p < 5 \times 10^{-5}$). Inset: QQ plot for genome-wide



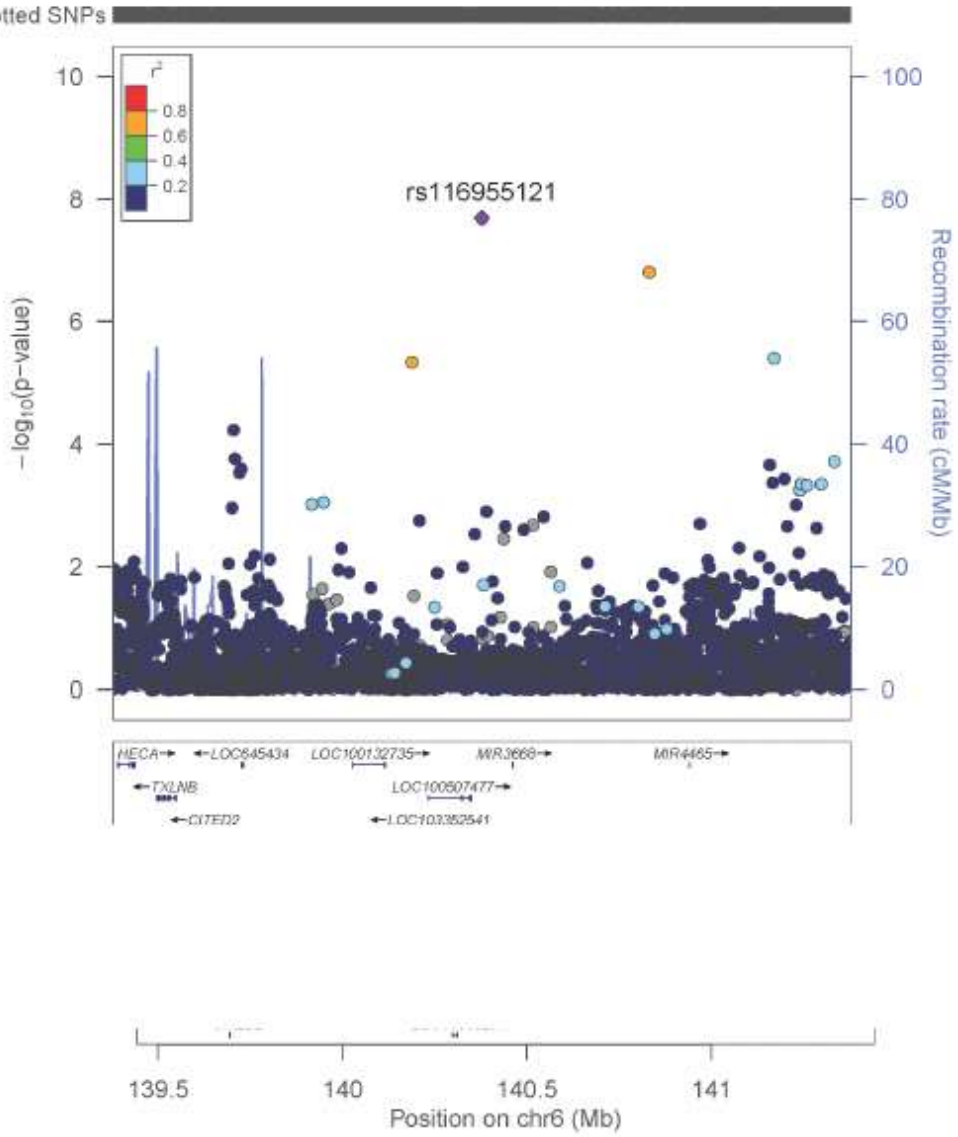
association with DSH. Red line = theoretical distribution under the null hypothesis of no association.

4.4.1.9 Supplementary Figure 5

Manhattan plot of GWAS of ordinal SIA in UK Biobank ($N = 108{,}090$). Dashed red line = genome wide significance threshold ($p < 5 \times 10^{-5}$). Inset: QQ plot for genome-wide association with SIA. Red line = theoretical distribution under the null hypothesis of no association.

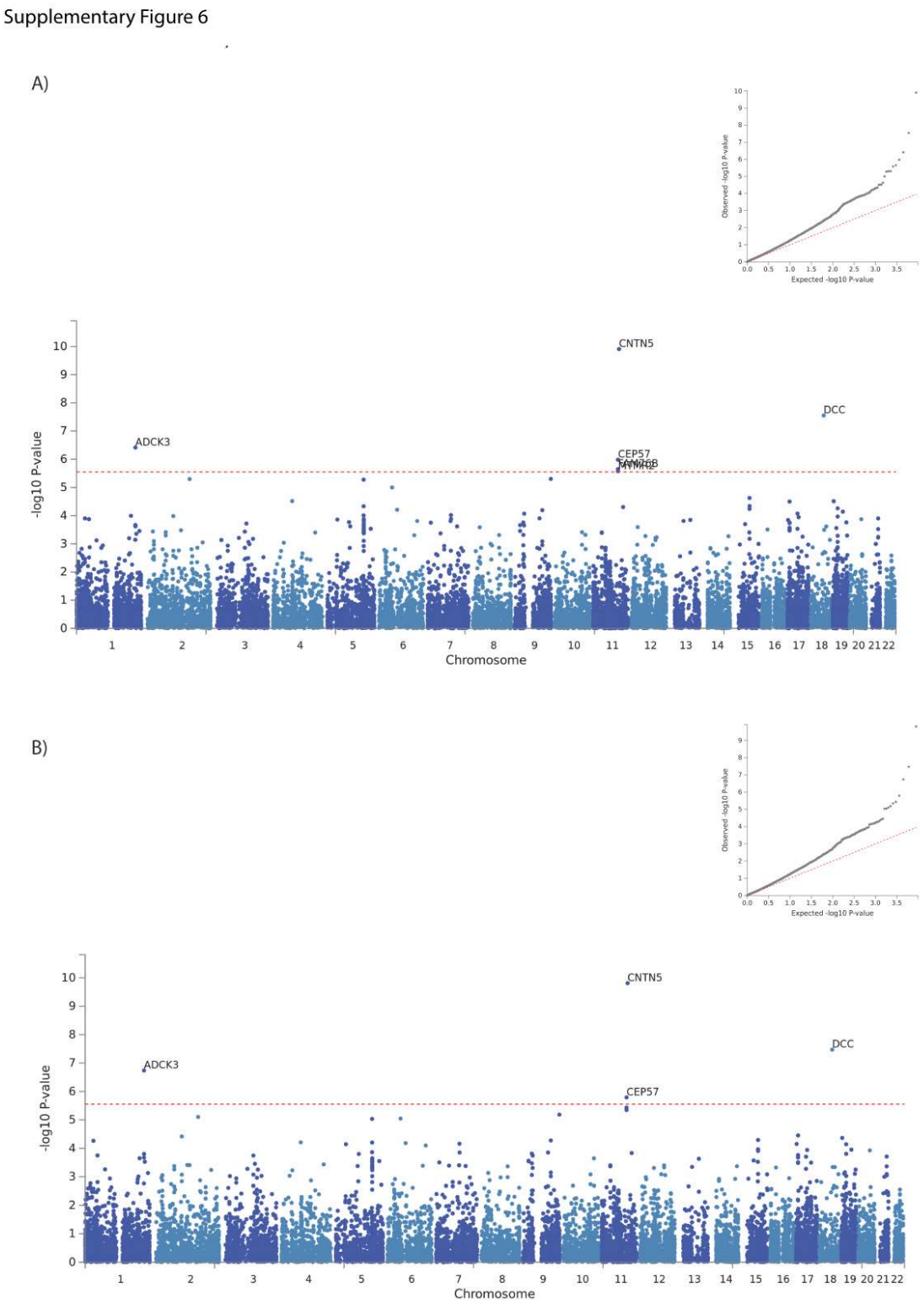


C)


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs116955121
10
8
6
4
2
0
100
80
60
40
20
0
−log10(p−value)
Recombination rate (cM/Mb)
HECA
LOC645434
LOC100132735
MIR3668
MIR4465
TXLNB
LOC100507477
CITED2
LOC103352541
139.5
140
140.5
141
Position on chr6 (Mb)

4.4.1.10 Supplementary Figure 6

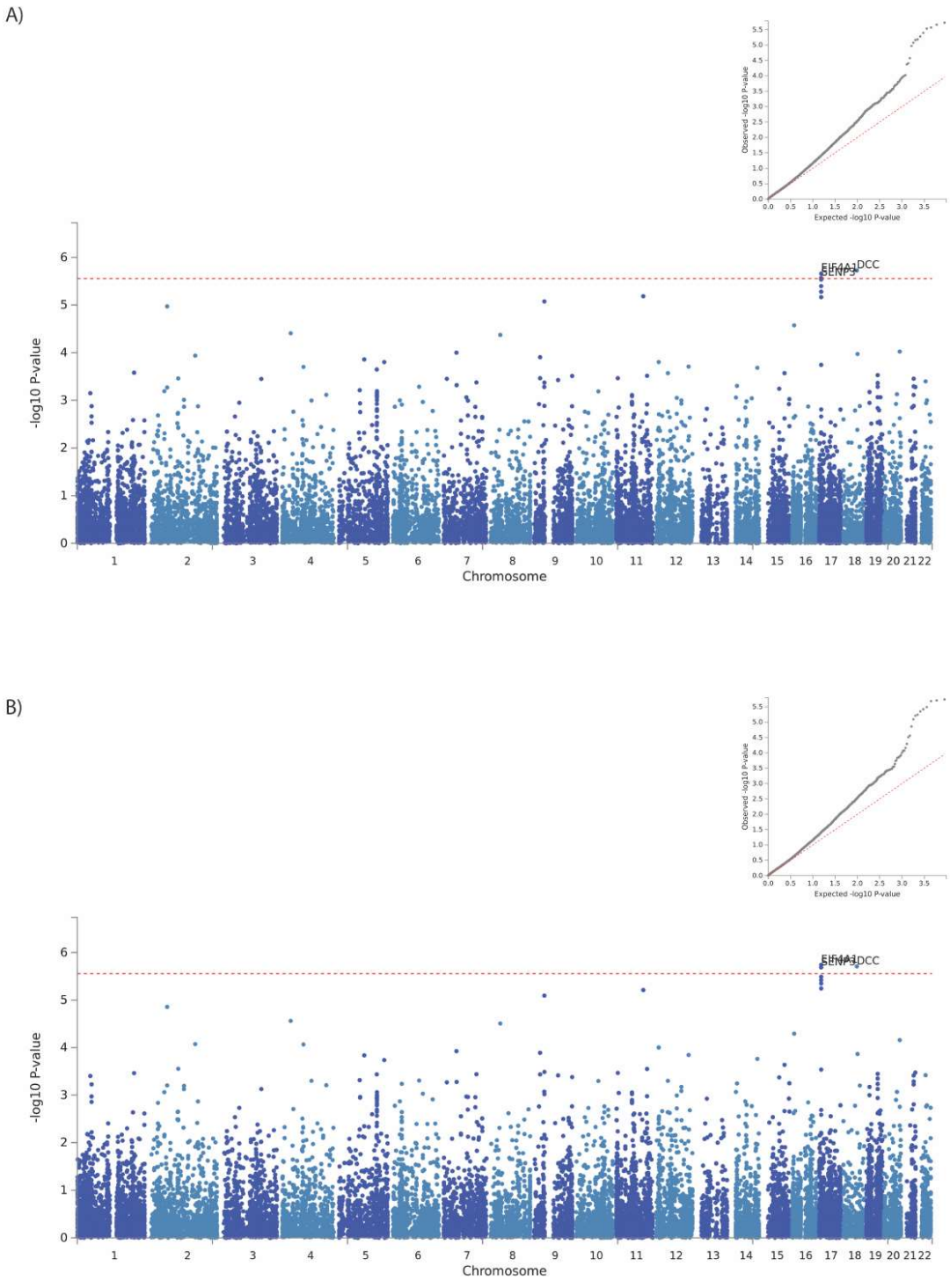
Manhattan plot of gene-based GWAS of ordinal suicide in UK Biobank ($N = 122{,}935$). Dashed red line = genome wide significance threshold ($p < 5 \times 10^{-5}$). Inset: QQ plot for genome-wide association with suicidality in UK Biobank. Red line = theoretical distribution under the null hypothesis of no association.



4.4.1.11 Supplementary Figure 7

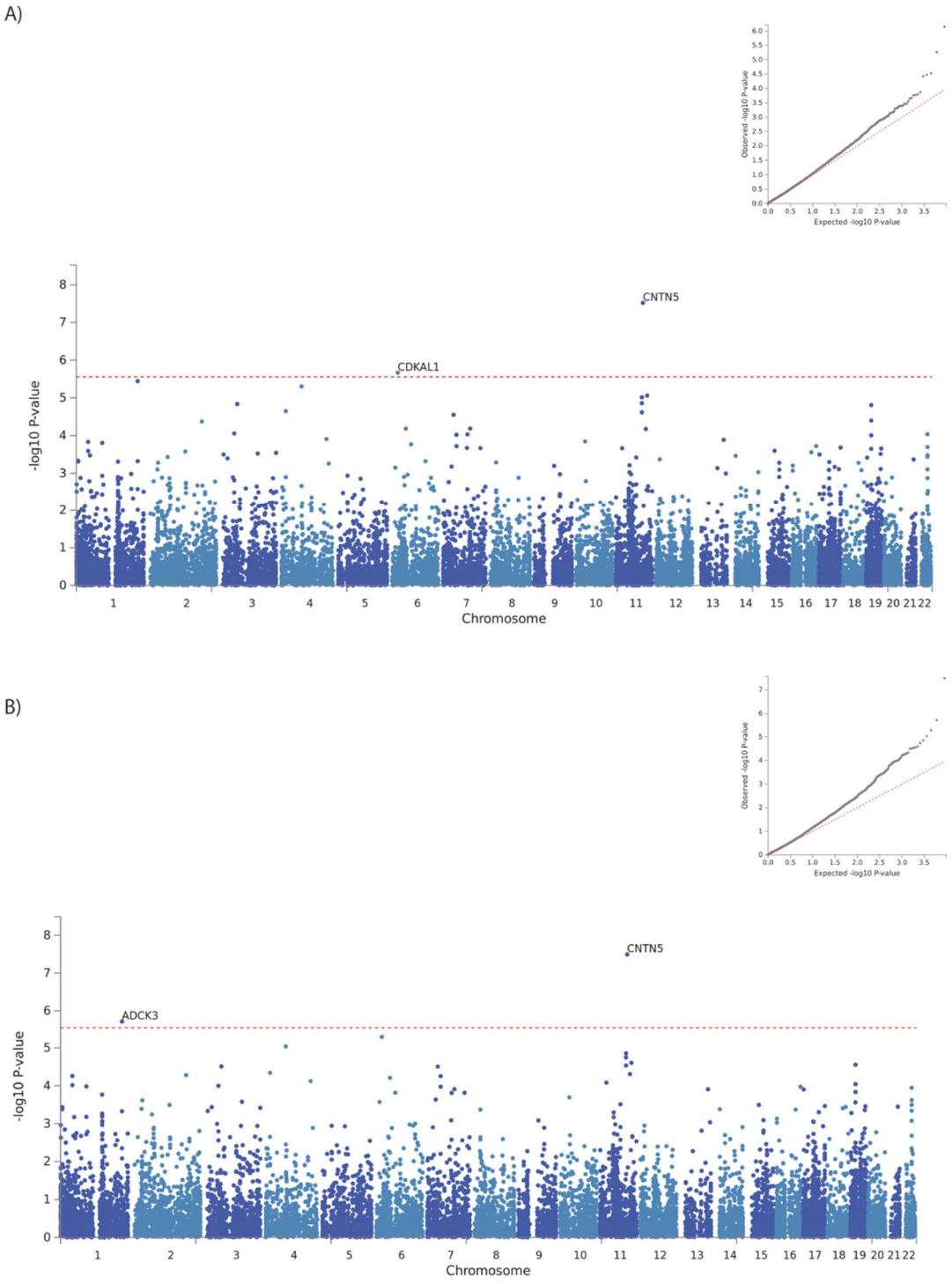
Manhattan plot of gene-based GWAS of ordinal DSH in UK Biobank (N = 100,234). Dashed red line = genome wide significance threshold ($p < 5 \times 10^{-5}$). Inset: QQ plot for genome-wide association with suicidality in UK Biobank. Red line = theoretical distribution under the null hypothesis of no association.

Supplementary Figure 7



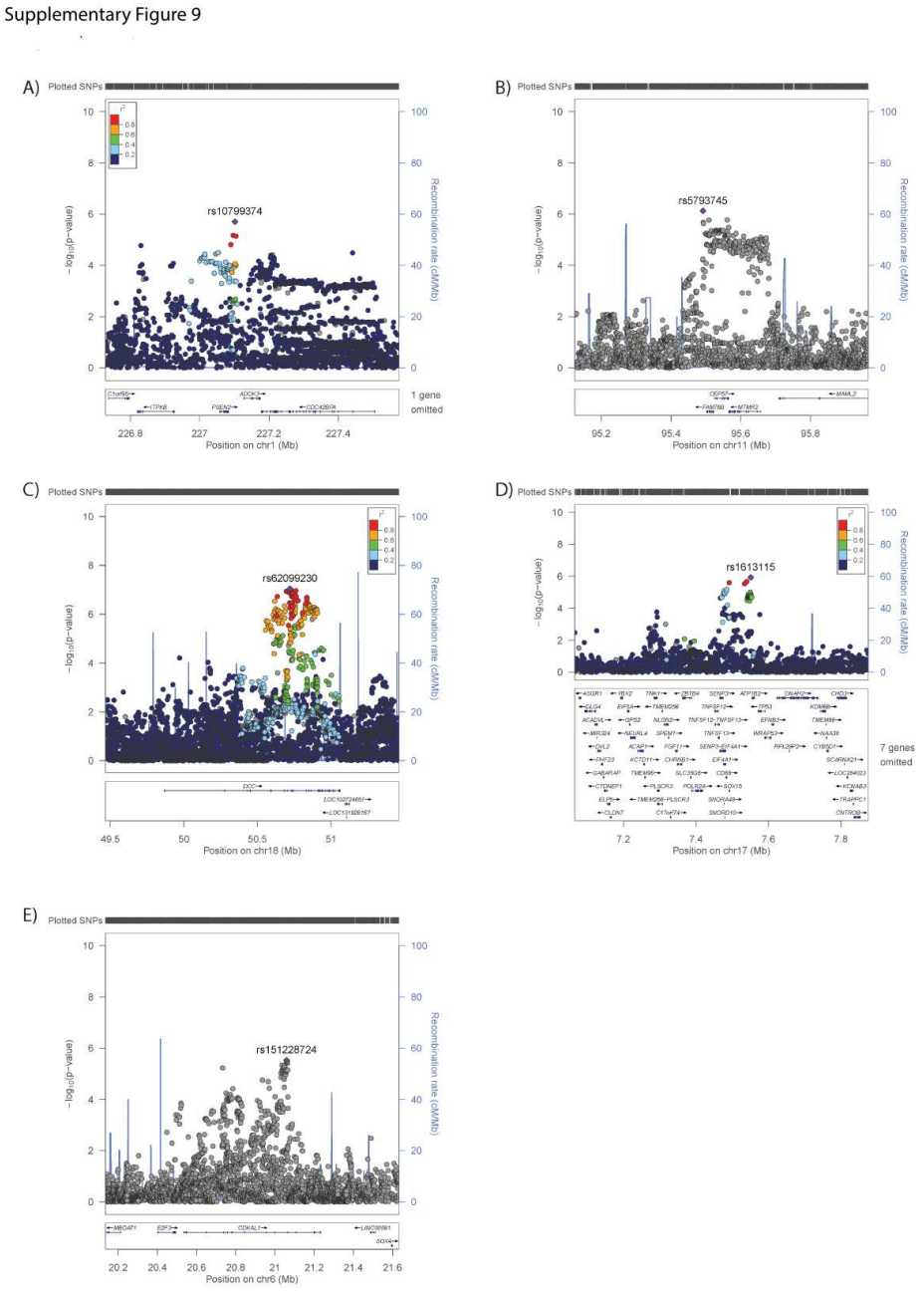
4.4.1.12 Supplementary Figure 8

Manhattan plot of gene-based GWAS of ordinal SIA in UK Biobank (N = 108,090). Dashed red line = genome wide significance threshold ($p < 5 \times 10^{-5}$). Inset: QQ plot for genome-wide association with suicidality in UK Biobank. Red line = theoretical distribution under the null hypothesis of no association.



4.4.1.13 Supplementary Figure 9

Regional plots for GWAS significant loci identified in the gene-based analyses. Highlighted genes for suicidality A) *ADCK3/COQ8A* on Chromosome 1, B) *CEP57-FAM76B-MTMR2* on Chromosome 11, C) *DCC* on Chromosome 18, For DSH D) *SENP3* on Chromosome 17 and for SIA E) *CDKAL1* on Chromosome 6. SNPs (each point) are aligned according to position (X axis) and strength of association (Y Axis, left); Purple colouring indicates the index SNP, with r^2 linkage disequilibrium with the index SNP being presented by colour as per the colour key; Rates of DNA recombination are presented as a pale blue line graph in the background (Y axis, right); Genes are presented by location (X axis) and direction of transcription (arrows).



4.4.1.14 Supplementary Table 5

Effect of genetic loading for suicidal behaviour on traits of relevance to psychiatric disorders

Gene	Name	Genecards summary
ADCK3/COQ8A	coenzyme Q8A	This gene encodes a mitochondrial protein similar to yeast ABC1, which functions in an electron-transferring membrane protein complex in the respiratory chain. It is not related to the family of ABC transporter proteins. Expression of this gene is induced by the tumour suppressor p53 and in response to DNA damage, and inhibiting its expression partially suppresses p53-induced apoptosis. Alternatively spliced transcript variants have been found; however, their full-length nature has not been determined.
AP001877.1		predicted gene
CDKAL1	CDK5 regulatory subunit associated protein 1 like 1	The protein encoded by this gene is a member of the methylthiotransferase family. The function of this gene is not known. Genome-wide association studies have linked single nucleotide polymorphisms in an intron of this gene with susceptibility to type 2 diabetes
CEP57	centrosomal protein 57	This gene encodes a cytoplasmic protein called Translokin. This protein localizes to the centrosome and has a function in microtubular stabilization. The N-terminal half of this protein is required for its centrosome localization and for its multimerization, and the C-terminal half is required for nucleating, bundling and anchoring microtubules to the centrosomes. This protein specifically interacts with fibroblast growth factor 2 (FGF2), sorting nexin 6, Ran-binding protein M and the kinesins KIF3A and KIF3B, and thus mediates the nuclear translocation and mitogenic activity of the FGF2. It also interacts with cyclin D1 and controls nucleocytoplasmic distribution of the cyclin D1 in quiescent cells. This protein is crucial for maintaining correct chromosomal number during cell division. Mutations in this gene cause mosaic variegated aneuploidy syndrome, a rare autosomal recessive disorder. Multiple alternatively spliced transcript variants encoding different isoforms have been identified.

CLTA	Clathrin Light Chain A	Clathrin is a large, soluble protein composed of heavy and light chains. It functions as the main structural component of the lattice-type cytoplasmic face of coated pits and vesicles which entrap specific macromolecules during receptor-mediated endocytosis. This gene encodes one of two clathrin light chain proteins which are believed to function as regulatory elements. Alternative splicing results in multiple transcript variants. Related pseudogenes have been identified on chromosomes 8 and 12. Diseases associated with CLTA include Leber Congenital Amaurosis. Among its related pathways are Clathrin derived vesicle budding and EPH-Ephrin signalling. Gene Ontology (GO) annotations related to this gene include structural molecule activity and clathrin heavy chain binding. An important paralog of this gene is CLTB. Acts as component of the TACC3/ch-TOG/clathrin complex proposed to contribute to stabilization of kinetochore fibres of the mitotic spindle by acting as inter-microtubule bridge.
CNTN5	Contactin 5	The protein encoded by this gene is a member of the immunoglobulin superfamily, and contactin family, which mediate cell surface interactions during nervous system development. This protein is a glycosylphosphatidylinositol (GPI)-anchored neuronal membrane protein that functions as a cell adhesion molecule. It may play a role in the formation of axon connections in the developing nervous system. Alternatively spliced transcript variants encoding different isoforms have been described for this gene. [provided by RefSeq, Aug 2011] CNTN5 (Contactin 5) is a Protein Coding gene. Diseases associated with CNTN5 include Actinomycosis. Among its related pathways are Metabolism of proteins and Post-translational modification- synthesis of GPI-anchored proteins. An important paralog of this gene is CNTN3. Contactins mediate cell surface interactions during nervous system development. Has some neurite outgrowth-promoting activity in the cerebral cortical neurons but not in hippocampal neurons. Probably involved in neuronal activity in the auditory system (By similarity).
DCAF10	DDB1 And CUL4 Associated Factor 10	A protein coding gene. May function as a substrate receptor for CUL4-DDB1 E3 ubiquitin-protein ligase complex.
DCC	DCC netrin 1 receptor	This gene encodes a netrin 1 receptor. The transmembrane protein is a member of the immunoglobulin superfamily of cell adhesion molecules, and mediates axon guidance of neuronal growth cones towards sources of netrin 1 ligand. The cytoplasmic tail interacts with the tyrosine kinases Src and focal adhesion kinase (FAK, also known as PTK2) to mediate axon attraction. The protein partially localizes to lipid rafts, and induces apoptosis in the absence of ligand. The protein functions as a tumour suppressor, and is frequently mutated or downregulated in colorectal cancer and oesophageal carcinoma.

EBLN3	endogenous Bornavirus-like nucleoprotein 3, pseudogene	pseudogene
EIF4A1	Eukaryotic translation initiation factor 4A1	Ubiquitous expression in many tissues
EXOSC3	Exosome Component 3	This gene encodes a non-catalytic component of the human exosome, a complex with 3'-5' exoribonuclease activity that plays a role in numerous RNA processing and degradation activities. Related pseudogenes of this gene are found on chromosome 19 and 21. Alternatively spliced transcript variants encoding different isoforms have been described. Diseases associated with EXOSC3 include Pontocerebellar Hypoplasia, Type 1B and Exosc3-Related Pontocerebellar Hypoplasia. Among its related pathways are CDK-mediated phosphorylation and removal of Cdc6 and Gene Expression. Gene Ontology (GO) annotations related to this gene include RNA binding and exoribonuclease activity. Non-catalytic component of the RNA exosome complex which has 3->5 exoribonuclease activity and participates in a multitude of cellular RNA processing and degradation events. In the nucleus, the RNA exosome complex is involved in proper maturation of stable RNA species such as rRNA, snRNA and snoRNA, in the elimination of RNA processing by-products and non-coding pervasive transcripts, such as antisense RNA species and promoter-upstream transcripts (PROMPTs), and of mRNAs with processing defects, thereby limiting or excluding their export to the cytoplasm. The RNA exosome may be involved in Ig class switch recombination (CSR) and/or Ig variable region somatic hypermutation (SHM) by targeting AICDA deamination activity to transcribed dsDNA substrates. In the cytoplasm, the RNA exosome complex is involved in general mRNA turnover and specifically degrades inherently unstable mRNAs containing AU-rich elements (AREs) within their 3 untranslated regions, and in RNA surveillance pathways, preventing translation of aberrant mRNAs. It seems to be involved in degradation of histone mRNA. The catalytic inactive RNA exosome core complex of 9 subunits (Exo-9) is proposed to play a pivotal role in the binding and presentation of RNA for ribonucleolysis, and to serve as a scaffold for the association with catalytic subunits and accessory proteins or complexes. EXOSC3 as peripheral part of the Exo-9 complex stabilizes the hexameric ring of RNase PH-domain subunits through contacts with EXOSC9 and EXOSC5.
FAM76B	family with sequence similarity 76 member B	

FBXO10	f-Box protein 10	Members of the F-box protein family, such as FBXO10, are characterized by an approximately 40-amino acid F-box motif. SCF complexes, formed by SKP1 (MIM 601434), cullin (see CUL1; MIM 603134), and F-box proteins, act as protein-ubiquitin ligases. F-box proteins interact with SKP1 through the F box, and they interact with ubiquitination targets through other protein interaction domains (Jin et al., 2004 [PubMed 15520277]).[supplied by OMIM, Mar 2008] FBXO10 (F-Box Protein 10) is a Protein Coding gene. Among its related pathways are Innate Immune System and Class I MHC mediated antigen processing and presentation. Gene Ontology (GO) annotations related to this gene include ubiquitin-protein transferase activity. An important paralog of this gene is FBXO11. Substrate-recognition component of the SCF (SKP1-CUL1-F-box protein)-type E3 ubiquitin ligase complex. The SCF(FBXO10) complex mediates ubiquitination and degradation of BCL2, an antiapoptotic protein, thereby playing a role in apoptosis by controlling the stability of BCL2.
FRMPD1	FERM And PDZ Domain Containing 1	FRMPD1 is a Protein Coding gene. An important paralog of this gene is FRMPD4. Stabilizes membrane-bound GPSM1, and thereby promotes its interaction with GNAI1.
GNE	glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase	the protein encoded by this gene is a bifunctional enzyme that initiates and regulates the biosynthesis of N-acetylneuraminic acid (NeuAc), a precursor of sialic acids. It is a rate-limiting enzyme in the sialic acid biosynthetic pathway. Sialic acid modification of cell surface molecules is crucial for their function in many biologic processes, including cell adhesion and signal transduction. Differential sialylation of cell surface molecules is also implicated in the tumorigenicity and metastatic behaviour of malignant cells. Mutations in this gene are associated with sialuria, autosomal recessive inclusion body myopathy, and Nonaka myopathy. Alternative splicing of this gene results in transcript variants encoding different isoforms. Diseases associated with GNE include Nonaka Myopathy and Sialuria. Among its related pathways are Metabolism of proteins and Transport to the Golgi and subsequent modification. Gene Ontology (GO) annotations related to this gene include hydrolase activity, hydrolysing O-glycosyl compounds and UDP-N-acetylglucosamine 2-epimerase activity. Required for normal sialylation in hematopoietic cells.

GRHPR	Glyoxylate And Hydroxypyruvate Reductase	This gene encodes an enzyme with hydroxypyruvate reductase, glyoxylate reductase, and D-glycerate dehydrogenase enzymatic activities. The enzyme has widespread tissue expression and has a role in metabolism. Type II hyperoxaluria is caused by mutations in this gene. Diseases associated with GRHPR include Hyperoxaluria, Primary, Type Ii and Primary Hyperoxaluria. Among its related pathways are Glyoxylate metabolism and glycine degradation and Viral mRNA Translation. Gene Ontology (GO) annotations related to this gene include protein homodimerization activity and oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor. Enzyme with hydroxy-pyruvate reductase, glyoxylate reductase and D-glycerate dehydrogenase enzymatic activities. Reduces hydroxypyruvate to D-glycerate, glyoxylate to glycolate oxidizes D-glycerate to hydroxypyruvate.
LINC00395	Long Intergenic Non-Protein Coding RNA 395	long non-coding RNA
MELK	maternal embryonic leucine zipper kinase	A protein coding gene. Diseases associated with MELK include Uterine Corpus Endometrial Carcinoma. Among its related pathways are Neuroscience. Gene Ontology (GO) annotations related to this gene include *calcium ion binding* and *protein kinase activity*. An important paralog of this gene is PRKAA2. Serine/threonine-protein kinase involved in various processes such as cell cycle regulation, self-renewal of stem cells, apoptosis and splicing regulation. Has a broad substrate specificity; phosphorylates BCL2L14, CDC25B, MAP3K5/ASK1 and ZNF622. Acts as an activator of apoptosis by phosphorylating and activating MAP3K5/ASK1. Acts as a regulator of cell cycle, notably by mediating phosphorylation of CDC25B, promoting localization of CDC25B to the centrosome and the spindle poles during mitosis. Plays a key role in cell proliferation and carcinogenesis. Required for proliferation of embryonic and postnatal multipotent neural progenitors. Phosphorylates and inhibits BCL2L14, possibly leading to affect mammary carcinogenesis by mediating inhibition of the pro-apoptotic function of BCL2L14. Also involved in the inhibition of spliceosome assembly during mitosis by phosphorylating ZNF622, thereby contributing to its redirection to the nucleus. May also play a role in primitive hematopoiesis.
MTMR2	myotubularin related protein 2	This gene is a member of the myotubularin family of phosphoinositide lipid phosphatases. The encoded protein possesses phosphatase activity towards phosphatidylinositol-3-phosphate and phosphatidylinositol-3,5-bisphosphate. Mutations in this gene are a cause of Charcot-Marie-Tooth disease type 4B, an autosomal recessive demyelinating neuropathy. Alternatively spliced transcript variants encoding multiple isoforms have been found for this gene

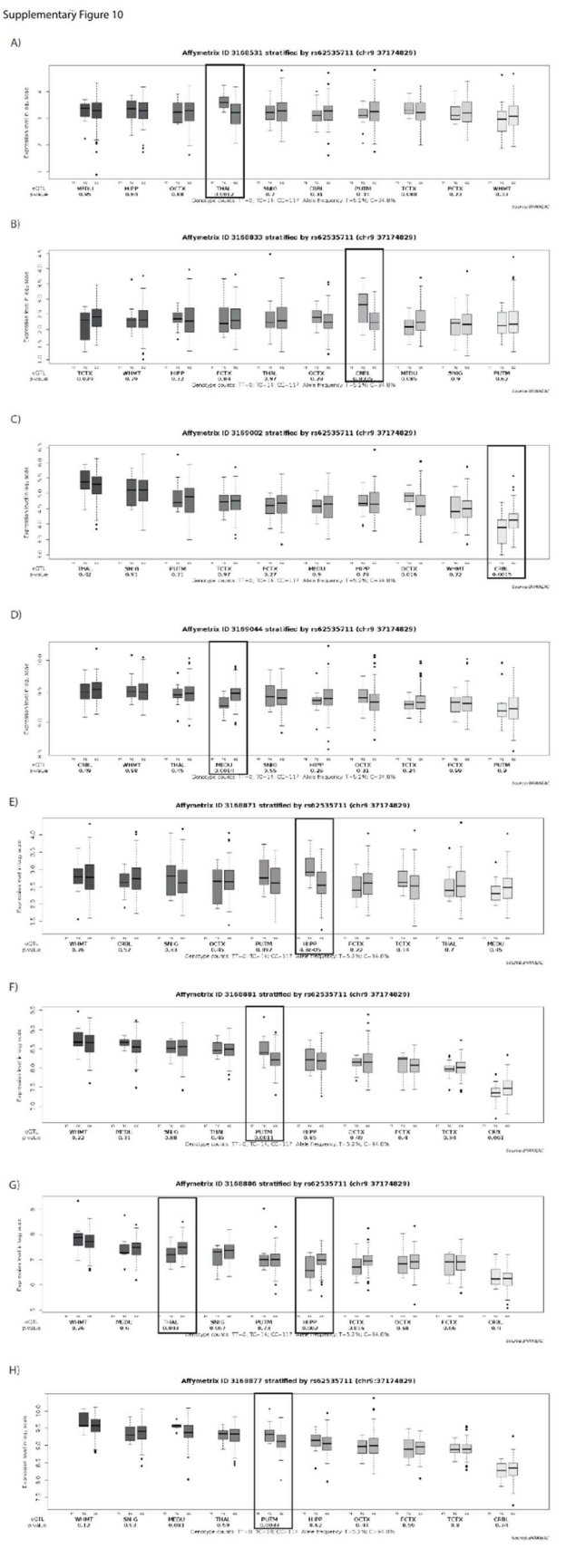
OR7E156p	Olfactory Receptor Family 7 Subfamily E Member 156 Pseudogene	pseudogene. Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms.
PAX5	paired box 5	A transcription factors, whose central feature is a highly conserved DNA-binding motif (the paired box). Paired box transcription factors are important regulators in early development, and alterations in the expression of their genes are thought to contribute to neoplastic transformation. This gene encodes the B-cell lineage specific activator protein that is expressed at early, but not late stages of B-cell differentiation. Its expression has also been detected in developing CNS and testis and so the encoded protein may also play a role in neural development and spermatogenesis. This gene is located at 9p13, which is involved in t(9;14)(p13;q32) translocations recurring in small lymphocytic lymphomas of the plasmacytoid subtype, and in derived large-cell lymphomas. This translocation brings the potent E-mu enhancer of the IgH gene into close proximity of the PAX5 promoter, suggesting that the deregulation of transcription of this gene contributes to the pathogenesis of these lymphomas. Alternative splicing results in multiple transcript variants encoding different isoforms. T
POLR1E	RNA Polymerase I Subunit E	POLR1E (RNA Polymerase I Subunit E) is a Protein Coding gene. Among its related pathways are Pyrimidine metabolism (KEGG) and Gene Expression. Gene Ontology (GO) annotations related to this gene include DNA-directed 5-3 RNA polymerase activity and RNA polymerase I transcription factor binding. DNA-dependent RNA polymerase catalyses the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates. Component of RNA polymerase I which synthesizes ribosomal RNA precursors. Appears to be involved in the formation of the initiation complex at the promoter by mediating the interaction between Pol I and UBTF/UBF (By similarity).

RNF38	ring finger protein 38	This gene encodes a protein with a coiled-coil motif and a RING-H2 motif (C3H2C2) at its carboxy-terminus. The RING motif is a zinc-binding domain found in a large set of proteins playing roles in diverse cellular processes including oncogenesis, development, signal transduction, and apoptosis. Multiple transcript variants encoding different isoforms have been found for this gene. Acts as an E3 ubiquitin-protein ligase able to ubiquitinate p53/TP53 which promotes its relocalization to discrete foci associated with PML nuclear bodies. Exhibits preference for UBE2D2 as a E2 enzyme. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Dec 2008]
SENP3	SUMO specific peptidase 3	The reversible posttranslational modification of proteins by the addition of small ubiquitin-like SUMO proteins (see SUMO1; MIM 601912) is required for numerous biologic processes. SUMO-specific proteases, such as SENP3, are responsible for the initial processing of SUMO precursors to generate a C-terminal diglycine motif required for the conjugation reaction. They also have isopeptidase activity for the removal of SUMO from high molecular mass SUMO conjugates
SHB	SH2 Domain Containing Adaptor Protein B	A protein coding gene. Diseases associated with SHB include Sulfhemoglobinemia. Among its related pathways are EPH-Ephrin signalling and Developmental Biology. Gene Ontology (GO) annotations related to this gene include SH3/SH2 adaptor activity. An important paralog of this gene is ENSG00000255872. Adapter protein which regulates several signal transduction cascades by linking activated receptors to downstream signalling components. May play a role in angiogenesis by regulating FGFR1, VEGFR2 and PDGFR signalling. May also play a role in T-cell antigen receptor/TCR signalling, interleukin-2 signalling, apoptosis and neuronal cells differentiation by mediating basic-FGF and NGF-induced signalling cascades. May also regulate IRS1 and IRS2 signalling in insulin-producing cells.
SLC25A51	Solute Carrier Family 25 Member 51	A protein coding gene. An important paralog of this gene is SLC25A52.
TOMM5	Translocase Of Outer Mitochondrial Membrane 5	TOMM5 (Translocase Of Outer Mitochondrial Membrane 5) is a Protein Coding gene. Among its related pathways are Pink/Parkin Mediated Mitophagy and Metabolism of proteins. Gene Ontology (GO) annotations related to this gene include *protein transporter activity*. An important paralog of this gene is ENSG00000256966.

TRMT10B	TRNA Methyltransferase 10B	TRMT10B is a Protein Coding gene. Gene Ontology (GO) annotations related to this gene include methyltransferase activity. An important paralog of this gene is TRMT10A. S-adenosyl-L-methionine-dependent guanine N(1)-methyltransferase that catalyses the formation of N(1)-methylguanine at position 9 (m1G9) in tRNAs (PubMed:23042678). Probably not able to catalyse formation of N(1)-methyladenine at position 9 (m1A9) in tRNAs (PubMed:23042678).
ZBTB5	Zinc Finger And BTB Domain Containing 5	ZBTB5 (Zinc Finger And BTB Domain Containing 5) is a Protein Coding gene. An important paralog of this gene is ZBTB3. May be involved in transcriptional regulation
ZCCHC7	Zinc Finger, CCHC Domain Containing 7	A protein coding gene. Related pathways include Deadenylation-dependent mRNA decay. Gene Ontology (GO) annotations related to this gene include nucleic acid binding.

4.4.1.15 Supplementary Figure 10

Genotype-specific gene expression of the Chromosome 9 lead SNP, rs62535711 on transcripts of *FRMPD1, MELK, TRMT10B, ZCCHC7* and *GRHPR*, (Supplementary Fig. 3 E–H) in cerebellar cortex (CRBL), frontal cortex (FCTX), hippocampus (HIPP), medulla (specifically inferior olivary nucleus, MEDU), occipital cortex (specifically primary visual cortex, OCTX), putamen (PUTM), substantia nigra (SNIG), thalamus (THAL), temporal cortex (TCTX) and intralobular white matter (WHMT).



4.4.1.16 Supplementary Table 6

Functions of genes in novel suicidality loci eQTLs in brain of predicted functional SNPs in the *CEP57-FAM76B* locus

SNP	bp	Gene Symbol	NES	Tissue
rs1150360	95512060	*AP001877.1*	0.45	Brain - Caudate (basal ganglia)
		AP001877.1	0.6	Brain - Cerebellar Hemisphere
		AP001877.1	0.56	Brain - Cerebellum
		AP001877.1	0.59	Brain - Frontal Cortex (BA9)
		AP001877.1	0.5	Brain - Hippocampus
		AP001877.1	0.53	Brain - Nucleus accumbens (basal ganglia)
		AP001877.1	0.48	Brain - Putamen (basal ganglia)
		AP001877.1	0.58	Brain - Spinal cord (cervical c-1)
		CEP57	0.21	Brain - Cerebellum
rs3824874	95657111	*AP001877.1*	0.7	Brain - Anterior cingulate cortex (BA24)
		AP001877.1	0.62	Brain - Caudate (basal ganglia)
		AP001877.1	0.72	Brain - Cerebellar Hemisphere
		AP001877.1	0.8	Brain - Cerebellum
		AP001877.1	0.67	Brain - Cortex
		AP001877.1	0.64	Brain - Frontal Cortex (BA9)
		AP001877.1	0.64	Brain - Hippocampus
		AP001877.1	0.69	Brain - Hypothalamus
		AP001877.1	0.69	Brain - Nucleus accumbens (basal ganglia)
		AP001877.1	0.47	Brain - Putamen (basal ganglia)
		AP001877.1	0.79	Brain - Spinal cord (cervical c-1)
rs644799	95564259	*AP001877.1*	0.71	Brain - Anterior cingulate cortex (BA24)
		AP001877.1	0.63	Brain - Caudate (basal ganglia)
		AP001877.1	0.77	Brain - Cerebellar Hemisphere
		AP001877.1	0.83	Brain - Cerebellum
		AP001877.1	0.7	Brain - Cortex
		AP001877.1	0.73	Brain - Frontal Cortex (BA9)
		AP001877.1	0.64	Brain - Hippocampus
		AP001877.1	0.68	Brain - Hypothalamus
		AP001877.1	0.68	Brain - Nucleus accumbens (basal ganglia)
		AP001877.1	0.55	Brain - Putamen (basal ganglia)
		AP001877.1	0.88	Brain - Spinal cord (cervical c-1)
		AP001877.1	0.58	Brain - Substantia nigra

4.4.1.17 Supplementary Table 7

Novel Suicidal behaviour loci and previous associations in the GWAS catalogue

Search Term	PUBMEDID	FIRST AUTHOR	DISEASE/TRAIT	STRONGEST SNP-RISK ALLELE	RAF	P-VALUE	OR or BETA
	22449649	Loo SK	**Intelligence**	rs1329573-?; rs7020413-?; rs3824344-?; rs3758171-?	NR	4.00E-08	
chr9 locus	22491018	Wang J	Response to tocilizumab in rheumatoid arthritis	rs1329568-?	0.03	8.00E-07	19.64
	26198764	Goes FS	**Schizophrenia**	rs7020830-T	NR	8.00E-06	1.053
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A; rs1329568-T		7.00E-07	1.64
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A; rs1329568-T		5.00E-08	
TOMM5	28753643	Yeo A	Lipoprotein phospholipase A2 activity in cardiovascular disease	rs57578064-A	0.01	1.00E-08	32.49
DCAF10	27629369	Gao J	**Loneliness (linear analysis)**	rs78173384-?		2.00E-06	0.374
	28928442	Tian C	Cold sores	rs776014-?	NR	7.00E-06	0.163
FRMPD1	29299148	Liu W	Cancer	rs2182318-?; rs7856656-?	NR	1.00E-06	1.39
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-?		4.00E-07	
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A	0.69	5.00E-07	1.32
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A; rs1329568-T		7.00E-07	1.64
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A; rs1329568-T		5.00E-08	
SHB	23934736	Yu B	Metabolite levels (Dihydroxy docosatrienoic acid)	rs3747547-C		5.00E-06	0.09
	27629369	Gao J	**Loneliness (linear analysis)**	rs149411702-?		2.00E-06	0.376
	20171287	Stein JL	**Brain structure**	rs7873102-?	0.38	6.00E-07	
	28060188	Hamet P	Type 2 diabetes (age of onset)	rs10973627-?	NR	4.00E-06	
PAX5	22449649	Loo SK	Intelligence	rs1329573-?; rs7020413-?; rs3824344-?; rs3758171-?	NR	4.00E-08	
	22013104	Melka MG	Obesity-related traits	rs16933812-?		5.00E-06	
	22013104	Melka MG	Obesity-related traits	rs16933812-?		9.00E-09	
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-?		4.00E-07	
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A	0.69	5.00E-07	1.32
	22491018	Wang J	Response to tocilizumab in rheumatoid arthritis	rs1329568-?	0.03	8.00E-07	19.64
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A; rs1329568-T		7.00E-07	1.64
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A; rs1329568-T		5.00E-08	
ELBN3	26198764	Goes FS	**Schizophrenia**	rs7020830-T	NR	8.00E-06	1.053
ZCCHC7	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-?		4.00E-07	

	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A	0.69	5.00E-07	1.32
GRHPR	29299148	Liu W	Cancer	rs2182318-?; rs7856656-?	NR	1.00E-06	1.39
FBXO10	27989323	Ahola-Olli AV	Macrophage inflammatory protein 1b levels	rs76582507-G	NR	3.00E-06	0.318
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-?		4.00E-07	
	25786224	Johnson EO	HIV-1 susceptibility	rs4878712-A	0.69	5.00E-07	1.32
chr11 locus	26503763	Song J	**Objective response to lithium treatment in bipolar disorder**	rs113262272-A	0.71	4.00E-06	1.93
	26911590	Mullin BH	Bone mineral density (femoral neck)	rs10893396-C	0.17	4.00E-07	0.15
	27629369	Gao J	**Loneliness (multivariate analysis)**	rs10893420-?	NR	7.00E-06	
CNTN5	23049088	Meng W	Myopia (pathological)	rs12803066-?	0.52	4.00E-06	
	23251661	Comuzzie AG	Obesity-related traits	rs11217223-G	0.09	1.00E-06	0.04
	23251661	Comuzzie AG	Obesity-related traits	rs11217223-G	0.09	6.00E-06	0.03
	27629369	Gao J	**Loneliness (multivariate analysis)**	rs10893420-?	NR	7.00E-06	
	26365420	Mbarek H	**Alcohol dependence**	rs117557854-A	0.02	2.00E-06	2.083
	17903304	Larson MG	Atrial fibrillation	rs10501920-?	NR	9.00E-06	
	17903297	Seshadri S	**Volumetric brain MRI**	rs952700-?	NR	6.00E-06	
	22610502	Kennedy RB	Immune response to smallpox (secreted IL-2)	rs11223581-G		9.00E-09	
	26911590	Mullin BH	Bone mineral density (femoral neck)	rs10893396-C	0.17	4.00E-07	0.15
	23667675	Tanikawa C	Menarche (age at onset)	rs12800752-T	0.8	3.00E-06	0.095
	28240269	Suhre K	Blood protein levels	rs1461672-T	0.06	2.00E-18	0.734
	29317604	Liu D	Plasma kynurenine levels in major depressive disorder	rs61475845-A	0.02	1.00E-06	2.2
	20889312	Wang KS	**Bipolar disorder and schizophrenia**	rs2509843-?	NR	2.00E-07	1.268
	24529757	Xie T	Amyotrophic lateral sclerosis (sporadic)	kgp11394149-?		2.00E-07	
	27989323	Ahola-Olli AV	Interleukin-2 receptor antagonist levels	rs73001149-C		9.00E-06	0.549
	29292387	Turley P	**Neuroticism**	rs1690816-C	0.31	3.00E-06	0.017
	29292387	Turley P	**Depressive symptoms (multi-trait analysis)**	rs586533-G	0.32	2.00E-09	0.014
	29292387	Turley P	**Neuroticism (multi-trait analysis)**	rs1690816-C	0.31	3.00E-09	0.018
	26503763	Song J	**Objective response to lithium treatment in bipolar disorder**	rs113262272-A	0.71	4.00E-06	1.93
	25646338	Mozaffarian D	Trans fatty acid levels	rs7952067-C	0.07	1.00E-06	0.005
	20708005	Chalasani N	Non-alcoholic fatty liver disease histology (lobular)	rs4237591-G	0.35	2.00E-06	0.33
	26634245	Lutz SM	Post bronchodilator FEV1/FVC ratio in COPD	rs1942108-A	0.49	5.00E-06	0.011
	27863252	Astle WJ	Lymphocyte counts	rs7939778-A	0.26	2.00E-10	0.026
	27863252	Astle WJ	Reticulocyte fraction of red cells	rs72996113-T	0.1	2.00E-17	0.051
	27863252	Astle WJ	Reticulocyte count	rs72996113-T	0.1	1.00E-17	0.052
	27863252	Astle WJ	Plateletcrit	rs1111890-G	0.37	2.00E-13	0.028

	27863252	Astle WJ	Red cell distribution width	rs717662-T	0.11	4.00E-12	0.04
	28898252	Wheeler E	Glycated hemoglobin levels	rs11224302-C	0.87	5.00E-07	0.015
	25673412	Shungin D	Waist-hip ratio	rs1394461-C	0.25	4.00E-08	0.035
	27089181	Okbay A	**Neuroticism**	rs2458167-A	0.3	2.00E-07	0.019
	29273806	Demenais F	Asthma	rs3758697-A	0.3	2.00E-06	1.33
	27089181	Okbay A	**Depression**	rs1690818-T	NR	6.00E-08	5.405
chr13 locus	25935106	Kim KW	Recalcitrant atopic dermatitis	rs9540294-G	0.08	1.00E-08	2.655
	25760438	Anderson D	Type 2 diabetes	rs11454281-?	NR	9.00E-06	0.6
	22318345	Cha PC	Gallbladder cancer	rs7504990-A	0.21	7.00E-08	6.95
	23251661	Comuzzie AG	Obesity-related traits	rs4940203-A	0.29	5.00E-07	0.05
	27149984	Degenhardt F	Coenzyme Q10 levels	rs74681568-G		2.00E-07	0.116
	24009623	Jiang J	Response to mTOR inhibitor (everolimus)	rs1460196-?	NR	9.00E-06	0.27
	23192594	Velez Edwards DR	Body mass index (ever vs never smoking interaction)	rs11876941-A	NR	5.00E-06	0.003
	28199695	Jones AV	Mosquito bite size	rs141670172-T	NR	7.00E-06	0.257
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs7245004-A	NR	7.00E-14	0.022
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs8083850-A	NR	6.00E-13	0.021
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs11663156-T	NR	1.00E-08	0.019
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs11665242-A	NR	9.00E-16	0.024
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs12607356-A	NR	2.00E-09	0.018
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs12960505-T	NR	2.00E-15	0.024
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs10502966-A	NR	1.00E-11	0.02
DCC	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs1972044-A	NR	7.00E-14	0.023
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs6508220-A	NR	5.00E-15	0.023
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs7506451-A	NR	3.00E-10	0.019
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs10221412-T	NR	4.00E-11	0.02
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs1078459-T	NR	3.00E-11	0.02
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs1367633-A	NR	2.00E-08	0.017
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs17417046-T	NR	3.00E-13	0.021
	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs4277413-A	NR	5.00E-13	0.021
	29147026	Chen CH	**Putamen volume**	rs11660938-?	NR	4.00E-12	41.504
	29187730	Ward J	**Mood instability**	rs8084280-T	0.51	3.00E-09	0.05
	29187730	Ward J	**Mood instability**	rs8084280-T	0.51	1.00E-07	0.047
	26025379	Ng E	Nickel levels	chr18:51115162-?		3.00E-06	1.51
	22318345	Cha PC	Gallbladder cancer	rs975334-C	0.16	9.00E-07	8.3
	22318345	Cha PC	Gallbladder cancer	rs13294589-G	0.11	2.00E-06	12.78

	22318345	Cha PC	Gallbladder cancer	rs6869388-C	0.04	7.00E-06	72.7
	22318345	Cha PC	Gallbladder cancer	rs10953615-C	0.11	9.00E-06	7.51
	21529783	Heath AC	**Alcoholism**	rs768048-C	0.12	8.00E-06	0.11
	28317148	Campo C	Bortezomib-induced peripheral neuropathy in multiple myeloma	rs17748074-A	NR	9.00E-06	1.96
	27393504	Zanetti KA	**Smoking behaviour (cigarettes smoked per day)**	rs1372626-?	NR	3.00E-06	
	28892062	Akiyama M	Body mass index	rs4129322-A	0.1	4.00E-08	0.018
	25524916	Palmer ND	Glucose homeostasis traits	rs2339345-?		8.00E-06	0.25
	29186694	Lam M	**Cognitive ability**	rs1431196-?	NR	4.00E-07	5.073
	29186694	Lam M	**Cognitive ability (multi-trait analysis)**	rs1431196-?	NR	6.00E-11	6.546
	26491034	Sarzynski MA	Response to exercise (triglyceride level interaction)	rs3906453-A	0.76	3.00E-06	0.39
	27015805	Pilling LC	Parental longevity (combined parental age at death)	rs67163261-?	NR	1.00E-06	0.039
	29292387	Turley P	**Depressive symptoms**	rs11663393-G	0.55	6.00E-09	0.014
	29292387	Turley P	**Depressive symptoms (multi-trait analysis)**	rs8084351-G	0.49	2.00E-12	0.015
	29292387	Turley P	**Subjective well-being (multi-trait analysis)**	rs8084351-G	0.48	4.00E-12	0.013
	29292387	Turley P	**Neuroticism (multi-trait analysis)**	rs8084351-G	0.49	1.00E-13	0.021
	25017104	Kottyan LC	Eosinophilic esophagitis	rs9956738-?	0.01	4.00E-07	2.472
	27002377	Irvin MR	Response to fenofibrate (total cholesterol levels)	chr18:50464126-?	NR	2.00E-06	0.017
	26198764	Goes FS	**Schizophrenia**	rs4632195-T	NR	4.00E-06	1.05
	26586795	Phipps AI	Survival in colorectal cancer (distant metastatic)	rs1372474-G	0.1	2.00E-06	1.53
	26586795	Phipps AI	Survival in colorectal cancer (distant metastatic)	rs1442089-C	0.09	2.00E-06	1.56
	28991256	Li Z	**Schizophrenia**	rs4632195-T	NR	2.00E-06	1.051
	27089181	Okbay A	**Depression**	rs62100776-A	NR	1.00E-08	5.664
	25607358	Hibar DP	**Subcortical brain region volumes**	rs62097986-A	0.44	1.00E-13	30.28
	27089181	Okbay A	**Neuroticism**	rs4632195-T	0.52	5.00E-07	0.018
	27225129	Okbay A	**Educational attainment (years of education)**	rs62100765-T	0.42	1.00E-10	0.015
PSMD14	26989097	Coleman JR	**Response to cognitive-behavioural therapy in anxiety disorder**	rs13432654-?	0.09	8.00E-06	
	25644384	Davies G	**Cognitive ability**	rs2303319-?	NR	9.00E-06	0.045
	25201988	Rietveld CA	**Educational attainment**	rs7309-A	0.49	2.00E-07	0.022
ACK3/ COQ8A	27863252	Astle WJ	Sum neutrophil eosinophil counts	rs17592479-A	0.51	7.00E-13	0.026
	27863252	Astle WJ	Sum basophil neutrophil counts	rs17592479-A	0.51	4.00E-13	0.026
	27863252	Astle WJ	Granulocyte count	rs17592479-A	0.51	1.00E-12	0.025
	27863252	Astle WJ	Granulocyte percentage of myeloid white cells	rs17592479-A	0.51	4.00E-12	0.025
	27863252	Astle WJ	Lymphocyte percentage of white cells	rs2297412-G	0.48	3.00E-11	0.024
	27863252	Astle WJ	Neutrophil count	rs17592479-A	0.51	3.00E-13	0.026

	27863252	Astle WJ	Neutrophil percentage of white cells	rs6426558-T	0.51	3.00E-14	0.027
CEP57	29326435	Hill WD	**Intelligence (multi-trait analysis)**	rs644799-A	NR	1.00E-09	0.018
	29186694	Lam M	**Cognitive ability (multi-trait analysis)**	rs644799-?	NR	2.00E-09	6.011
	24390342	Okada Y	Rheumatoid arthritis	rs4409785-C	0.15	1.00E-11	1.12
	24390342	Okada Y	Rheumatoid arthritis	rs4409785-C	0.17	4.00E-09	1.12
FAM76B	29083406	Ferreira MA	Allergic disease (asthma, hay fever or eczema)	rs59593577-C	0.87	2.00E-11	1.053

4.4.1.18 Supplementary Table 8

Previously reported suicidal behaviour-associated SNPs

Previously reported associations							
PMID	FIRST AUTHOR	DISEASE/ TRAIT	CHR	BP	STRONGEST SNP-RISK ALLELE	RAF	P-VALUE
20877300	Perroud N	Suicidal ideation	4	21,475,367	rs358592-?	0.7	3.00E-06
20877300	Perroud N		8	28,065,871	rs4732812-?	0.73	3.00E-06
20877300	Perroud N		9	72,272,787	rs11143230-C	0.35	8.00E-07
20877300	Perroud N		9	72,272,787	rs11143230-C	0.35	7.00E-06
21041247	Perlis RH	Suicide risk in mood disorders	2	46,093,955	rs12373805-A	NR	9.00E-06
21041247	Perlis RH		10	30,201,775	rs2462021-C	NR	8.00E-06
21041247	Perlis RH		10	95,362,484	rs4918918-T	NR	3.00E-06
21041247	Perlis RH		21	39,649,825	rs10854398-C	NR	6.00E-06
21423239	Willour VL	Suicide attempts in bipolar disorder	2	112,496	rs300774-A	0.18	**5.00E-08**
21423239	Willour VL		11	33,566,664	rs10437629-?	NR	4.00E-06
21423239	Willour VL		12	128,610,527	rs7296262-T	0.51	1.00E-06
24964207	Mullins N	Suicide attempts in depression or bipolar disorder	4	15,993,502	rs17387100-G	0.08	8.00E-07
24964207	Mullins N		7	150,339,575	rs17173608-G	0.06	2.00E-07
24964207	Mullins N		11	113,249,490	rs3781878-?	0.73	2.00E-06
24964207	Mullins N		12	66,422,359	rs10748045-G	0.35	1.00E-06
25917933	Zai CC	Suicide in bipolar disorder	8	56,592,754	rs2610025-A		5.00E-06
25917933	Zai CC		8	79,191,197	rs10448044-C		3.00E-06
25917933	Zai CC		10	32,704,340	rs7079041-A		2.00E-06

25917933	Zai CC		18	68,547,459	rs7244261-T		4.00E-06
26079190	Galfalvy H		1	213,424,498	rs320461-A	0.23	4.00E-06
26079190	Galfalvy H	Suicide	7	35,254,361	rs336284-A	0.45	2.00E-07
26079190	Galfalvy H		8	10,677,867	rs7011192-A	0.11	4.00E-06
26079190	Galfalvy H	Suicide attempts in major depressive disorder	2	115,733,670	rs4308128-A	0.43	4.00E-06
26079190	Galfalvy H		10	70,758,081	rs6480463-G	0.38	2.00E-06
26079190	Galfalvy H	Suicide behaviour	14	24,146,226	rs4575-G	0.27	8.00E-06
26079190	Galfalvy H		15	34,813,456	rs11852984-C	0.2	2.00E-06
26079190	Galfalvy H		8	98,883,177	rs3019286-A	0.36	8.00E-06
26079190	Galfalvy H	Suicide ideation score in major depressive disorder	1	190,088,550	rs2419374-A	0.19	1.00E-06
26079190	Galfalvy H		20	8,233,139	rs6055685-A	0.18	8.00E-07
26079190	Galfalvy H		5	112,267,016	rs13358904-G	0.21	5.00E-06
30116032	Erlangsen A	Suicide attempts, with and without psychiatric diagnoses	1	81,428,767	rs72940689-A	0.06	2.78E−07
30116032	Erlangsen A		5	153,290,253	rs2085865-A	0.31	1.06E−07
30116032	Erlangsen A		9	18,290,857	rs7862648-G	0.22	9.80E−07
30116032	Erlangsen A		12	32,640,591	rs112595860-G	0.22	4.55E−07
30116032	Erlangsen A		20	47,193,719	rs4809706-G	0.63	2.19E−07
30116032	Erlangsen A		22	36,255,928	rs150801052-A	0.02	6.78E−07
28902444	Stein MB		6	84,770,179	rs2497117-A		1.58E-08
28902444	Stein MB		6	84,771,964	rs2497118-A		1.70E-08
28902444	Stein MB	Suicide attempt in USA soldiers	6	84,772,469	rs2480192-T		1.32E-08
28902444	Stein MB		6	84,772,961	rs2497119-A		1.18E-08
28902444	Stein MB		6	84,794,805	rs142060512-T		3.55E-08
28902444	Stein MB		6	84,803,043	rs116923768-A		2.02E-09
28902444	Stein MB		6	84,809,043	chr6_84809043_D-I2		4.68E-09

28902444	Stein MB		6	84,820,786	rs116878613-T	4.12E-09
28902444	Stein MB		6	84,898,516	rs117975834-C	2.12E-08
28902444	Stein MB		6	84,914,920	rs78022606-A	4.14E-08
28902444	Stein MB		6	84,935,294	chr6_84935294_D-I5	4.96E-10
28902444	Stein MB		6	84,935,441	rs12524136-T	5.24E-10
30655502	Levey DF	suicide attempt severity	15		rs72740088-T	7.49E-08
30655502	Levey DF		1		rs61520094-T	8.14E-07
30655502	Levey DF		9		rs10867557-A	4.73E-07
30655502	Levey DF		12		rs1677091-A	1.07E-08
30655502	Levey DF		12		rs860447-T	6.10E-07

4.5 Novel genome-wide associations for anhedonia, genetic correlation with psychiatric disorders, and polygenic association with brain

Ward et al. *Translational Psychiatry* (2019)9:327
https://doi.org/10.1038/s41398-019-0635-y

Translational Psychiatry

ARTICLE Open Access

Novel genome-wide associations for anhedonia, genetic correlation with psychiatric disorders, and polygenic association with brain structure

Joey Ward[1], Laura M. Lyall[1], Richard A. I. Bethlehem[2], Amy Ferguson[1], Rona J. Strawbridge[1,3], Donald M. Lyall[1], Breda Cullen[1], Nicholas Graham[1], Keira J. A. Johnston[1], Mark E. S. Bailey[4], Graham K. Murray[2] and Daniel J. Smith[1]

Abstract

Anhedonia is a core symptom of several psychiatric disorders but its biological underpinnings are poorly understood. We performed a genome-wide association study of state anhedonia in 375,275 UK Biobank participants and assessed for genetic correlation between anhedonia and neuropsychiatric conditions (major depressive disorder, schizophrenia, bipolar disorder, obsessive compulsive disorder and Parkinson's Disease). We then used a polygenic risk score approach to test for association between genetic loading for anhedonia and both brain structure and brain function. This included: magnetic resonance imaging (MRI) assessments of total grey matter volume, white matter volume, cerebrospinal fluid volume, and 15 cortical/subcortical regions of interest; diffusion tensor imaging (DTI) measures of white matter tract integrity; and functional MRI activity during an emotion processing task. We identified 11 novel loci associated at genome-wide significance with anhedonia, with a SNP heritability estimate (h_2SNP) of 5.6%. Strong positive genetic correlations were found between anhedonia and major depressive disorder, schizophrenia and bipolar disorder; but not with obsessive compulsive disorder or Parkinson's Disease. Polygenic risk for anhedonia was associated with poorer brain white matter integrity, smaller total grey matter volume, and smaller volumes of brain regions linked to reward and pleasure processing, including orbito-frontal cortex. In summary, the identification of novel anhedonia-associated loci substantially expands our current understanding of the biological basis of state anhedonia and genetic correlations with several psychiatric disorders confirm the utility of this phenotype as a transdiagnostic marker of vulnerability to mental illness. We also provide the first evidence that genetic risk for state anhedonia influences brain structure, including in regions associated with reward and pleasure processing.

Introduction

Anhedonia refers to reduced capacity to experience pleasure in situations that individuals would normally enjoy, and has been a focus of psychiatry research for decades[1,2]. It constitutes a core symptom of several neuropsychiatric disorders, including major depressive disorder (MDD), schizophrenia, bipolar disorder and obsessive compulsive disorder (OCD), as well as Parkinson's Disease (PD)[3–6]. Along with a direct negative association with quality of life and subjective wellbeing, anhedonia is associated with multiple negative health-related behaviours, such as smoking, illicit drug use, and low physical activity, even in the absence of psychiatric disorder[7,8]. In line with a Research Domain Criteria (RDoC) approach[9], state anhedonia can be measured and studied as a dimensional psychopathological trait.

Anhedonia has been closely linked to the function and structure of reward circuitry in the brain (primarily frontal, striatal and limbic regions). These neurobiological associations are consistent with the view that anhedonia reflects

Correspondence: Daniel J. Smith (daniel.smith@glasgow.ac.uk)
[1]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK
[2]Department of Psychiatry, University of Cambridge, Cambridge, UK
Full list of author information is available at the end of the article
These authors contributed equally: Joey Ward, Laura M. Lyall

© The Author(s) 2019
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

dysfunction in reward processing[10,11]. Measures of anhedonia have been associated with altered functional activity during reward-based tasks within frontal cortical regions (medial frontal cortex) and subcortical striatal regions (caudate and putamen)[12], and with reduced volumes in a similar set of frontal and striatal regions[13,14]. Anhedonia is also associated with reduced white matter integrity[15,16].

The genetic underpinnings of anhedonia are largely unknown. Several GWAS of disorders where anhedonia is a feature have been reported, such as MDD and schizophrenia[17,18]. However, to date only small underpowered GWAS studies of anhedonia have been published. A study of 759 patients with MDD identified 18 SNPs associated with an 'interest-activity' measure of anhedonia[19]. The largest study to date is a mega-analysis of three studies of young people from the UK and Sweden, with a total sample size of 6579; a single locus was associated with anhedonia in the discovery sample, but not in the replication sample[20]. A Finnish study examined genetic associations with physical and social anhedonia, as assessed with the Chapman scales[21], in 3820 individuals but no genome-wide significant loci were identified[22]. Genetic loci associated with anhedonia have therefore not yet been reliably identified in large clinical or general population samples[23], and association tests of genetic risk for anhedonia with brain structure and function have not yet been performed.

Here we report a large GWAS of state anhedonia within the UK Biobank cohort. We also use a polygenic risk score (PRS) approach to assess whether genetic loading for anhedonia is associated with brain structure and brain function.

Methods

UK Biobank sample

UK Biobank is a large cohort of over half a million UK residents, aged between 39 and 73 years at baseline assessment[24]. The cohort was designed to assess how genetic, environmental and lifestyle factors influence a range of morbidities in middle and older age. Baseline assessments occurred over a 4-year recruitment 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 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 have over 95% of content in common[22]. Pre-imputation quality control, imputation (both 1000Genomes and HRC Reference Panels) and post-imputation cleaning were conducted centrally by UK Biobank (described in the UK Biobank release documentation, please see URLs in appendix for details).

Phenotyping

As part of the comprehensive baseline assessment participants were asked: *"Over the past two weeks, how often have you had little interest or pleasure in doing things?"* (Data field 2060). Respondents could choose from the following answers: *"not at all"; "several days"; "more than half the days"; and "nearly every day"*. These responses were coded as 0, 1, 2 and 3 respectively. This question on anhedonia is derived from the Patient Health Questionnaire-9 (PHQ-9), a well-validated screening instrument for MDD[25]. To maximise numbers available for downstream magnetic resonance imaging (MRI) analyses, we excluded from the primary GWAS those participants with any available MRI data ($n = 20{,}174$). Additional exclusion criteria for the GWAS included individuals in whom: over 10% of genetic data were missing; self-reported sex did not match genetic sex; sex chromosome aneuploidy was reported; where heterozygosity value was a clear outlier; and participants not of European ancestry ($n = 73{,}385$).

Genetic association and heritability

Genetic association with the measure of anhedonia was performed using BOLT-LMM[26,27], which accounts for population structure and sample relatedness by including a genetic relatedness matrix within the models. Models were further adjusted for age, sex, and genotyping array. SNPs included in the analysis were filtered by MAF > 0.01, Hardy–Weinberg Equilibrium $p > 1 \times 10^{-6}$, and imputation score > 0.3. BOLT-LMM was also used to provide a SNP-heritability estimate and an estimate of λ_{GC}.

Genetic correlations

Linkage Disequilibrium Score Regression (LDSR)[28] was carried out using LDSC on GWAS summary statistics from several published studies, to obtain genetic correlations with psychiatric disorders where anhedonia is known to be a feature (MDD[29], schizophrenia[30], bipolar disorder[30], and OCD[31]), as well as Parkinson's Disease[32]. We left the intercept unconstrained to allow for sample overlap between the sample used here and the outputs of the other GWAS.

Polygenic risk score generation

Polygenic risk scores (PRS) were created using LDpred[33]. LDpred differs from the more common

pruning and threshold (P&T) PRS because it generates a single risk score for the trait of interest derived from as many loci as possible. A training set was created to obtain the LD structure by using 1000 unrelated Biobank participants who had passed the same genetic QC as those used in the GWAS (but who were excluded from the GWAS because they did not respond to the anhedonia question and had not provided brain imaging data). These training data were then used for the construction of anhedonia PRS in those participants for who brain imaging data were available.

Brain imaging variables

A number of structural and functional brain MRI measures have been made available by UK Biobank as Imaging Derived Phenotypes (IDPs)[34]. These measures are: total volume (mm^3) of brain grey matter, white matter and cerebrospinal fluid (CSF), each normalised for head size; volumes (grey matter or total) of 15 cortical and subcortical regions of interest (ROIs); diffusion tensor imaging (DTI) measures of white matter integrity (fractional anisotropy (FA) and mean diffusivity (MD)); and functional MRI activity during an emotion processing task (the Hariri face shape task)[35] in an amygdala mask and group-defined mask consisting of occipito-temporal and amygdala regions. For a more detailed description of these variables, and for details of MRI acquisition and pre-processing and IDP selection, please see supplementary methods.

Polygenic risk score and brain imaging analyses

MRI data were available for 20,174 UK Biobank participants. PRS/MRI analyses were conducted in a subset of 17,120 participants who had available MRI data and who were not included in the GWAS, after exclusion of participants who did not meet genetic quality control criteria ($n = 2479$) or who self-reported a developmental or neurological disorder at either the baseline assessment or the imaging visit ($n = 575$) (please see Table S1 for exclusions). For each MRI outcome, data points with values more than 3 standard deviations from the sample mean were excluded.

Models were adjusted for age at MRI visit, (age at MRI)[2], sex, genotype array, the first eight genetic principal components, and lateral, transverse and longitudinal scanner position covariates. Total tissue volume measures (total grey matter, white matter and ventricular CSF volumes) were normalised for head size prior to adjustment for the above covariates. ROI analyses were additionally adjusted for total brain volume (calculated by summing total grey matter, white matter and ventricular cerebrospinal fluid (CSF) volume); and fMRI analyses were also adjusted for head motion during the emotion processing task. False Discovery Rate (FDR) correction was applied[36,37].

Results

Demographics

The GWAS was performed on 375,724 UK Biobank participants, of whom 203,322 (54.1%) were female. The age of the sample ranged from 39 to 73 and the mean age was 57 years (S.D. = 8.01). In response to the anhedonia question, 299,232 (79.64%) answered *"not at all"*; 60,212 (16.0%) reported *"several days"*; 9405 (2.5%) reported *"more than half the days"* and 6876 (1.8%) reported *"nearly every day"*.

MRI analyses were conducted in 17,120 participants, 52.4% (8978) of whom were female. The mean age (at the time of MRI) of these participants was 62.7 years (S.D. = 7.46; range = 45–80 years). Of the 16,783 participants with available MRI data who answered the anhedonia question, 13,810 (82.3%) responded *"not at all"*, 2469 (14.7%) reported feelings of anhedonia for *"several days"*, 297 (1.8%) for *"more than half the days"*, and 207 (1.2%) *"nearly every day"*.

Genome-wide association study findings

GWAS results are presented as a Manhattan plot in Fig. 1, and details of genome-wide significant loci are provided in Table S2. In all, there were 1100 SNPs that were genome-wide significant ($p < 5 \times 10^{-8}$), and, of these, represented 11 independent loci on 9 different chromosomes (Table S2; Figs. S1–S11). An independent signal was defined as the region of $r^2 > 0.1$ within a 500 MB window from the most significant SNP below genome-wide significance.

Some inflation of the GWAS results was observed ($\lambda_{GC} = 1.15$), however considering the sample size this is expected to have had a negligible impact on findings. There was evidence for a polygenic component (LDSR intercept = 1.03, S.E. = 0.005) and no evidence for undue inflation of the test statistics due to unaccounted population stratification.

Genome-wide association study replication

To help validate the phenotype, we attempted a replication in the sample that was excluded from the GWAS. Only one of the eleven loci identified achieved nominal significance (CHR6:27731402, $\beta = -0.02$, $p = 0.015$). This would not survive correction for multiple testing. Of the 11 top hit SNPs, 8 of 11 coefficients were in the same direction as in the discovery set (Table S3).

Additionally, we determined if the PRS of anhedonia in those excluded from the GWAS was associated with their state anhedonia scores. The model was adjusted for the first 4 genetic principle components, genotyping chip, age, sex and Townsend score. We detected a small but highly significant association between the anhedonia PRS and the anhedonia phenotype scores ($\beta = 0.0007$, $p = 1.17 \times 10^{-14}$).

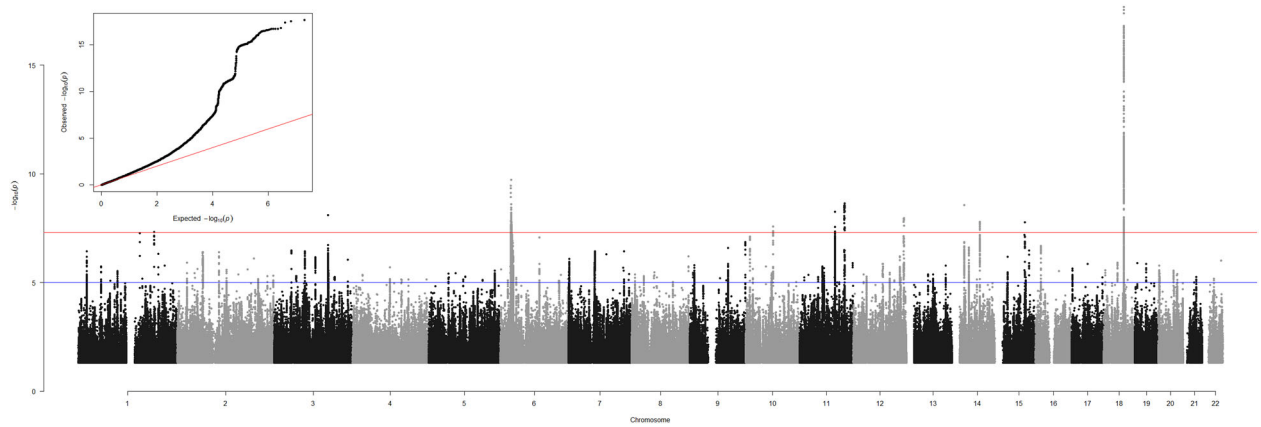


Fig. 1 Anhedonia GWAS results. Results are presented as a Manhattan Plot and as a QQ plot (inset). Association analysis *p*-values for each SNP are plotted (as –log10(*p*)) vs. chromosomal position. The red and blue lines indicate the genome-wide significant and suggestive *p*-value thresholds, respectively. The QQ plot shows observed vs expected *p*-values for every SNP.

Genetic correlations

Using LDSR we assessed whether genetic predisposition to anhedonia overlapped with that for several psychiatric disorders or traits. Anhedonia had significant genetic correlation with MDD (r_g 0.77, $q = 1.84 \times 10^{-139}$), schizophrenia ($r_g$ 0.28, $q = 5.28 \times 10^{-15}$) and bipolar disorder (r_g 0.12, $q = 0.002$) but not with OCD or Parkinson's disease ($q = 0.6$ and $q = 0.97$ respectively) (Table 1).

PRS and brain MRI structural and functional outcomes

Associations between PRS for anhedonia and total brain tissue volumes are presented in Table S4. Greater polygenic risk for anhedonia was associated with lower total grey matter volume, but not with total white matter volume or with total ventricular CSF volume.

Associations were then assessed between PRS for anhedonia and volumes of 15 cortical and subcortical regions of interest (ROIs were derived a priori from meta-analysis and literature review; please see Table S5 and Supplementary Methods). Greater genetic risk score for anhedonia was associated with smaller volumes for insular cortex, orbitofrontal cortex, middle frontal gyrus and anterior temporal fusiform cortex. Uncorrected associations were also observed with nucleus accumbens, medial frontal cortex and caudate volumes, but these did not survive FDR correction. The locations of the ROIs showing significant association with the PRS are displayed in Supplementary Fig. S12, and a point-range plot showing the association evidence for ROI volumes is displayed in Fig. 2a.

In subsequent analyses of association with white matter integrity, greater PRS for anhedonia was not associated with general factors of either higher mean diffusivity (MD) or fractional anisotropy (FA) (Table S6). PRS associations were assessed with 15 individual white matter tracts (bilateral tracts were combined in the same models and models were adjusted for hemisphere; see Supplementary Methods, Supplementary Tables S7 and S8 for tract-specific FA and MD results, respectively). Higher PRS for anhedonia was associated with lower FA in four of the fifteen tracts (anterior thalamic radiation, forceps minor, medial lemniscus, posterior thalamic radiation) (Fig. 2b). Higher PRS for anhedonia was also associated with higher MD in 9 of the 15 tracts (all apart from acoustic radiation, corticospinal tract, forceps major, inferior longitudinal fasciculus, medial lemniscus and parahippocampal part of cingulum) (Fig. 2c). There was no association between PRS for anhedonia and functional MRI activity during the emotion-processing task (median Blood Oxygen Level Dependent signal for the face vs. shape contrast) (Table S9).

Table 1 Genetic correlations of anhedonia with psychiatric phenotypes.

Trait	r_g	S.E.	Z	p	q
MDD	0.771	0.0306	25.2033	3.68E-140	1.84E-139
Schizophrenia	0.28	0.0353	7.9345	2.11E-15	5.28E-15
Bipolar	0.122	0.038	3.2187	0.00129	0.002
Parkinson's Disease	0.0584	0.0837	0.6978	0.4853	0.60
OCD	−0.0023	0.0618	−0.0372	0.9704	0.97

rg genetic correlation with mood instability, *S.E.* 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.

Discussion

These analyses represent the largest genetic association study of state anhedonia performed to date. We identified eleven genetic loci associated with anhedonia in the UK

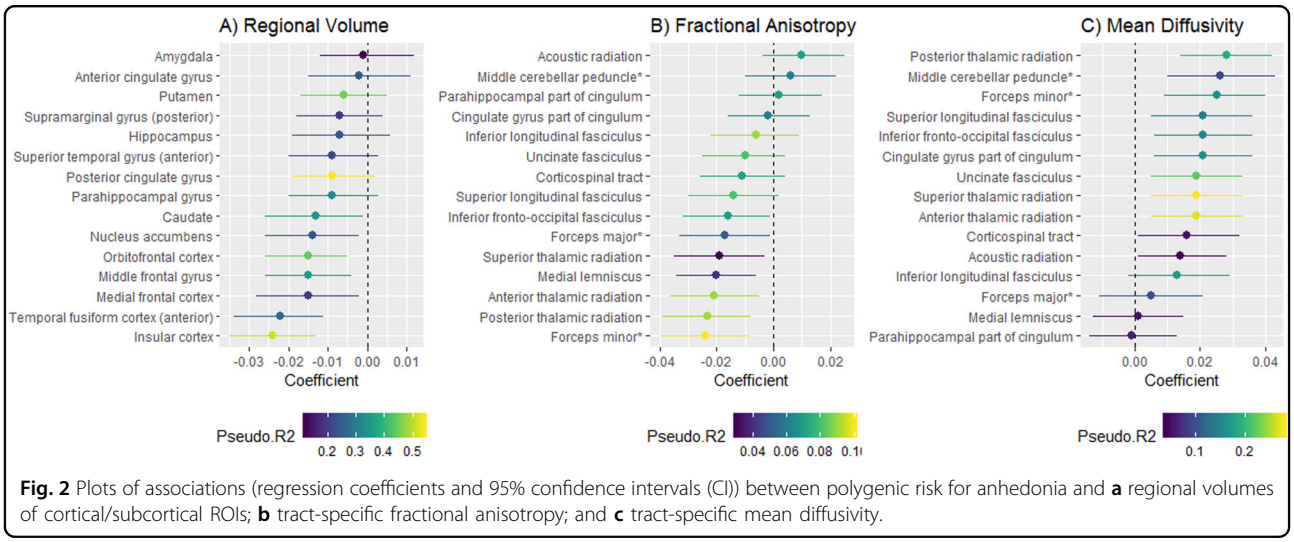


Fig. 2 Plots of associations (regression coefficients and 95% confidence intervals (CI)) between polygenic risk for anhedonia and **a** regional volumes of cortical/subcortical ROIs; **b** tract-specific fractional anisotropy; and **c** tract-specific mean diffusivity.

general population. Within each associated region there were a number of genes that could have a functional impact on anhedonia and the pleasure cycle (considered in detail below). Consistent with an RDoC approach focusing on transdiagnostic symptoms and traits, we found strong genetic correlations between anhedonia and MDD, schizophrenia and bipolar disorder, but not between anhedonia and OCD or Parkinson's Disease. Despite the lack of correlation with Parkinson's Disease, several of the loci identified in our GWAS include genes with known association with Parkinson's Disease (see below). We also report the first investigation of associations between genetic loading for anhedonia and both brain structure and brain function.

Genes within anhedonia-associated loci

There are multiple genes in each QTL identified in the GWAS. Here we discuss potentially relevant genes based on their known function in the scientific literature. Within the chromosome 1 locus there are multiple *RGS* genes, most notably *RGS1* and *RGS2*, encoding regulators of G-protein signalling that show prominent expression in the brain[38]. *RGS2* has previously been identified as a modulator of *LRRK2*[39] expression, a gene known to be a genetic cause of Parkinson's Disease. *RGS2* has also been associated with symptom severity in schizophrenia[40] and lower expression of *RGS2* may be related to depression-like behaviours in animal models[41].

EPHB1 on chromosome 3 encodes an ephrin receptor tyrosine kinase identified in a GWAS of antidepressant response[42], and is associated with symptoms of schizophrenia in Chinese Han populations[43], and with susceptibility to Parkinson's Disease[44].

At the more centromeric locus on chromosome 11, *GRM5* encodes a metabotropic glutamate receptor that has been extensively studied in relation to MDD[45,46] and schizophrenia[47,48]. Assessments in mice have also found that agonists of *GRM5* attenuate Parkinsonian motor deficits via striatal dopamine depletion[49]. The Genotype Tissue Expression (GTEx) database[38] shows prominent expression of *GRM5* in the brain, especially within the nucleus accumbens, a region with a major role in the prediction of reward[50,51].

Another likely candidate gene at this chromosome 11 locus is *DISC1FP1*. *DISC1* (Disrupted in Schizophrenia 1), a gene located on chromosome 1, is part of a chromosome 1:11 translocation that increases risk of schizophrenia, schizoaffective disorder and bipolar disorder[52]. *DISC1FP1* is disrupted by this translocation, impacting on both intracellular NADH oxidoreductase activities and protein translation[53].

At the more telomeric locus on chromosome 11, *NCAM1* (neural cell adhesion molecule 1) has been implicated as a potential link between depressive symptoms and brain structure[54], specifically decreased FA. *NCAM1* has been reported to be present at increased levels in the amygdala[38], another brain region associated with the pleasure cycle, in depressed subjects[55]. However, we did not find any association between PRS for anhedonia and either amygdala volume or functional activity in the amygdala during an emotion processing task. It is possible that *NCAM1* may exert its effects via expression in other brain regions.

Another gene of interest at this second chromosome 11 locus is *DRD2*, encoding the dopamine D2 receptor. There is an extensive literature on the importance of *DRD2* for the psychopharmacology of both schizophrenia and MDD[56,57].

The chromosome 12 locus contains only *LOC* and *LINC* genes. These encode different classes of functional RNA but little is known about their function beyond possible post-transcriptional regulation of other gene products.

The more centromeric locus on chromosome 14 contains *PRKD1*, which encodes a serine/threonine-protein

kinase identified in a GWAS of schizophrenia[18]. The more telomeric locus on chromosome 14 contains the gene *SLC8A3*, a gene involved in maintaining Ca^{2+} homoeostasis within a variety of tissues, including neurons. *SLC8A3* may also play a role in Parkinson's Disease[58].

ISLR2 on chromosome 15 is involved in neurodevelopment[59] but its role in psychiatric and neurological traits is not well characterised. Another candidate gene within this locus, *NRG4*, encodes a neuregulin protein that activates type-1 growth factor receptors. Recent work has shown that NRG4 acts a regulator of the growth and elaboration of pyramidal neuron dendrites in the developing neocortex[60]. Pyramidal neurons have been directly implicated in the pathophysiology of schizophrenia[61].

Finally, on chromosome 18, *DCC* (the most significant hit) encodes a netrin 1 receptor with a role in axon guidance, and has been previously reported to be associated with anhedonic phenotypes in mice and humans[62] and potentially with schizophrenia pathogenesis[63]. We have previously identified *DCC* in GWAS of mood instability[64], suicidality[65] and multisite chronic pain[66].

Brain structure and function

Our findings on the relationship between genetic loading for anhedonia and brain structure and function are of considerable interest. Greater levels of anhedonia, in healthy and clinical populations, have been linked to altered functional activity (and less consistently to reduced volume) in frontal/striatal regions involved in reward or pleasure processing[12–14,67–71]. The regions most consistently implicated include nucleus accumbens, caudate, putamen, medial frontal cortex and orbitofrontal cortex.

We found that increased genetic risk for anhedonia was associated with smaller volumes of the orbitofrontal cortex the (Fig. 2a; Supplementary Fig. 1) a region involved in representation of reward value and social reward dependence[70,72]. We also noted uncorrected associations of PRS for anhedonia and smaller volumes of nucleus accumbens (involved in reward value and pleasure processing[69,73]) and medial frontal cortex (linked to pleasure processing[69]), but these associations did not survive correction for multiple comparisons. These cortical/subcortical volume findings are therefore consistent with an association between genetic risk for anhedonia and reward/pleasure processing[73].

Increased genetic risk for anhedonia was also associated with smaller volumes of insular cortex (associated with emotion processing[74]) and fusiform cortex. Most ROIs in our analysis were selected on the basis that they were associated with smaller volumes in MDD vs. healthy controls in the most recent MDD brain structure meta-analysis[75]. Our findings are consistent with two scenarios: either, genes for anhedonia partly mediate the association between the ROI volumes and MDD, or these genes exhibit horizontal pleiotropy and affect both phenotypes via separate mechanisms. Consistent with our finding of widespread associations of genetic risk for anhedonia with cortical/subcortical volumes, we also found an association with reduced total grey matter volume, adjusted for head size.

Evidence of reduced white matter integrity (FA) in several tracts in individuals scoring higher on measures of anhedonia has been reported[16,76]. We found that higher values for the anhedonia PRS were associated with higher MD (reflecting poorer white matter integrity), in most individual tracts, but not in a general factor of MD. Several tracts also showed reduced FA.

We did not, however, find any association between PRS for state anhedonia and functional brain activity. UK Biobank ROIs were selected based on average responses during an emotion processing task; it is therefore plausible that effects would emerge with the use of a reward processing task, or by applying whole-brain voxel-wise analyses, to include further reward/pleasure processing regions.

Strengths and limitations

This study is the largest GWAS of state anhedonia to date, and substantially contributes new knowledge on the biology of this important transdiagnostic symptom. We conducted analyses within a large population-based cohort. Our primary analysis included individuals with mental health histories, but these represent only a small proportion of the total sample, which is an order of magnitude larger than any previous study of this kind. Subclinical anhedonia is common, and associated with increased risk of later mental illness[77]. This idea is reinforced by the significant genetic correlations with other psychiatric disorders. Additionally, we used an ordered ordinal phenotype, resulting in more power to detect associations than with the more common dichotomised ('non-anhedonic' vs. 'anhedonic') analyses[78]. Identification of loci associated with population-level anhedonia may be important from a personalised medicine perspective, for example, in terms of developing stratified medicine approaches to identify individuals at high-risk of developing psychiatric disorders.

Notably, however, the UK Biobank cohort has a degree of selection bias. In general, volunteers are typically healthier and of higher socioeconomic status and higher education level than the general population[79], so reported levels of anhedonia in this sample may be lower than the population rate. Therefore, it is plausible that the strength of associations we identified may be an underestimate of the true population value; there may also be an inflated type 2 error rate.

The measure of anhedonia employed here was a single question from a depression screening instrument, the PHQ-9, assessing frequency of symptoms anhedonia

within the preceding two weeks. This item therefore measures state anhedonia at a single time point, and clearly could be influenced by environmental factors, such as season or current health status[80,81]. However, we have assumed here that despite transient environmental and physical factors, individuals who are prone to trait anhedonia will be more likely to report higher frequency of anhedonia at any given time point, and it seems likely that a recent anhedonia phenotype will be enriched for individuals with stronger genetic predisposition towards anhedonia. In line with this assumption, the anhedonia item of the PHQ-9 at a single time point has demonstrated utility in predicting longitudinal brain structural change[82].

Additionally, our attempts at replication of the GWAS results met with limited success, with no real replication of the top hits. However, the PRS analysis was able to detect a significant correlation with state anhedonia in the replication sample. We attribute the lack of replication success in the GWAS to lack of power (small replication sample size) and small effect sizes for individual SNPs.

The GWAS treated the ordered ordinal responses to the frequency of anhedonia question as linear. This approach is likely to have had a minimal impact on the BOLT-LMM GWAS results. Although the distance between points on the anhedonia scale is not evenly spaced, each point can reasonably be considered to be at least several days of anhedonia in the preceding two weeks than the point before it.

The anhedonia measure did not enable determination of specific anhedonia subtypes. Existing validated instruments typically divide anhedonia into physical and social subscales[21], or into anticipatory vs. consummatory components of pleasure[83]. Future studies using more detailed anhedonia scales may be of use in examining the extent of genetic overlap between anhedonia subtypes.

Conclusion

We report the largest GWAS to date of state anhedonia, a common symptom associated with several psychiatric disorders. We identified 11 novel genetic loci and our findings indicate substantial genetic overlap between anhedonia and several psychiatric disorders including MDD, schizophrenia and bipolar disorder. PRS analyses revealed association between genetic loading for anhedonia and smaller volumes of several brain regions, and poorer white matter integrity. Taken together, these findings provide important insights into the neurobiology of an important but under-studied psychiatric symptom and strongly support the proposition that genetic predisposition to anhedonia influences brain structure and function.

URLs

UK Biobank genetic data release information - https://data.bris.ac.uk/datasets/3074krb6t2frj29yh2b03x3wxj/UK%20Biobank%20Genetic%20Data_MRC%20IEU%20Quality%20Control%20version%201.pdf. Genotyping and quality control information – http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf.

Acknowledgements

J.W. is supported by the JMAS Sim Fellowship for depression research from the Royal College of Physicians of Edinburgh (173558). A.F. is supported by an MRC Doctoral Training Programme Studentship at the University of Glasgow (MR/K501335/1). R.J.S. is supported by a UKRI Innovation- HDR-UK Fellowship (MR/S003061/1). K.J.A.J. is supported by an MRC Doctoral Training Programme Studentship at the Universities of Glasgow and Edinburgh. D.J.S. acknowledges the support of a Lister Prize Fellowship (173096) and the MRC Mental Health Data Pathfinder Award (MC_PC_17217). R.A.I.B. is supported by a British Academy Post-Doctoral Fellowship.

Author details

[1]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK. [2]Department of Psychiatry, University of Cambridge, Cambridge, UK. [3]Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden. [4]School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

Code availability

BOLT-LMM software is available from https://data.broadinstitute.org/alkesgroup/BOLT-LMM/downloads/. LDpred software is available from https://bitbucket.org/bjarni_vilhjalmsson/ldpred/src/master/

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41398-019-0635-y).

Received: 16 August 2019 Revised: 18 September 2019 Accepted: 20 October 2019
Published online: 04 December 2019

References

1. Andreasen, N. C. & Olsen, S. Negative v positive schizophrenia. Definition and validation. *Arch. Gen. Psychiatry* **39**, 789–794 (1982).
2. Heinz, A., Schmidt, L. G. & Reischies, F. M. Anhedonia in schizophrenic, depressed, or alcohol-dependent patients-neurobiological correlates. *Pharmacopsychiatry* **27**(Suppl 1), 7–10 (1994).
3. Dichter, G. Anhedonia in Unipolar Major Depressive Disorder: A Review. *Open Psychiatry J.* **4**, 1–9 (2010).
4. Pelizza, L. & Ferrari, A. Anhedonia in schizophrenia and major depression: state or trait? *Ann. Gen. Psychiatry* **8**, 22 (2009).
5. Abramovitch, A., Pizzagalli, D. A., Reuman, L. & Wilhelm, S. Anhedonia in obsessive-compulsive disorder: beyond comorbid depression. *Psychiatry Res* **216**, 223–229 (2014).
6. Loas, G., Krystkowiak, P. & Godefroy, O. Anhedonia in Parkinson's disease: an overview. *J. Neuropsychiatry Clin. Neurosci.* **24**, 444–451 (2012).
7. Leventhal, A. M. et al. Anhedonia associated with stimulant use and dependence in a population-based sample of American adults. *Exp. Clin. Psychopharmacol.* **18**, 562–569 (2010).
8. Leventhal, A. M. Relations between anhedonia and physical activity. *Am. J. health Behav.* **36**, 860–872 (2012).
9. Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* **11**, 126 (2013).

10. Der-Avakian, A. & Markou, A. The neurobiology of anhedonia and other reward-related deficits. *Trends Neurosci.* **35**, 68–77 (2012).
11. Heshmati, M. & Russo, S. J. Anhedonia and the brain reward circuitry in depression. *Curr. Behav. Neurosci. Rep.* **2**, 146–153 (2015).
12. Zhang, B. et al. Mapping anhedonia-specific dysfunction in a transdiagnostic approach: an ALE meta-analysis. *Brain Imaging Behav.* **10**, 920–939 (2016).
13. Pizzagalli, D. A. et al. Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *Am. J. Psychiatry* **166**, 702–710 (2009).
14. Wacker, J., Dillon, D. G. & Pizzagalli, D. A. The role of the nucleus accumbens and rostral anterior cingulate cortex in anhedonia: integration of resting EEG, fMRI, and volumetric techniques. *Neuroimage* **46**, 327–337 (2009).
15. Henderson, S. E. et al. A preliminary study of white matter in adolescent depression: relationships with illness severity, anhedonia, and irritability. *Front Psychiatry* **4**, 152 (2013).
16. Keedwell, P. A. et al. Cingulum white matter in young women at risk of depression: the effect of family history and anhedonia. *Biol. Psychiatry* **72**, 296–302 (2012).
17. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
18. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
19. Ren, H. et al. Genes associated with anhedonia: a new analysis in a large clinical trial (GENDEP). *Translational. Psychiatry* **8**, 150 (2018).
20. Pain, O. et al. Genome-wide analysis of adolescent psychotic-like experiences shows genetic overlap with psychiatric disorders. *Am. J. Med. Genet. Part B, Neuropsychiatr. Genet. : Off. Publ. Int. Soc. Psychiatr. Genet.* **177**, 416–425 (2018).
21. Chapman, L. J., Chapman, J. P. & Raulin, M. L. Scales for physical and social anhedonia. *J. Abnorm. Psychol.* **85**, 374–382 (1976).
22. Ortega-Alonso, A. et al. Genome-Wide Association Study of Psychosis Proneness in the Finnish Population. *Schizophrenia Bull.* **43**, 1304–1314 (2017).
23. Ronald, A. & Pain, O. A systematic review of genome-wide research on psychotic experiences and negative symptom traits: new revelations and implications for psychiatry. *Hum. Mol. Genet.* **27**(R2), R136–R152 (2018).
24. Sudlow, C. 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.* **12**, e1001779 (2015).
25. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern Med.* **16**, 606–613 (2001).
26. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet* **47**, 284–290 (2015).
27. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
28. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
29. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
30. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705–1715.e1716 (2018).
31. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
32. Nalls M. A. et al. Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk. *bioRxiv* 2019: 388165.
33. Vilhjálmsson Bjarni, J. et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
34. Elliott L. T. et al. The genetic basis of human brain structure and function: 1,262 genome-wide associations found from 3,144 GWAS of multimodal brain imaging phenotypes from 9,707 UK Biobank participants. *bioRxiv* 2017: 178806.
35. Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F. & Weinberger, D. R. The amygdala response to emotional stimuli: a comparison of faces and scenes. *Neuroimage* **17**, 317–323 (2002).
36. Benjamini, Y. & Hochberg, Y. *Controlling False Discov. Rate - A Practical Powerful Approach Mult. Test.* **57**, 289–300 (1995).
37. Kim, K. I. & van de Wiel, M. A. Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinforma.* **9**, 114 (2008).
38. Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
39. Dusonchet, J. et al. A Parkinson's disease gene regulatory network identifies the signaling protein RGS2 as a modulator of LRRK2 activity and neuronal toxicity. *Hum. Mol. Genet* **23**, 4887–4905 (2014).
40. Campbell, D. B. et al. Association of RGS2 and RGS5 variants with schizophrenia symptom severity. *Schizophr. Res.* **101**, 67–75 (2008).
41. Lifschytz, T. et al. Relationship between Rgs2 gene expression level and anxiety and depression-like behaviour in a mutant mouse model: serotonergic involvement. *Int J. Neuropsychopharmacol.* **15**, 1307–1318 (2012).
42. Ising, M. et al. A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch. Gen. psychiatry* **66**, 966–975 (2009).
43. Su, L. et al. Association of EPHB1 rs11918092 and EFNB2 rs9520087 with psychopathological symptoms of schizophrenia in Chinese Zhuang and Han populations. *Asia Pac. Psychiatry* **8**, 306–308 (2016).
44. Kim, J. M. et al. SNPs in axon guidance pathway genes and susceptibility for Parkinson's disease in the Korean population. *J. Hum. Genet* **56**, 125–129 (2011).
45. Chandley, M. J. et al. Elevated gene expression of glutamate receptors in noradrenergic neurons from the locus coeruleus in major depression. *Int J. Neuropsychopharmacol.* **17**, 1569–1578 (2014).
46. Paul, I. A. & Skolnick, P. Glutamate and depression: clinical and preclinical studies. *Ann. N. Y Acad. Sci.* **1003**, 250–272 (2003).
47. Devon, R. S. et al. The genomic organisation of the metabotropic glutamate receptor subtype 5 gene, and its association with schizophrenia. *Mol. Psychiatry* **6**, 311–314 (2001).
48. Matosin, N. et al. Alterations of mGluR5 and its endogenous regulators Norbin, Tamalin and Preso1 in schizophrenia: towards a model of mGluR5 dysregulation. *Acta Neuropathol.* **130**, 119–129 (2015).
49. Black, Y. D. et al. Protective effect of metabotropic glutamate mGluR5 receptor elimination in a 6-hydroxydopamine model of Parkinson's disease. *Neurosci. Lett.* **486**, 161–165 (2010).
50. Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L. & Hommer, D. Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* **12**, 3683–3687 (2001).
51. McClure, S. M., York, M. K. & Montague, P. R. The neural substrates of reward processing in humans: the modern role of FMRI. *Neuroscientist* **10**, 260–268 (2004).
52. Hodgkinson, C. A. et al. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am. J. Hum. Genet.* **75**, 862–872 (2004).
53. Ji, B. et al. Inhibition of protein translation by the DISC1-Boymaw fusion gene from a Scottish family with major psychiatric disorders. *Hum. Mol. Genet.* **23**, 5683–5705 (2014).
54. Petrovska, J. et al. The NCAM1 gene set is linked to depressive symptoms and their brain structural correlates in healthy individuals. *J. Psychiatr. Res.* **91**, 116–123 (2017).
55. Irwin, W. et al. Human amygdala activation detected with echo-planar functional magnetic resonance imaging. *Neuroreport* **7**, 1765–1769 (1996).
56. Gluskin, B. S. & Mickey, B. J. Genetic variation and dopamine D2 receptor availability: a systematic review and meta-analysis of human in vivo molecular imaging studies. *Transl. psychiatry* **6**, e747–e747 (2016).
57. Glantz, L. A. et al. Pro-apoptotic Par-4 and dopamine D2 receptor in temporal cortex in schizophrenia, bipolar disorder and major depression. *Schizophr. Res.* **118**, 292–299 (2010).
58. Wood-Kaczmar, A., Deas, E., Wood, N. W. & Abramov, A. Y. The role of the mitochondrial NCX in the mechanism of neurodegeneration in Parkinson's disease. *Adv. Exp. Med. Biol.* **961**, 241–249 (2013).
59. Abudureyimu, S. et al. Essential Role of Linx/Islr2 in the Development of the Forebrain Anterior Commissure. *Sci. Rep.* **8**, 7292 (2018).
60. Paramo, B., Wyatt, S. & Davies, A. M. An essential role for neuregulin-4 in the growth and elaboration of developing neocortical pyramidal dendrites. *Exp. Neurol.* **302**, 85–92 (2018).
61. Garey, L. When cortical development goes wrong: schizophrenia as a neurodevelopmental disease of microcircuits. *J. Anat.* **217**, 324–333 (2010).

62. Torres-Berrío, A. et al. DCC Confers Susceptibility to Depression-like Behaviors in Humans and Mice and Is Regulated by miR-218. *Biol. Psychiatry* **81**, 306–315 (2017).
63. Wang, Z. et al. Axon guidance pathway genes are associated with schizophrenia risk. *Exp. Ther. Med.* **16**, 4519–4526 (2018).
64. Ward, J. 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. *Transl. Psychiatry* **7**, 1264 (2017).
65. Strawbridge, R. J. et al. Identification of novel genome-wide associations for suicidality in UK Biobank, genetic correlation with psychiatric disorders and polygenic association with completed suicide. *EBioMedicine* **41**, 517–525 (2019).
66. Johnston K. J. A. et al. Genome-wide Association Study of Multisite Chronic Pain in UK Biobank. *bioRxiv* 2018: 502807.
67. Auerbach, R. P. et al. Neuroanatomical Prediction of Anhedonia in Adolescents. *Neuropsychopharmacol. : Off. Publ. Am. Coll. Neuropsychopharmacol.* **42**, 2087–2095 (2017).
68. Harvey, P. O., Pruessner, J., Czechowska, Y. & Lepage, M. Individual differences in trait anhedonia: a structural and functional magnetic resonance imaging study in non-clinical subjects. *Mol. Psychiatry* **12**, 767–775 (2007). 703.
69. Sabatinelli, D., Bradley, M. M., Lang, P. J., Costa, V. D. & Versace, F. Pleasure rather than salience activates human nucleus accumbens and medial prefrontal cortex. *J. Neurophysiol.* **98**, 1374–1379 (2007).
70. Lebreton, M. et al. The brain structural disposition to social interaction. *Eur. J. Neurosci.* **29**, 2247–2252 (2009).
71. Pornpattananangkul, N., Leibenluft, E., Pine, D. S. & Stringaris, A. Association Between Childhood Anhedonia and Alterations in Large-Scale Resting-State Networks and Task-Evoked Activation. *JAMA Psychiatry* **76**, 624–633 (2019).
72. Kringelbach, M. L. The human orbitofrontal cortex: linking reward to hedonic experience. *Nat. Rev. Neurosci.* **6**, 691–702 (2005).
73. Gorwood, P. Neurobiological mechanisms of anhedonia. *Dialogues Clin. Neurosci.* **10**, 291–299 (2008).
74. Murphy, F. C., Nimmo-Smith, I. & Lawrence, A. D. Functional neuroanatomy of emotions: A meta-analysis. *Cogn., Affect., Behav. Neurosci.* **3**, 207–233 (2003).
75. Wise, T. et al. Common and distinct patterns of grey-matter volume alteration in major depression and bipolar disorder: evidence from voxel-based meta-analysis. *Mol. Psychiatry* **22**, 1455–1463 (2017).
76. Yang, X. H. et al. White matter microstructural abnormalities and their association with anticipatory anhedonia in depression. *Psychiatry Res. Neuroimaging* **264**, 29–34 (2017).
77. Horwath, E., Johnson, J., Klerman, G. L. & Weissman, M. M. What are the public health implications of subclinical depressive symptoms? *Psychiatr. Q* **65**, 323–337 (1994).
78. Auerbach, R. P., Pagliaccio, D. & Pizzagalli, D. A. Toward an Improved Understanding of Anhedonia. *JAMA Psychiatry* **76**, 571–573 (2019).
79. Fry, A. et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
80. Husain, M. & Roiser, J. P. Neuroscience of apathy and anhedonia: a transdiagnostic approach. *Nat. Rev. Neurosci.* **19**, 470–484 (2018).
81. Lyall, L. M. et al. Seasonality of depressive symptoms in women but not in men: A cross-sectional study in the UK Biobank cohort. *J. Affect Disord.* **229**, 296–305 (2018).
82. Grool, A. M. et al. Location and progression of cerebral small-vessel disease and atrophy, and depressive symptom profiles: the Second Manifestations of ARTerial disease (SMART)-Medea study. *Psychol. Med.* **42**, 359–370 (2012).
83. Gooding, D. C. & Pflum, M. J. The assessment of interpersonal pleasure: introduction of the Anticipatory and Consummatory Interpersonal Pleasure Scale (ACIPS) and preliminary findings. *Psychiatry Res.* **215**, 237–243 (2014).

4.5.1 Appendix E

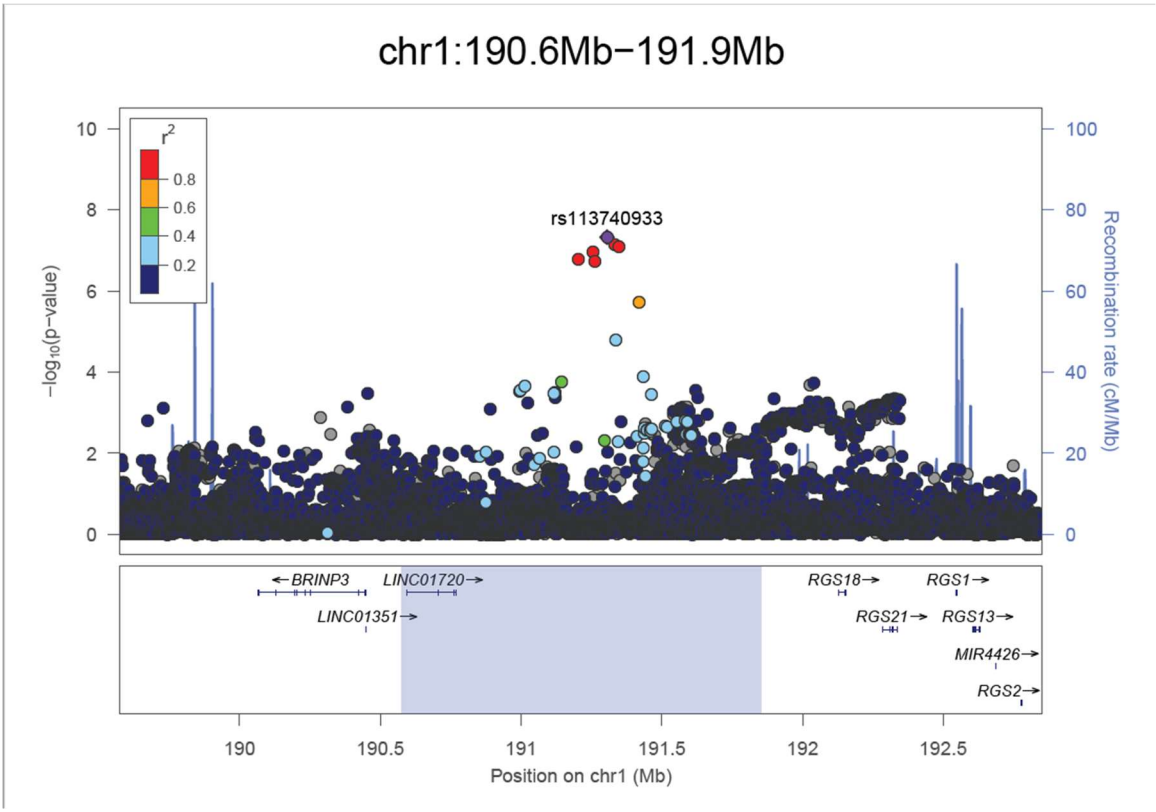
4.5.1.1 Table S1 List of excluded (self-reported) conditions

List of excluded (self-reported) conditions
Benign neuroma
Brain abscess/intracranial abscess
Brain cancer/primary malignant brain tumour
Brain haemorrhage
Cerebral aneurysm
Cerebral palsy
Chronic/degenerative neurological problem
dementia/Alzheimer's/cognitive impairment
Encephalitis
Epilepsy
Fracture skull/head
Head injury
Ischaemic stroke
Meningeal cancer/malignant meningioma
Meningioma/benign meningeal tumour
Meningitis
Motor Neurone Disease
Multiple Sclerosis
Nervous system infection
Neurological injury/trauma
Other demyelinating disease (not Multiple Sclerosis)
Other neurological problem
Parkinson's Disease
Spina Bifida
Stroke
Subarachnoid haemorrhage
Subdural haemorrhage/haematoma
Transient ischaemic attack

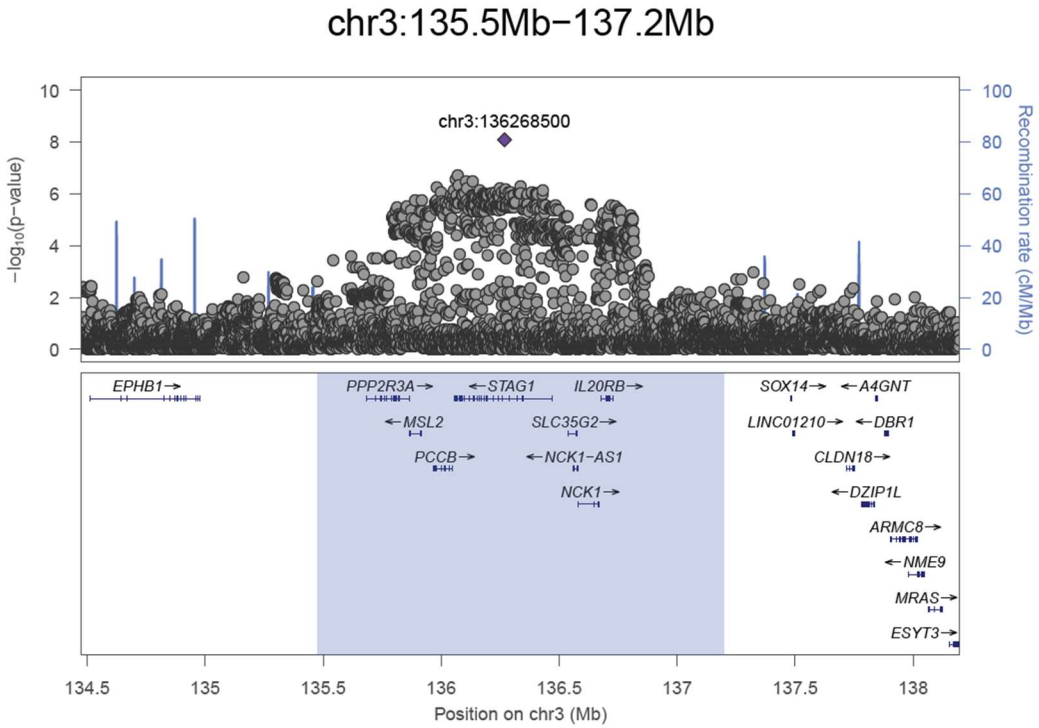
4.5.1.2 Table S2 – Top Hits – Discovery

SNP	CHR	BP	A1	A2	BETA	S.E.	p	start_BP	stop_BP	info	Candidate Genes
rs113740933	1	191305574	A	G	-0.0178355	0.0032652	4.70E-08	190575572	191852635	0.989	*RGS1, RGS2*
3:136268500_GT_G	3	136268500	GT	G	0.0084754	0.0014691	8.00E-09	135476532	137198063	0.973	*EPHB1,*
6:27731402_ATTTTTTTTTTTTTTTTTTTTTTTTTT_A	6	27731402	ATTTTTTTTTTTTTTTTTTTTTTTTTTTT	A	-0.0100565	0.0015786	1.90E-10	25323143	32064726	0.971	
rs756739942	10	68530607	AAATT	A	0.0094145	0.0016926	2.70E-08	68380593	68626126	0.998	*CTNNA3*
11:88755654_CA_C	11	88755654	CA	C	0.0083944	0.0014412	5.70E-09	88175768	89206451	0.904	*GRM5, DISC1FP1*
11:112900785_TA_T	11	112900785	TA	T	-0.0083567	0.0013992	2.30E-09	112758467	113062983	0.996	*NCAM1, DRD2*
rs35416728	12	127312142	G	A	0.0081883	0.001432	1.10E-08	127244817	127359960	0.989	
14:30164440_CA_C	14	30164440	CA	C	-0.0087454	0.0014705	2.70E-09	29998952	30300361	0.912	*PRKD1*
rs1965449	14	69680441	T	C	-0.0080697	0.0014282	1.60E-08	69365255	69791308	0.997	*SLC8A3*
15:75180130_CT_C	15	75180130	CT	C	0.0112911	0.0020003	1.70E-08	74607257	75474936	0.981	*ISLR2, NRG4*
rs72923287	18	50759318	C	T	-0.0119671	0.0013677	2.10E-18	50201613	51061399	0.994	*DCC*

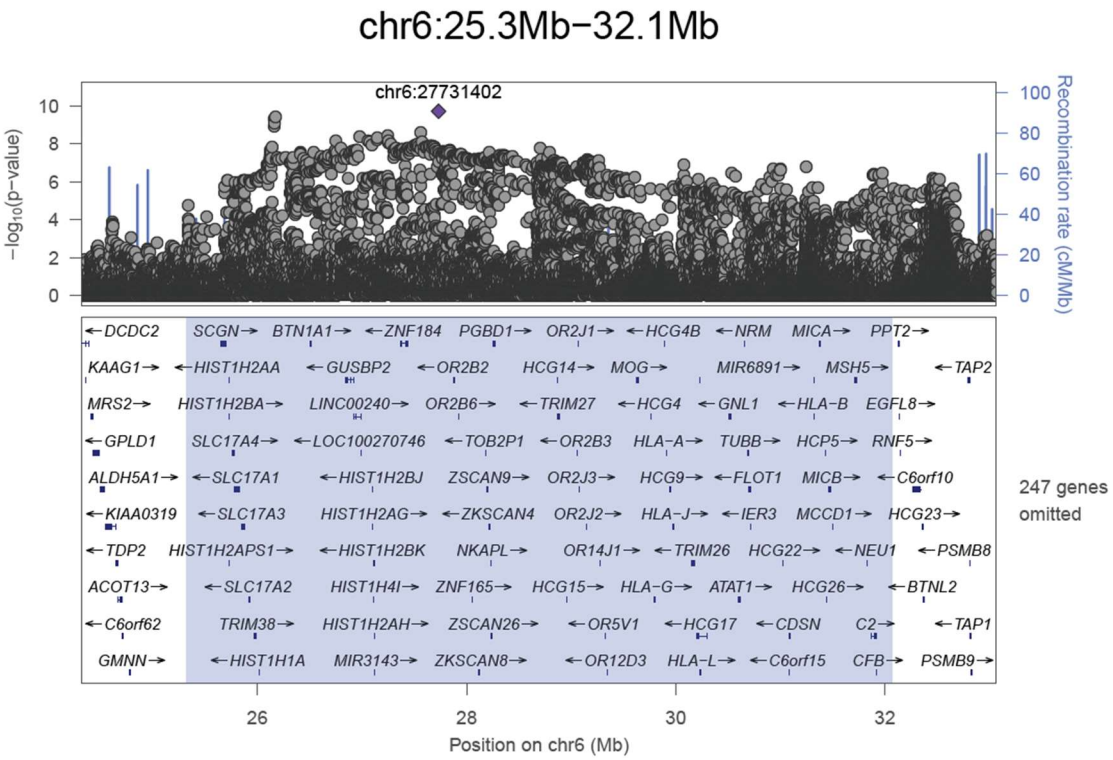
4.5.1.3 Figure S1 LocusZoom plot of anhedonia associated loci on chromosome 1



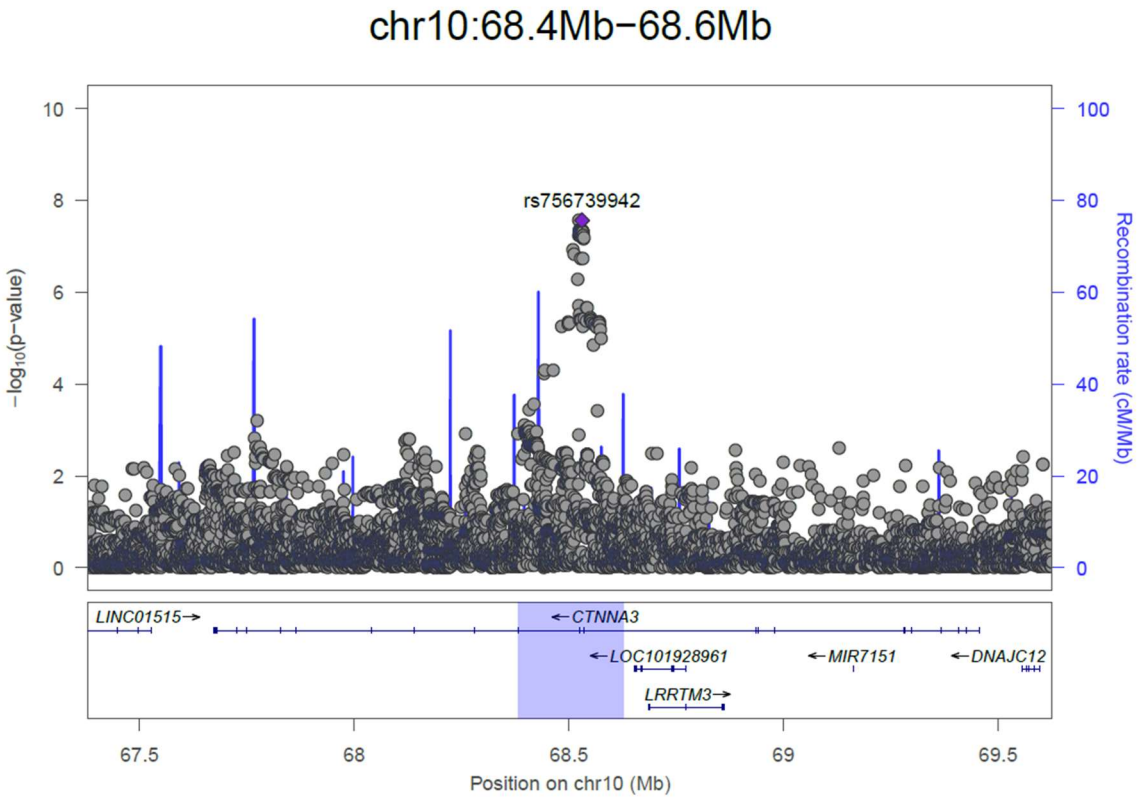
4.5.1.4 Figure S2 LocusZoom plot of anhedonia associated loci on chromosome 3



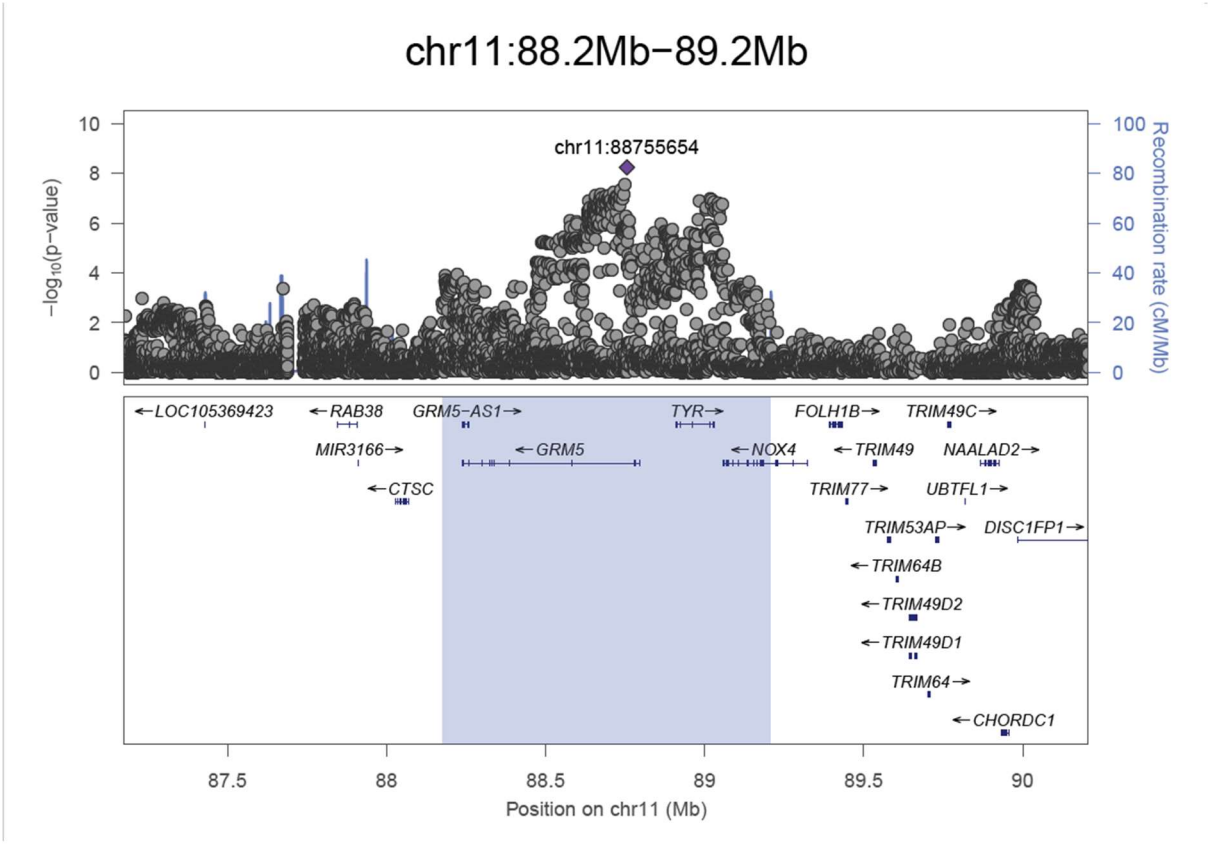
4.5.1.5 Figure S3 LocusZoom plot of anhedonia associated loci on chromosome 6



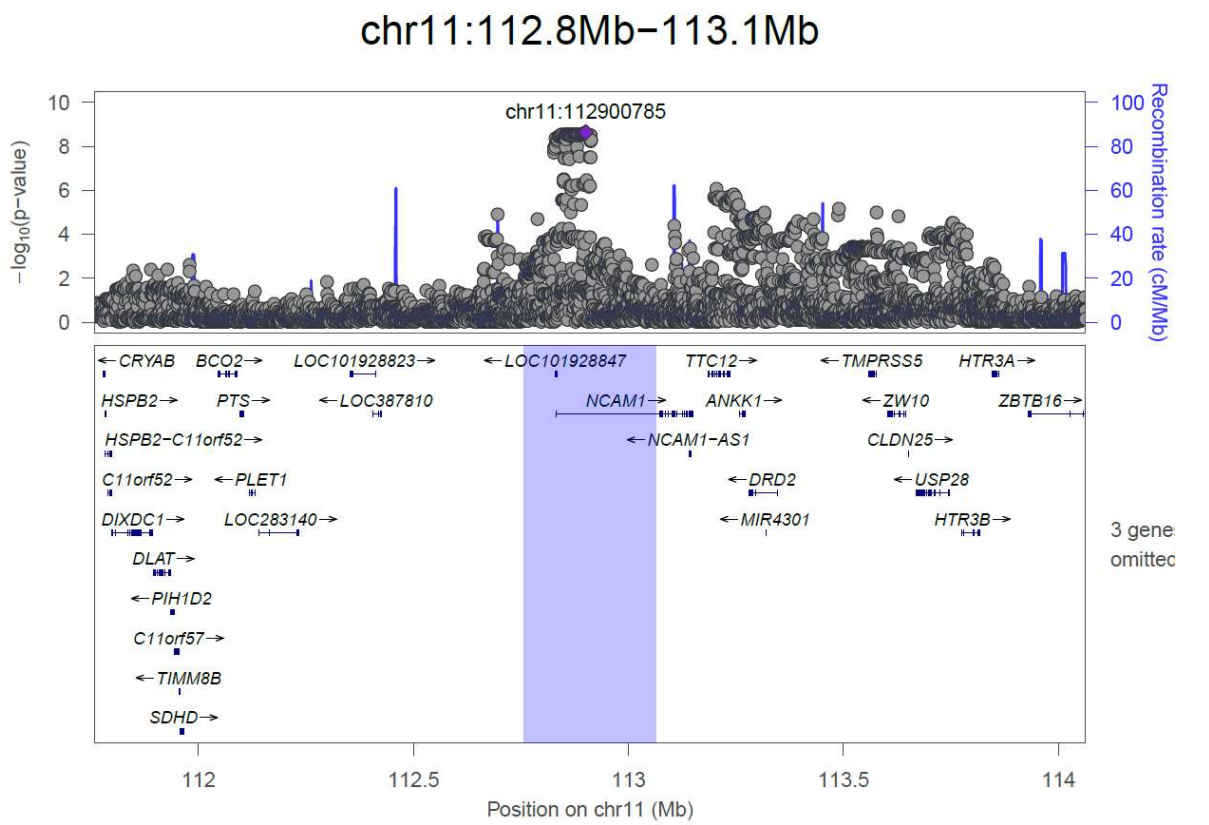
4.5.1.6 Figure S4 LocusZoom plot of anhedonia associated loci on chromosome 10



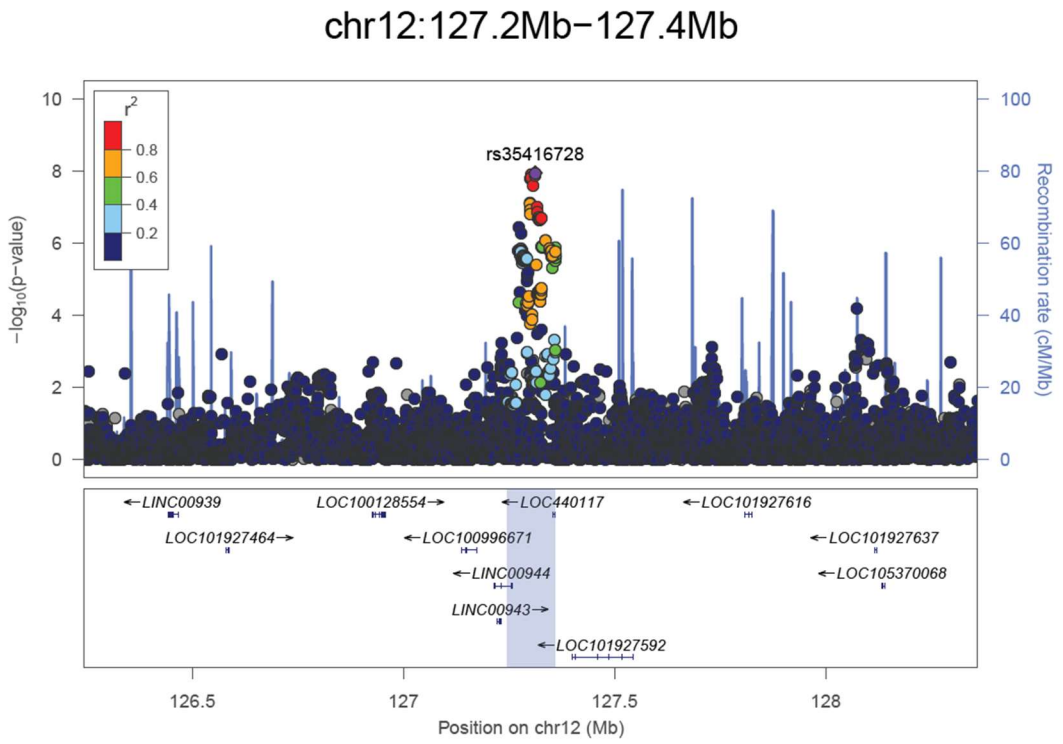
4.5.1.7 Figure S5 LocusZoom plot of anhedonia associated loci on chromosome 11



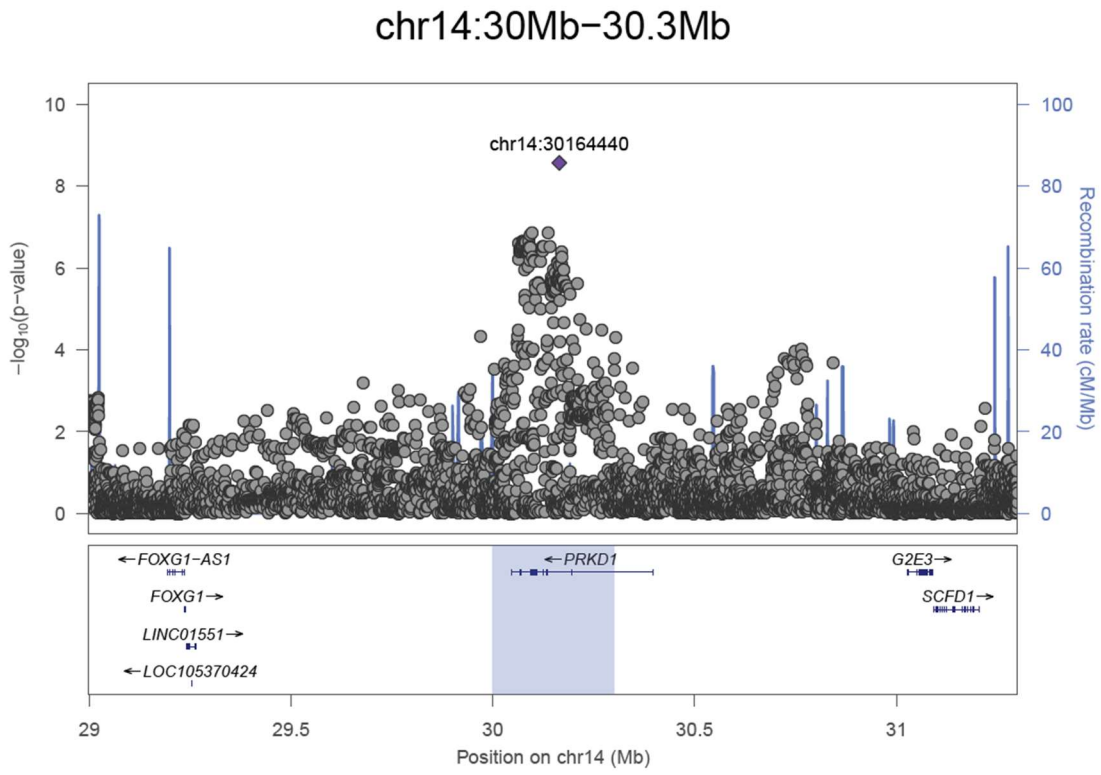
4.5.1.8 Figure s6 LocusZoom plot of anhedonia associated loci on chromosome 11



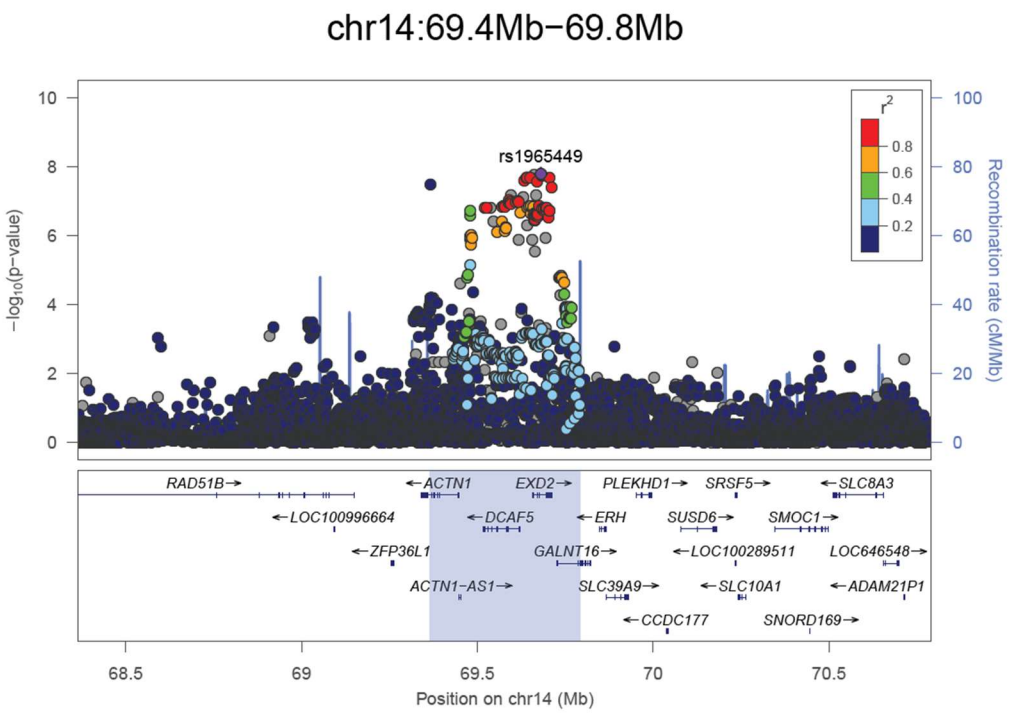
4.5.1.9 Figure S7 LocusZoom plot of anhedonia associated loci on chromosome 12



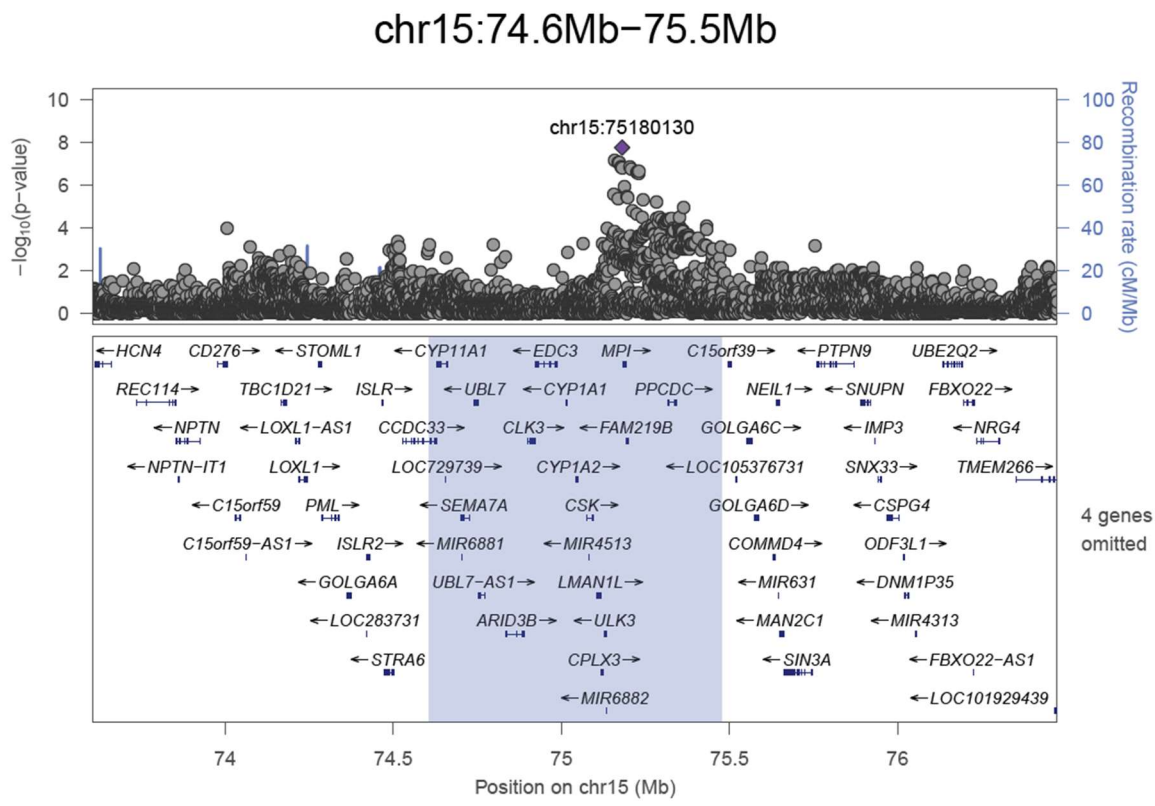
4.5.1.10 Figure S8 LocusZoom plot of anhedonia associated loci on chromosome 14



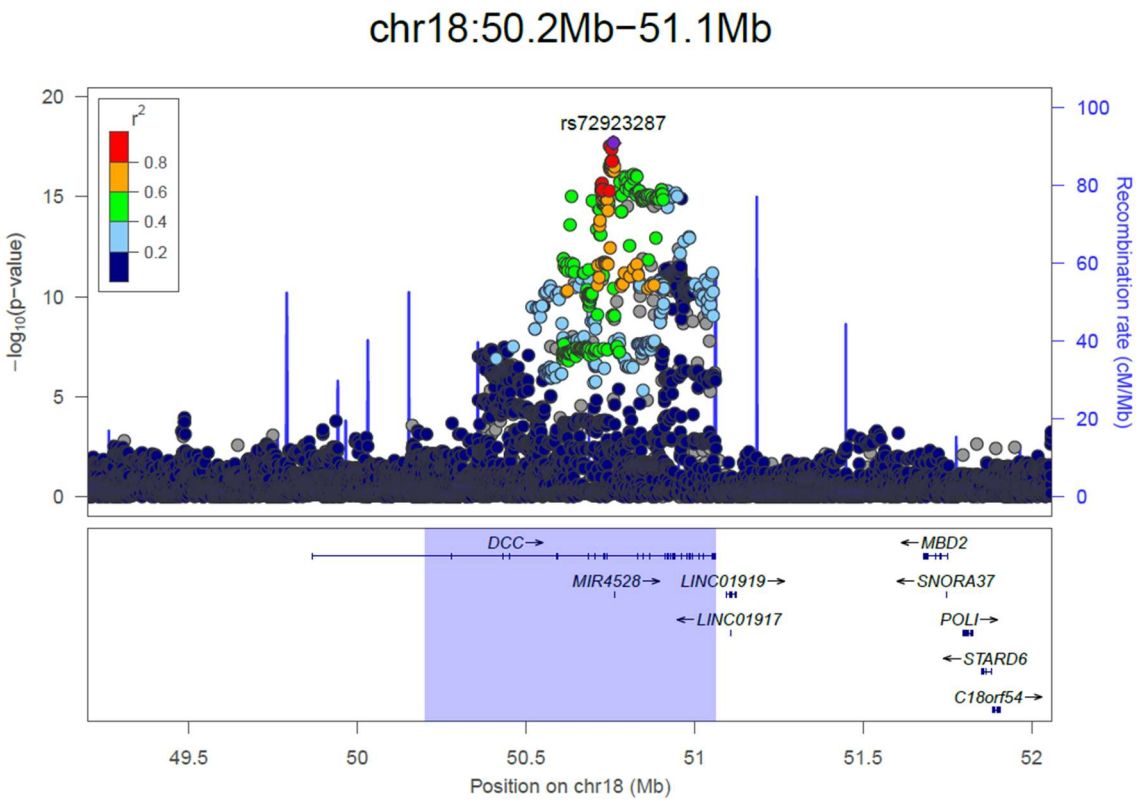
4.5.1.11 Figure S9 LocusZoom plot of anhedonia associated loci on chromosome 14



4.5.1.12 Figure S10 LocusZoom plot of anhedonia associated loci on chromosome 15



4.5.1.13 Figure S11 LocusZoom plot of anhedonia associated loci on chromosome 18



4.5.1.14 Table S3 – Top Hits - Replication

SNP	CHR	BP	A1	A2	BETA	S.E.	p
rs113740933	1	191305574	A	G	-1.86E-02	0.0160925	0.25
3:136268500_GT_G	3	136268500	GT	G	1.09E-02	0.0072202	0.13
6:27731402_ATTTTTTTTTTTTTTTTTTTTTTTTTTTT_A	6	27731402	ATTTTTTTTTTTTTTTTTTTTTTTTTTTT	A	-1.89E-02	0.0077679	0.015
rs756739942	10	68530607	AAATT	A	-7.69E-05	0.0084288	0.99
11:88755654_CA_C	11	88755654	CA	C	-8.04E-04	0.0070401	0.91
11:112900785_TA_T	11	112900785	TA	T	-1.90E-03	0.0068869	0.78
rs35416728	12	127312142	G	A	4.97E-03	0.0070003	0.48
14:30164440_CA_C	14	30164440	CA	C	-1.14E-02	0.0071944	0.11
rs1965449	14	69680441	T	C	2.85E-03	0.0070102	0.68
15:75180130_CT_C	15	75180130	CT	C	8.53E-03	0.0097378	0.38
rs72923287	18	50759318	C	T	-7.35E-03	0.0067357	0.28

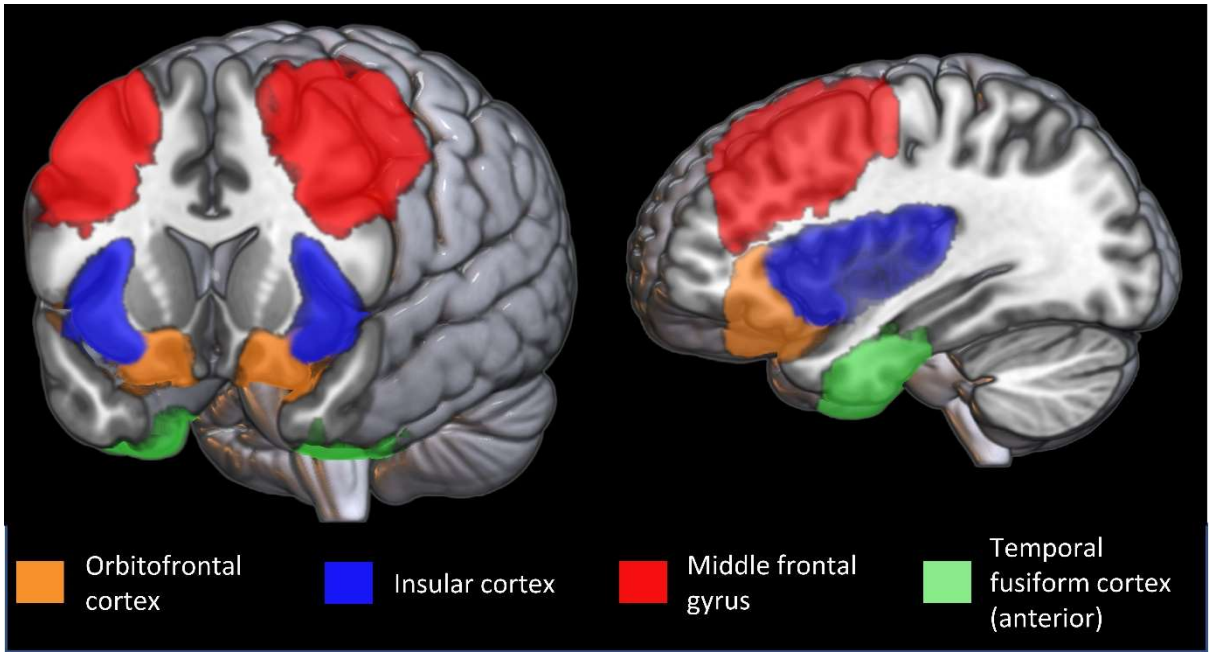
4.5.1.15 Table S4 – anhedonia PRS associations with total brain volumes

Outcome	N	Coefficient (95% CI)	p	q	R^2
Grey matter volume	17071	-0.025 (-0.036, -0.013)	<0.001	**<0.001**	0.45
White matter volume	17065	0.011 (-0.004, 0.026)	0.164	0.246	0.11
Ventricular CSF volume	16885	-0.001 (-0.014, 0.013)	0.920	0.920	0.26

4.5.1.16 Table S5 – anhedonia PRS associations with ROIs

Outcome	N	Coefficient (95% CI)	p	q	Pseudo R^2
Nucleus accumbens	17105	-0.014 (-0.026, -0.002)	0.027	0.068	0.23
Insular cortex	17080	-0.024 (-0.035, -0.013)	<0.001	**<0.001**	0.51
Medial frontal cortex	17097	-0.015 (-0.028, -0.002)	0.027	0.068	0.19
Orbitofrontal cortex	17090	-0.015 (-0.026, -0.005)	0.004	**0.021**	0.44
Middle frontal gyrus	17104	-0.015 (-0.026, -0.004)	0.007	**0.027**	0.35
Amygdala	17103	-0.001 (-0.012, 0.012)	0.970	0.970	0.12
Supramarginal gyrus (posterior)	17112	-0.007 (-0.018, 0.004)	0.196	0.267	0.19
Hippocampus	17080	-0.007 (-0.019, 0.006)	0.294	0.367	0.23
Parahippocampal gyrus	17099	-0.009 (-0.020, 0.003)	0.155	0.233	0.32
Posterior cingulate gyrus	17022	-0.009 (-0.019, 0.002)	0.097	0.183	0.54
Caudate	17075	-0.013 (-0.026, -0.001)	0.041	0.088	0.34
Putamen	17079	-0.006 (-0.017, 0.005)	0.319	0.368	0.45
Anterior cingulate gyrus	16977	-0.002 (-0.015, 0.011)	0.751	0.805	0.24
Superior temporal gyrus (anterior)	17107	-0.009 (-0.020, 0.003)	0.132	0.220	0.21
Temporal fusiform cortex (anterior)	17107	-0.022 (-0.034, -0.011)	<0.001	**0.001**	0.26

4.5.1.17 Figure S12 brain regions associated with anhedonia PRS



4.5.1.18 Table S6 - anhedonia PRS associations with General Factors

Outcome	N	Coefficient (95% CI)	p	q	R^2
*g*FA	14481	-0.038 (-0.095, 0.020)	0.197	0.197	0.08
*g*MD	14191	0.055 (-0.003, 0.133)	0.063	0.126	0.16

4.5.1.19 Table S7 – anhedonia PRS associations with FA

Outcome	N	Coefficient (95% CI)	p	q	Pseudo R^2
Acoustic radiation	15505	0.010 (-0.004, 0.025)	0.160	0.229	0.07
Anterior thalamic radiation	15488	-0.021 (-0.036, -0.005)	0.008	**0.032**	0.09
Cingulate gyrus part of cingulum	15508	-0.002 (-0.016, 0.013)	0.828	0.828	0.06
Corticospinal tract	15504	-0.011 (-0.026, 0.004)	0.161	0.229	0.07
Forceps major*	15444	-0.017 (-0.033, -0.001)	0.047	0.100	0.05
Forceps minor*	15467	-0.024 (-0.040, -0.008)	0.004	**0.027**	0.10
Inferior fronto-occipital fasciculus	15488	-0.016 (-0.032, -0.001)	0.044	0.100	0.07
Inferior longitudinal fasciculus	15486	-0.006 (-0.022, 0.009)	0.426	0.532	0.09
Medial lemniscus	15501	-0.020 (-0.034, -0.006)	0.006	**0.031**	0.04
Middle cerebellar peduncle*	15415	0.006 (-0.010, 0.022)	0.476	0.549	0.06
Parahippocampal part of cingulum	15456	0.002 (-0.012, 0.017)	0.736	0.789	0.07
Posterior thalamic radiation	15484	-0.023 (-0.039, -0.008)	0.003	**0.027**	0.09
Superior longitudinal fasciculus	15475	-0.014 (-0.030, 0.002)	0.078	0.146	0.08
Superior thalamic radiation	15478	-0.019 (-0.035, -0.003)	0.024	0.071	0.03
Uncinate fasciculus	15504	-0.010 (-0.025, 0.004)	0.168	0.229	0.08

4.5.1.20 Table S8 – anhedonia PRS associations with MD

Outcome	N	Coefficient (95% CI)	p	q	Pseudo R^2
Acoustic radiation	15504	0.014 (0.001, 0.028)	0.049	0.066	0.04
Anterior thalamic radiation	15437	0.019 (0.005, 0.033)	0.008	**0.016**	0.27
Cingulate gyrus part of cingulum	15484	0.021 (0.006, 0.036)	0.005	**0.016**	0.16
Corticospinal tract	15497	0.016 (0.001, 0.032)	0.040	0.060	0.05
Forceps major*	15416	0.005 (-0.011, 0.021)	0.543	0.626	0.10
Forceps minor*	15443	0.025 (0.009, 0.040)	0.002	**0.009**	0.16
Inferior fronto-occipital fasciculus	15472	0.021 (0.006, 0.036)	0.008	**0.016**	0.15
Inferior longitudinal fasciculus	15465	0.013 (-0.002, 0.029)	0.079	0.099	0.17
Medial lemniscus	15496	0.001 (-0.013, 0.015)	0.892	0.929	0.05
Middle cerebellar peduncle*	15423	0.026 (0.010, 0.043)	0.001	**0.009**	0.09
Parahippocampal part of cingulum	15448	-0.001 (-0.014, 0.013)	0.929	0.929	0.06
Posterior thalamic radiation	15452	0.028 (0.014, 0.042)	<0.001	**0.002**	0.19
Superior longitudinal fasciculus	15430	0.021 (0.005, 0.036)	0.010	**0.016**	0.14
Superior thalamic radiation	15422	0.019 (0.005, 0.033)	0.009	**0.016**	0.28
Uncinate fasciculus	15489	0.019 (0.005, 0.033)	0.007	**0.016**	0.22

4.5.1.21 Table S9 – BOLD

Outcome	N	Coefficient (95% CI)	p	q	R^2
Median BOLD – amygdala mask	14833	-0.001 (-0.002, 0.001)	0.484	0.484	0.02
Median BOLD – group-defined mask (including Temporo-occipital regions and amygdala)	14856	-0.004 (-0.008, 0.001)	0.129	0.259	0.04

5 Discussion

5.1 Technical considerations

This thesis by publication shows the development of techniques in genetic analysis over the past several years. The thesis began with investigating whether genetic loading for a clinical diagnosis and one of the 'big five' personality traits (Digman, 1990) (neuroticism) could be used to responders to SSRIs. Here the P&T PRS method was used for the estimation of genetic loading for the two phenotypes. This was the first type of risk scoring that takes account of the polygenic nature of traits and accounts for loci that don't meet the formal threshold for significance in the discovery GWAS. To overcome the fact that not all loci that impact on a phenotype will reach genome wide significance, several scores were created at ever less stringent p value thresholds. Although the creation of several scores allows for accounting of sub-genome wide significant loci, it does create some additional issues.

Firstly, not all loci that get incorporated into the scores will be true positive loci that do have a causal impact on the trait but failed to achieve statistical significance due to lack of statistical power or poor phenotyping etc. By lowering the p value threshold for which SNPs get incorporated into the score more false negatives but so will true negatives. These true negatives will add noise to the score and reduce its power. Secondly, there is a multiple testing correction issue. Say for example 5 scores are created and tested against an outcome. These tests are not truly independent, so how is a threshold for significance determined? It is possible to use a conservative Bonferroni correction but this could be considered too stringent, especially when dealing with the rather small associations that are found in PRS analyses. Thirdly, it creates a situation where one person can have several different genetic loadings for a single trait. In paper one the cohorts were split into quintiles and only the top and bottom quintile were compared, however those who appeared in the top and bottom quintile for each of the respective regressions changed depending on the p value cut off used for the score.

Papers two, three and four also use the P&T methodology but used the whole range of scores, not just those with the highest and lowest loading. Although this overcame the issue of

changing who in the cohort was used in each of the analyses, there was still the issue that where an individual happens to appear in the distributions of scores will change for each regression.

The implementation of LDpred in paper five helps overcome the latter two issues identified with P&T. Here the use of an infinitesimal model where all independent loci of the genome are incorporated into the score leads to the creation of one score per trait. Therefore, an individual is assigned one genetic loading and, as there is only one score there is no need to correct for multiple testing. As noted earlier, the use of LDpred also confers another large advantage over P&T due to the use of an expected beta per locus and not just selecting the most significant SNP from each locus leading to a large increase in the power of the score.

There is also an evolution of the GWAS methodology used throughout the thesis. Paper two uses samples of unrelated individuals in a logistic regression for a case/control analysis and paper four uses a cumulative link function which is an extension of a logistic regression. The use of a logistic regression is favoured over the simpler allelic association as it allows for the model to be adjusted for covariates, in this case: age, sex, genotyping chip and the first eight principal components. While this methodology was fruitful with the identification of loci in each analysis, there are some methodological issues that arise.

Firstly, how many principal components should be included in the model? Most of the variance is explained by the first few principal components but not adjusting for a sufficient number of principal components will lead to variance being attributed to SNPs rather than hidden population stratification. Secondly, these methodologies work by testing each SNP in turn, starting at the top of chromosome one and finishing at the bottom of chromosome twenty-two. However, this clearly isn't how biological systems work. All the SNPs are acting simultaneously. Testing SNPs one by one will give each SNP a higher proportion of variance than if all SNPs are considered together.

The use of BOLT-LMM in papers three and five helped to address these issues. BOLT-LMM makes use of a genetic relationship matrix, GRM, which allows the model to account for

relatedness between individuals in the sample and thereby eliminates the need to exclude those who are too closely related. The GRM also accounts for population stratification of the sample used, which eliminates the need for the inclusion of principal components as covariates.

BOLT-LMM also builds a global model of genotyped SNPs and then tests imputed SNPs against this model. This is better than testing SNPs individually because the genotypes for which there is direct evidence can be used alongside the GRM to model a system more reflective of a real biological system.

5.2 Biological context of findings

One of the primary focuses of this thesis was the use of non-clinical phenotypes and how they can be used in the discovery of genetic variants that have relevance to clinical outcomes. The first GWAS of mood instability was of relatively modest sample size and detected four independent loci. Additionally, the mood instability phenotype showed significant genetic correlation with MDD and nominal significant correlations between schizophrenia and anxiety disorder. Surprisingly, the correlation with BD was null ($p = 0.27$). The lack of a significant correlation was attributed to a lack of power either in our study and/or theirs. It was for this reason the analysis was repeated on the full cohort once the genetic data had been made available.

By trebling the sample size and using more sophisticated methodologies, over ten-fold more loci were identified. This increase in statistical power not only also led to significant genetic correlations with all six psychiatric outcomes tested but allowed for further downstream analysis of eQTLs and PLN. The PLN analysis identified a community of genes that included melatonin receptors, that are involved in circadian rhythmicity and whose dysregulation has been shown in clinical mood disorders (Srinivasan et al., 2006, Pacchierotti et al., 2001).

The exact mechanisms by which melatonin may be affecting patients with MDD is unclear as studies show that both reduced and elevated melatonin levels can be observed. Subjects may show decreased levels of melatonin (Brown et al., 1985, Claustrat et al., 1984) which is thought

to be driven by lack of the precursor substrate serotonin and tryptophan(Arendt, 1989), or increased levels of melatonin (Rubin et al., 1992) though to correspond with the lethargic symptoms. Other studies point to a phase shift of the diurnal cycle being the driving factor (Crasson et al., 2004).

It is not just MDD that shows disruption in the melatonin regulation but also in bipolar disorder (Kennedy et al., 1996), schizophrenia (Monteleone et al., 1997) and even suicidal behaviour (Stanley and Brown, 1988). Higher levels of melatonin during the manic phase compared to the euthymic and depressive phases have been observed in bipolar subjects (Lewy et al., 1979) whilst being lower overall compared to controls (Kennedy et al., 1996). In schizophrenia, treatment with melatonin can leads to exacerbation of psychotic symptoms (Altschule, 1957) however more recent studies have failed to establish a clear link between the two (Morera-Fumero and Abreu-Gonzalez, 2013). The community of genes that contained melatonin receptors identified in the PLN may well partly explain the significant genetic correlations of mood instability with MDD, bipolar disorder and schizophrenia.

The discussion section of papers two through five speculates on how candidate genes found in significant loci in the GWAS had previously identified roles brain development and function as well as their implication in various mood disorders. It is not possible to state the exact mechanisms by which these genes may affect the phenotypes due to our current lack understanding of brain development and function. Also, as there are often multiple genes found in the associated loci it is not possible to tell whether the one or several genes are impacting the phenotype. However, there is one locus identified in both the mood instability GWAS and also the suicidality GWAS which contained only a single gene, netrin receptor 1 (DCC).

DCC is a transmembrane protein that localises to the cell membrane and detects presence of netrin 1 ligands in some neurons causing axon guidance towards the ligand whilst conversely causing apoptosis in its absence but also having an opposite effect in other neurons in the

presence of UNC5 (Moore et al., 2007). RNAseqDB (Wang et al., 2018) shows transcripts of DCC are expressed in a wide variety of tissues throughout the body but proteomicsDb (Schmidt et al., 2017) shows only minor presence in most tissues except in foetal brain as well as frontal cortex, a region associated with personality (DeYoung et al., 2010), or more specifically brain functions such as executive function, theory of mind, social cognition (Chow, 2000). It is therefore not surprising that this gene should associate with a wide variety of cognitive phenotypes and as such warrant's further investigation as a potential drug target in the treatment of psychiatric ailments.

6 Limitations

The discussion sections of each of papers in this thesis have sought to clarify and contextualise genetic loci previously identified in the field of psychiatric genetics at the clinical level using these increasingly sophisticated techniques, but; there remain limitations to the work. The most notable is that of using self-reported and unvalidated measures of the traits in question. For example, there are derived scales for mood instability (Marwaha et al., 2014) that would capture the mood instability phenotype more completely than the binary question used in papers two and three. The same is true for anhedonia (Chapman et al., 1976); validated objective instruments have been developed that capture a fuller description of the complex trait of anhedonia. The reason for these scales not being used is that the analyses were limited by the variables that UK Biobank had collected. It may have been possible to ask Biobank to send out questionnaires with these validated instruments but that would have been prohibitively expensive and time-consuming.

Another major limitation is that the UK Biobank, despite being a general population cohort, is not fully reflective of the wider population from which it was drawn. It has a higher socio-economic profile from that of the UK as a whole and is generally healthier as a result. This could lead to an underestimation of effect sizes and, as such, more false negatives in analyses.

It is also worth noting that the loci identified by the GWAS in this thesis were not replicated in an independent cohort. Several other measures were taken to validate the output of the GWAS. All the GWAS papers used LDSR to establish whether the genetics of the traits under investigation showed similarity to that of clinical diagnoses of which the trait analysed was a component. Paper two used PRS to determine how genetic loading of these traits correlated with individuals with a diagnosis of a psychiatric condition.

Two slightly different attempts at the more traditional method replication were made. The first in paper three used a subsample of the biobank cohort to validate the four significant loci identified in paper two. Although the top SNPs from each loci did achieve significance after Bonferroni correction, as noted in the paper, the sample used may be subject to the same underlying biases as the original sample.

The second attempt in paper five was to use those who were excluded from the GWAS as they had valid MRI data but also had a valid anhedonia phenotype. Unfortunately, as is so often the case in genomic analysis, only 1 SNP achieved nominal significance, but it would not have survived correction for multiple testing. I attribute the lack of significant replication of the loci identified in the main analysis is due to lack of power as the replication sample was only 15,000. It will only be once new cohorts of sufficient size (such as All Of Us, A UK Biobank-style cohort in the USA of 1 million US citizens) will true independent replication of these loci be able to be performed.

7 Future work

Paper one touches on stratified medicine in psychiatry. It is highly unlikely that a single risk score would have any significant degree of predictive power to be clinically useful, given that PRS usually explain a few percent of the variance of a phenotype. The use of multiple PRS in a single model will require more sophisticated modelling techniques than a standard linear or logistic modelling can provide. A technique showing promise would be that of GLM-NET. This

technique can select for highly correlated variables in the model and allows for the constraining of coefficients within certain predefined limits.

The outputs derived from papers two to four as well as others will be used in such models. Larger cohorts such as All Of Us will be available in the coming years, allowing for prediction modelling on a scale where real, clinically useful models can be generated. This however is not the final stage in the implementation of genetics in patient diagnosis. The final obstacle in the implementation of the use of risk scores in clinical diagnoses will the implementation of genotyping of patients by medical practitioners and health care systems. It is not currently standard practice to genotype patients. Several hurdles such as sample collection, genetic imputation protocols and generalisability of imputation across ancestry groups as well as regulations on the storage and access of data will need to be overcome.

The work presented in this thesis makes no real claim on the biological mechanisms that link the loci to the traits. The estimated heritability of these phenotypes is only around 10% of the phenotype, so even larger sample sizes and ever more sophisticated techniques will be required to further uncover the lower penetrance SNPs, genes and pathways that generate the phenotype. The methodologies used throughout the thesis have advanced our understanding of the genetic components of the traits under investigation better than the basic clinical case/control analyses performed by large consortia in the past. Stratified medicine in psychiatry is still in its infancy. Real progress in determining the causes and remedies for psychiatric morbidities will require ever more specific and detailed scales, measured with greater accuracy, and the development of new techniques that closer reflect underlying biological systems. Only then will it be possible not just to get the right treatment to the right patient at the right time leading to faster remission times and higher remission rates but potentially lead to the prevention of these morbidities arising in the first place.

8 References

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

ALTSCHULE, M. D. 1957. Some effects of aqueous extracts of acetone-dried beef-pineal substance in chronic schizophrenia. *N Engl J Med,* 257**,** 919-22.

AMARE, A. T., SCHUBERT, K. O., HOU, L., CLARK, S. R., PAPIOL, S., HEILBRONNER, U., DEGENHARDT, F., TEKOLA-AYELE, F., HSU, Y. H., SHEKHTMAN, T., ADLI, M., AKULA, N., AKIYAMA, K., ARDAU, R., ARIAS, B., AUBRY, J. M., BACKLUND, L., BHATTACHARJEE, A. K., BELLIVIER, F., BENABARRE, A., BENGESSER, S., BIERNACKA, J. M., BIRNER, A., BRICHANT-PETITJEAN, C., CERVANTES, P., CHEN, H. C., CHILLOTTI, C., CICHON, S., CRUCEANU, C., CZERSKI, P. M., DALKNER, N., DAYER, A., DEL ZOMPO, M., DEPAULO, J. R., ETAIN, B., FALKAI, P., FORSTNER, A. J., FRISEN, L., FRYE, M. A., FULLERTON, J. M., GARD, S., GARNHAM, J. S., GOES, F. S., GRIGOROIU-SERBANESCU, M., GROF, P., HASHIMOTO, R., HAUSER, J., HERMS, S., HOFFMANN, P., HOFMANN, A., JAMAIN, S., JIMENEZ, E., KAHN, J. P., KASSEM, L., KUO, P. H., KATO, T., KELSOE, J., KITTEL-SCHNEIDER, S., KLIWICKI, S., KONIG, B., KUSUMI, I., LAJE, G., LANDEN, M., LAVEBRATT, C., LEBOYER, M., LECKBAND, S. G., TORTORELLA, A., MANCHIA, M., MARTINSSON, L., MCCARTHY, M. J., MCELROY, S., COLOM, F., MITJANS, M., MONDIMORE, F. M., MONTELEONE, P., NIEVERGELT, C. M., NOTHEN, M. M., NOVAK, T., O'DONOVAN, C., OZAKI, N., OSBY, U., PFENNIG, A., POTASH, J. B., REIF, A., REININGHAUS, E., ROULEAU, G. A., RYBAKOWSKI, J. K., SCHALLING, M., SCHOFIELD, P. R., SCHWEIZER, B. W., SEVERINO, G., SHILLING, P. D., SHIMODA, K., SIMHANDL, C., SLANEY, C. M., SQUASSINA, A., STAMM, T., STOPKOVA, P., MAJ, M., TURECKI, G., et al. 2018a. Association of Polygenic Score for Schizophrenia and HLA Antigen and Inflammation Genes With Response to Lithium in Bipolar Affective Disorder: A Genome-Wide Association Study. *JAMA Psychiatry,* 75**,** 65-74.

AMARE, A. T., SCHUBERT, K. O. T., SCHULZE, T. G. & BAUNE, B. T. 2018b. Polygenic scores for major depressive disorder and depressive symptoms predict response to lithium in patients with bipolar disorder. *bioRxiv***,** 449363.

ARENDT, J. 1989. Melatonin: a new probe in psychiatric investigation? *Br J Psychiatry,* 155**,** 585-90.

BIERNACKA, J. M., SANGKUHL, K., JENKINS, G., WHALEY, R. M., BARMAN, P., BATZLER, A., ALTMAN, R. B., AROLT, V., BROCKMOLLER, J., CHEN, C. H., DOMSCHKE, K., HALL-FLAVIN, D. K., HONG, C. J., ILLI, A., JI, Y., KAMPMAN, O., KINOSHITA, T., LEINONEN, E., LIOU, Y. J., MUSHIRODA, T., NONEN, S., SKIME, M. K., WANG, L., BAUNE, B. T., KATO, M., LIU, Y. L., PRAPHANPHOJ, V., STINGL, J. C., TSAI, S. J., KUBO, M., KLEIN, T. E. & WEINSHILBOUM, R. 2015. The International SSRI Pharmacogenomics Consortium (ISPC): a genome-wide association study of antidepressant treatment response. *Transl Psychiatry,* 5**,** e553.

BIOBANK, U. 2018. UK Biobank. Genotyping of 500,000 UK Biobank participants. Description of sample processing workflow and preparation of DNA for genotyping, 11 September 2015. Available at: .

BROWN, R., KOCSIS, J. H., CAROFF, S., AMSTERDAM, J., WINOKUR, A., STOKES, P. E. & FRAZER, A. 1985. Differences in nocturnal melatonin secretion between melancholic depressed patients and control subjects. *Am J Psychiatry,* 142**,** 811-6.

CHAPMAN, L. J., CHAPMAN, J. P. & RAULIN, M. L. 1976. Scales for physical and social anhedonia. *Journal of Abnormal Psychology,* 85**,** 374-382.

CHENEVERT, T. L., BRUNBERG, J. A. & PIPE, J. G. 1990. Anisotropic diffusion in human white matter: demonstration with MR techniques in vivo. *Radiology,* 177**,** 401-405.

CHOW, T. W. 2000. Personality in frontal lobe disorders. *Current psychiatry reports,* 2**,** 446-451.

CLAUSTRAT, B., CHAZOT, G., BRUN, J., JORDAN, D. & SASSOLAS, G. 1984. A chronobiological study of melatonin and cortisol secretion in depressed subjects: plasma melatonin, a biochemical marker in major depression. *Biol Psychiatry,* 19**,** 1215-28.

CONSORTIUM, G. T. 2013. The Genotype-Tissue Expression (GTEx) project. *Nature genetics,* 45**,** 580-585.

CRASSON, M., KJIRI, S., COLIN, A., KJIRI, K., L'HERMITE, M., ANSSEAU, M. & LEGROS, J.-J. 2004. Serum melatonin and urinary 6-sulfatoxymelatonin in major depression. *Psychoneuroendocrinology,* 29**,** 1-12.

CUTHBERT, B. N. & INSEL, T. R. 2013. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med,* 11**,** 126.

DEYOUNG, C. G., HIRSH, J. B., SHANE, M. S., PAPADEMETRIS, X., RAJEEVAN, N. & GRAY, J. R. 2010. Testing Predictions From Personality Neuroscience: Brain Structure and the Big Five. *Psychological Science,* 21**,** 820-828.

DIGMAN, J. M. 1990. Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology,* 41**,** 417-440.

DSM-5 2013. *Diagnostic and statistical manual of mental disorders : DSM-5*, Fifth edition. Arlington, VA : American Psychiatric Association, [2013].

DUDBRIDGE, F. 2013. Power and Predictive Accuracy of Polygenic Risk Scores. *PLOS Genetics,* 9**,** e1003348.

DUNCAN, L. E., RATANATHARATHORN, A., AIELLO, A. E., ALMLI, L. M., AMSTADTER, A. B., ASHLEY-KOCH, A. E., BAKER, D. G., BECKHAM, J. C., BIERUT, L. J., BISSON, J., BRADLEY, B., CHEN, C. Y., DALVIE, S., FARRER, L. A., GALEA, S., GARRETT, M. E., GELERNTER, J. E., GUFFANTI, G., HAUSER, M. A., JOHNSON, E. O., KESSLER, R. C., KIMBREL, N. A., KING, A., KOEN, N., KRANZLER, H. R., LOGUE, M. W., MAIHOFER, A. X., MARTIN, A. R., MILLER, M. W., MOREY, R. A., NUGENT, N. R., RICE, J. P., RIPKE, S., ROBERTS, A. L., SACCONE, N. L., SMOLLER, J. W., STEIN, D. J., STEIN, M. B., SUMNER, J. A., UDDIN, M., URSANO, R. J., WILDMAN, D. E., YEHUDA, R., ZHAO, H., DALY, M. J., LIBERZON, I., RESSLER, K. J., NIEVERGELT, C. M. & KOENEN, K. C. 2018. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry,* 23**,** 666-673.

HAMILTON, M. 1960. A rating scale for depression. *J Neurol Neurosurg Psychiatry,* 23**,** 56-62.

HONTI, F., MEADER, S. & WEBBER, C. 2014. Unbiased functional clustering of gene variants with a phenotypic-linkage network. *PLoS Comput Biol,* 10**,** e1003815.

KENNEDY, S. H., KUTCHER, S. P., RALEVSKI, E. & BROWN, G. M. 1996. Nocturnal melatonin and 24-hour 6-sulphatoxymelatonin levels in various phases of bipolar affective disorder. *Psychiatry Res,* 63**,** 219-22.

LE BIHAN, D., BRETON, E., LALLEMAND, D., GRENIER, P., CABANIS, E. & LAVAL-JEANTET, M. 1986. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology,* 161**,** 401-7.

LEWY, A. J., WEHR, T. A., GOLD, P. W. & GOODWIN, F. K. 1979. PLASMA MELATONIN IN MANIC-DEPRESSIVE ILLNESS. *In:* USDIN, E., KOPIN, I. J. & BARCHAS, J. (eds.) *Catecholamines: Basic and Clinical Frontiers.* Pergamon.

LOH, P.-R., BHATIA, G., GUSEV, A., FINUCANE, H. K., BULIK-SULLIVAN, B. K., POLLACK, S. J., DE CANDIA, T. R., LEE, S. H., WRAY, N. R., KENDLER, K. S., O'DONOVAN, M. C., NEALE, B. M., PATTERSON, N., PRICE, A. L. & SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C. 2015a. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics,* 47**,** 1385-1392.

LOH, P.-R., KICHAEV, G., GAZAL, S., SCHOECH, A. P. & PRICE, A. L. 2018. Mixed-model association for biobank-scale datasets. *Nature Genetics,* 50**,** 906-908.

LOH, P. R., TUCKER, G., BULIK-SULLIVAN, B. K., VILHJALMSSON, B. J., FINUCANE, H. K., SALEM, R. M., CHASMAN, D. I., RIDKER, P. M., NEALE, B. M., BERGER, B., PATTERSON, N. & PRICE, A. L. 2015b. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet,* 47**,** 284-90.

MAJOR DEPRESSIVE DISORDER WORKING GROUP OF THE PSYCHIATRIC, G. C. 2013. A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular psychiatry,* 18**,** 10.1038/mp.2012.21.

MARWAHA, S., HE, Z., BROOME, M., SINGH, S. P., SCOTT, J., EYDEN, J. & WOLKE, D. 2014. How is affective instability defined and measured? A systematic review. *Psychological Medicine,* 44**,** 1793-1808.

MONTELEONE, P., NATALE, M., LA ROCCA, A. & MAJ, M. 1997. Decreased nocturnal secretion of melatonin in drug-free schizophrenics: no change after subchronic treatment with antipsychotics. *Neuropsychobiology,* 36**,** 159-63.

MOORE, S. W., TESSIER-LAVIGNE, M. & KENNEDY, T. E. 2007. Netrins and Their receptors. *In:* BAGNARD, D. (ed.) *Axon Growth and Guidance.* New York, NY: Springer New York.

MORERA-FUMERO, A. L. & ABREU-GONZALEZ, P. 2013. Role of melatonin in schizophrenia. *International journal of molecular sciences,* 14**,** 9037-9050.

PACCHIEROTTI, C., IAPICHINO, S., BOSSINI, L., PIERACCINI, F. & CASTROGIOVANNI, P. 2001. Melatonin in Psychiatric Disorders: A Review on the Melatonin Involvement in Psychiatry. *Frontiers in Neuroendocrinology,* 22**,** 18-32.

PRICE, A. L., PATTERSON, N. J., PLENGE, R. M., WEINBLATT, M. E., SHADICK, N. A. & REICH, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet,* 38**,** 904-909.

RUBIN, R. T., HEIST, E. K., MCGEOY, S. S., HANADA, K. & LESSER, I. M. 1992. Neuroendocrine aspects of primary endogenous depression. XI. Serum melatonin measures in patients and matched control subjects. *Arch Gen Psychiatry,* 49**,** 558-67.

RUSH, A. J., TRIVEDI, M. H., WISNIEWSKI, S. R., NIERENBERG, A. A., STEWART, J. W., WARDEN, D., NIEDEREHE, G., THASE, M. E., LAVORI, P. W., LEBOWITZ, B. D., MCGRATH, P. J., ROSENBAUM, J. F., SACKEIM, H. A., KUPFER, D. J., LUTHER, J. & FAVA, M. 2006. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry,* 163**,** 1905-17.

SADOCK, K. 2000. *Kaplan & Sadock's comprehensive textbook of psychiatry*, Seventh edition. Philadelphia : Lippincott Williams & Wilkins, [2000] ©2000.

SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature,* 511**,** 421-427.

SCHMIDT, T., SAMARAS, P., FREJNO, M., GESSULAT, S., BARNERT, M., KIENEGGER, H., KRCMAR, H., SCHLEGL, J., EHRLICH, H.-C., AICHE, S., KUSTER, B. & WILHELM, M. 2017. ProteomicsDB. *Nucleic Acids Research,* 46**,** D1271-D1281.

SKLAR, P. 2011. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet,* 43**,** 977-983.

SRINIVASAN, V., SMITS, M., SPENCE, W., LOWE, A. D., KAYUMOV, L., PANDI-PERUMAL, S. R., PARRY, B. & CARDINALI, D. P. 2006. Melatonin in mood disorders. *World J Biol Psychiatry,* 7**,** 138-51.

STANLEY, M. & BROWN, G. M. 1988. Melatonin levels are reduced in the pineal glands of suicide victims. *Psychopharmacol Bull,* 24**,** 484-8.

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P., DANESH, J., DOWNEY, P., ELLIOTT, P., GREEN, J., LANDRAY, M., LIU, B., MATTHEWS, P., ONG, G., PELL, J., SILMAN, A., YOUNG, A., SPROSEN, T., PEAKMAN, T. & COLLINS, R. 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med,* 12**,** e1001779.

VILHJÁLMSSON, BJARNI J., YANG, J., FINUCANE, HILARY K., GUSEV, A., LINDSTRÖM, S., RIPKE, S., GENOVESE, G., LOH, P.-R., BHATIA, G., DO, R., HAYECK, T., WON, H.-H., RIPKE, S., NEALE, BENJAMIN M., CORVIN, A., WALTERS, JAMES T. R., FARH, K.-H., HOLMANS, PETER A., LEE, P., BULIK-SULLIVAN, B., COLLIER, DAVID A., HUANG, H., PERS, TUNE H., AGARTZ, I., AGERBO, E., ALBUS, M., ALEXANDER, M., AMIN, F., BACANU, SILVIU A., BEGEMANN, M., BELLIVEAU, RICHARD A., JR., BENE, J., BERGEN, SARAH E., BEVILACQUA, E., BIGDELI, TIM B., BLACK, DONALD W., BRUGGEMAN, R., BUCCOLA, NANCY G., BUCKNER, RANDY L., BYERLEY, W., CAHN, W., CAI, G., CAMPION, D., CANTOR, RITA M., CARR, VAUGHAN J., CARRERA, N., CATTS, STANLEY V., CHAMBERT, KIMBERLY D., CHAN, RAYMOND C. K., CHEN, RONALD Y. L., CHEN, ERIC Y. H., CHENG, W., CHEUNG, ERIC F. C., CHONG, SIOW A., CLONINGER, C. R., COHEN, D., COHEN, N., CORMICAN, P., CRADDOCK, N., CROWLEY, JAMES J., CURTIS, D., DAVIDSON, M., DAVIS, KENNETH L., DEGENHARDT, F., DEL FAVERO, J., DELISI, LYNN E., DEMONTIS, D., DIKEOS, D., DINAN, T., DJUROVIC, S., DONOHOE, G., DRAPEAU, E., DUAN, J., DUDBRIDGE, F., DURMISHI, N., EICHHAMMER, P., ERIKSSON, J., ESCOTT-PRICE, V., ESSIOUX, L., FANOUS, AYMAN H., FARRELL, MARTILIAS S., FRANK, J.,

FRANKE, L., FREEDMAN, R., FREIMER, NELSON B., FRIEDL, M., FRIEDMAN, JOSEPH I., FROMER, M., GENOVESE, G., GEORGIEVA, L., GERSHON, ELLIOT S., GIEGLING, I., GIUSTI-RODRGUEZ, P., GODARD, S., GOLDSTEIN, JACQUELINE I., GOLIMBET, V., GOPAL, S., GRATTEN, J., GROVE, J., DE HAAN, L., et al. 2015. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics,* 97**,** 576-592.

WANG, Q., ARMENIA, J., ZHANG, C., PENSON, A. V., REZNIK, E., ZHANG, L., MINET, T., OCHOA, A., GROSS, B. E., IACOBUZIO-DONAHUE, C. A., BETEL, D., TAYLOR, B. S., GAO, J. & SCHULTZ, N. 2018. Unifying cancer and normal RNA sequencing data from different sources. *Scientific Data,* 5**,** 180061.

WOOLRICH, M. W., JBABDI, S., PATENAUDE, B., CHAPPELL, M., MAKNI, S., BEHRENS, T., BECKMANN, C., JENKINSON, M. & SMITH, S. M. 2009. Bayesian analysis of neuroimaging data in FSL. *Neuroimage,* 45**,** S173-86.

WRAY, N. R., RIPKE, S., MATTHEISEN, M., TRZASKOWSKI, M., BYRNE, E. M., ABDELLAOUI, A., ADAMS, M. J., AGERBO, E., AIR, T. M., ANDLAUER, T. M. F., BACANU, S. A., BAEKVAD-HANSEN, M., BEEKMAN, A. F. T., BIGDELI, T. B., BINDER, E. B., BLACKWOOD, D. R. H., BRYOIS, J., BUTTENSCHON, H. N., BYBJERG-GRAUHOLM, J., CAI, N., CASTELAO, E., CHRISTENSEN, J. H., CLARKE, T. K., COLEMAN, J. I. R., COLODRO-CONDE, L., COUVY-DUCHESNE, B., CRADDOCK, N., CRAWFORD, G. E., CROWLEY, C. A., DASHTI, H. S., DAVIES, G., DEARY, I. J., DEGENHARDT, F., DERKS, E. M., DIREK, N., DOLAN, C. V., DUNN, E. C., ELEY, T. C., ERIKSSON, N., ESCOTT-PRICE, V., KIADEH, F. H. F., FINUCANE, H. K., FORSTNER, A. J., FRANK, J., GASPAR, H. A., GILL, M., GIUSTI-RODRIGUEZ, P., GOES, F. S., GORDON, S. D., GROVE, J., HALL, L. S., HANNON, E., HANSEN, C. S., HANSEN, T. F., HERMS, S., HICKIE, I. B., HOFFMANN, P., HOMUTH, G., HORN, C., HOTTENGA, J. J., HOUGAARD, D. M., HU, M., HYDE, C. L., ISING, M., JANSEN, R., JIN, F., JORGENSON, E., KNOWLES, J. A., KOHANE, I. S., KRAFT, J., KRETZSCHMAR, W. W., KROGH, J., KUTALIK, Z., LANE, J. M., LI, Y., LI, Y., LIND, P. A., LIU, X., LU, L., MACINTYRE, D. J., MACKINNON, D. F., MAIER, R. M., MAIER, W., MARCHINI, J., MBAREK, H., MCGRATH, P., MCGUFFIN, P., MEDLAND, S. E., MEHTA, D., MIDDELDORP, C. M., MIHAILOV, E., MILANESCHI, Y., MILANI, L., MILL, J., MONDIMORE, F. M., MONTGOMERY, G. W., MOSTAFAVI, S., MULLINS, N., NAUCK, M., NG, B., et al. 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet,* 50**,** 668-681.