Genes associated with anhedonia: a new analysis in a large clinical trial (GENDEP)

Ren, Hongyan; Fabbri, Chiara; Uher, Rudolf; Rietschel, Marcella; Mors, Ole; Henigsberg, Neven; Hauser, Joanna; Zobel, Astrid; Maier, Wolfgang; Dernovsek, Mojca Z.; ...

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:105:522445

ARTICLE

Genes associated with anhedonia: a new analysis in a large clinical trial (GENDEP)

Hongyan Ren[1], Chiara Fabbri[2], Rudolf Uher[3], Marcella Rietschel [4], Ole Mors[5], Neven Henigsberg [6], Joanna Hauser[7], Astrid Zobel[8], Wolfgang Maier[8], Mojca Z. Dernovsek[9], Daniel Souery[10], Annamaria Cattaneo[11], Gerome Breen[2], Ian W. Craig[2], Anne E. Farmer[2], Peter McGuffin[2], Cathryn M. Lewis [2] and Katherine J. Aitchison [1,2]

Abstract

A key feature of major depressive disorder (MDD) is anhedonia, which is a predictor of response to antidepressant treatment. In order to shed light on its genetic underpinnings, we conducted a genome-wide association study (GWAS) followed by investigation of biological pathway enrichment using an anhedonia dimension for 759 patients with MDD in the GENDEP study. The GWAS identified 18 SNPs associated at genome-wide significance with the top one being an intronic SNP (rs9392549) in *PRPF4B* (pre-mRNA processing factor 4B) located on chromosome 6 ($P = 2.07 \times 10^{-9}$) while gene-set enrichment analysis returned one gene ontology term, axon cargo transport (GO: 0008088) with a nominally significant P value (1.15×10^{-5}). Furthermore, our exploratory analysis yielded some interesting, albeit not statistically significant genetic correlation with Parkinson's Disease and nucleus accumbens gray matter. In addition, polygenic risk scores (PRSs) generated from our association analysis were found to be able to predict treatment efficacy of the antidepressants in this study. In conclusion, we found some markers significantly associated with anhedonia, and some suggestive findings of related pathways and biological functions, which could be further investigated in other studies.

Introduction

Major depressive disorder (MDD) is chronic illness which affects 350 million people world-wide according to an estimate by the World Health Organization (WHO); it is characterized by depressed mood, diminished interest, impaired cognitive function, and somatic symptoms, such as disturbed sleep or appetite. The aetiology of MDD is multifactorial with a heritability estimated to be approximately 35%[1,2]. It is generally recognized that MDD is a common illness involving multiple common genetic variants with small to moderate effect size[3]. Indeed, several large cohort-based genome-wide association studies (GWASs) in recent years have identified signals which shed new light on our

understanding of MDD, for example, implicating the presynaptic protein piccolo and alpha-1 subunit of a voltage-dependent calcium channel in the pathogenesis of MDD[4,5] and shared genetic risk for MDD, bipolar disorder and schizophrenia[6,7]. However, current studies still fall far short of accounting for all of the genetic variation in MDD with robust replicated findings. One of the possible reasons could be that the majority of these studies chose a dichotomous phenotype such as diagnosis as their outcome measure, with the currently limited understanding of the disorder leading to heterogeneity in diagnostic ascertainment[8]. Of interest, using a polygenic risk score (PRS) for schizophrenia (SCZ), Whalley et al. (2016) identified a subgroup of patients with MDD that had a higher polygenic risk of SCZ than others; this subgroup of MDD patients also showed an attenuated level of distress and neuroticism[9]. Instead of a case-control design, some studies choose quantitative traits (QTs) related to illness to increase the

Correspondence: Katherine J. Aitchison (kaitchis@ualberta.ca)
[1]Psychiatry and Medical Genetics, University of Alberta, Edmonton, AB, Canada
[2]MRC Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK
Full list of author information is available at the end of the article.

power of the analysis. Quantitative variables have a higher information content than categorical variables; association studies using QTs can therefore increase the statistical power four to eight-fold, with a resultant proportional reduction of the required sample size[10]. For example, one study used hippocampal atrophy measured by MRI as a QT for Alzheimer's disease in a GWAS of only moderate sample size and nonetheless identified novel candidate loci attaining genome-wide significance[11,12].

Different kinds of studies have long indicated that anhedonia is a fundamental feature of MDD[13,14]. DSM-IV-TR defines anhedonia as diminished interest or pleasure in response to stimuli that were previously perceived as rewarding during a premorbid state[15]. Moreover, anhedonia has been shown to be able to predict a longer time to remission and fewer depression-free days[16,17]. Specifically, using the same dataset as in our present study, Uher et al. showed that out of the six disease dimensions (mood, anxiety, pessimism, interest-activity, sleep, and appetite), the interest-activity dimension (anhedonia) at baseline was the only dimension able to predict poor treatment outcome in the later time points[17]. Both twin and family studies demonstrate that 44% of anhedonia is attributable to genetic factors, especially additive genetic effects, and first-degree relatives of patients with MDD display anhedonia-related phenotypes when compared to controls[18,19]. Although different threads of evidence have validated anhedonia as a QT of MDD, no genetic or genomic study has yet been carried out to identify candidate loci associated with this key feature of MDD.

Our study used a dimensional score of anhedonia to conduct a GWAS and to estimate the heritability of this phenotype accounted for by common variants, aiming to shed new light on our understanding of MDD.

Materials and methods

Patient recruitment

Seven hundred and ninety-six people (296 males, 500 females) with unipolar depression of at least moderate severity according to ICD-10 (International Classification of Diseases, 10th revision, Mental and Behavioural Disorders, Research Criteria) and DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, fourth edition) criteria[20] were recruited from eight European countries in the GENDEP project[21]. All patients were of European ancestry without a family history of schizophrenia, bipolar disorder or a current dependency on alcohol or drugs. For further details about the GENDEP project, see Uher et al.[21,22].

Phenotype definition

Uher et al. conducted factor analysis of depression severity data generated from three measures: the

Montgomery-Asberg Depression Scale (MADRS); the Hamilton Depression Rating Scale (HDRS) and the Beck Depression Inventory (BDI). Although these measures had previously been widely used in studies of depression, prior to GENDEP, no study had used all three measures simultaneously. Six dimensions with continuous factor scores representing the different aspects of the psychopathology of depression were extracted from initial questionnaire estimates[23]. Of these six dimensions, the interest-activity score at baseline (which had higher information loadings from items in the three measures relevant to anhedonia, such as "inability to feel", "lassitude" in the MADRS; "sexual interest" in the HAMD-17; "enjoyment" and "interest in people" in the BDI) was found to significantly predict response to treatment with the antidepressants used in the study[17]. In this analysis, we used the baseline interest-activity score as our outcome measure for GWAS.

DNA extraction and genotyping

DNA was extracted from blood samples collected in ethylenediaminetetraacetic acid (EDTA) tubes using standard procedures[24]; genotyping was performed in the Centre National de Génotypage using the Illumina Human610-quad bead chip (Illumina) as described[25].

Genotype quality control and population stratification analysis

Standard steps were taken for quality control of genomic data in PLINK 1.09[26] and data were excluded on failure to pass the following thresholds: consistency of gender information between genomic data and demographic data, SNP genotyping rate ≥ 95%, individual genotyping rate ≥ 97%, Hardy–Weinberg equilibrium test ($P \geq 0.001$), minor allele frequency (MAF) ≥ 0.01. Furthermore, using both PLINK 1.09 and KING[27], pairwise identity-by-state (IBS) was calculated and outliers or subjects showing unknown familial relationship with others (proportion IBD > 0.05) were subsequently excluded[26,28].

Population stratification analysis

Population stratification analysis was conducted using EIGENSTRAT[29], which employs principal component analysis (PCA) to capture hidden population structure in genomic data. Prior to the analysis, data were pruned to make sure adjacent SNPs were in no more than weak linkage disequilibrium (LD) with each other (PLINK command: --indep-pairwise 50 10 0.5)[30]. This generated 20 principal components (PCs) which were controlled for as covariates in the subsequent association analysis.

Imputation of missing genotypes using the 1000 Genomes dataset

Following quality control steps, imputation was carried out on genomic data. We employed IMPUTE2 + SHAPEIT2 to impute using the 1000 Genomes phase 3 dataset as the reference dataset[31,32]. Before imputation, the physical position of SNPs was updated using UCSC Liftover tool (https://genome.ucsc.edu/)[33] to the haploid human genome build 19 (hg19). Following imputation, the same quality control steps were used to clean the resultant imputed data.

GWAS using a linear mixed model (LMM)

In order to test for genotype–phenotype association while controlling for potential confounding factors such as population structure, family structure, and cryptic relatedness simultaneously, we used factored spectrally transformed LMM (FaST-LMM) for our association study[34]. In brief, the LMM log likelihood of the phenotype data, y (dimension $n \times 1$; n denoting the cohort size), given fixed effects X (dimension $n \times d$; d denoting the number of fixed effects in a single model, including the offset, the covariates, and the SNP to be tested), can be written as

$$LL\left(\delta_e^2, \delta_g^2, \boldsymbol{\beta}\right) = logN\left(\boldsymbol{y}|\boldsymbol{X}\beta; \delta_g^2\boldsymbol{K} + \delta_e^2\boldsymbol{I}\right) \qquad (1)$$

where N ($\boldsymbol{r}|\boldsymbol{m}$; Σ) denotes a normal distribution of variable $\boldsymbol{r}$ with mean $\boldsymbol{m}$ and covariance matrix Σ; $\boldsymbol{K}$ (dimension $n \times n$) is the genetic similarity matrix; $\boldsymbol{I}$ is the identity matrix; δ_e^2 is the magnitude of the residual variance; δ_g^2 is the magnitude of the genetic variance; and $\boldsymbol{\beta}$ (dimension $\boldsymbol{d} \times 1$) denotes the weight of the fixed effects.

The "Fa" in FaST-LMM stands for factorization. Let S be genetic similarity matrix, as the covariance matrix of the normal distribution becomes a diagonal matrix $\boldsymbol{S} + \delta\boldsymbol{I}$ (spectral decomposition), the log likelihood can be rewritten as the sum over n terms. Factorization dramatically increases the size of datasets that can be analyzed with LMM, and additionally enhances the speed and feasibility of the analysis.

In our analysis, we chose the continuous interest-activity score as our outcome measure, controlling for gender, age, years of education, recruitment centres and the first 20 PCs from EIGENSTRAT as covariates.

Replication analysis using STAR*D

Following our initial findings, we used data from the Sequenced Treatment Alternatives to Relieve Depression Study (STAR*D) to replicate our primary results. Detailed information about the STAR*D including its demographic characteristics and genomic profile have been previously described[35–37]. In brief, 1351 patients with MDD were recruited with the phenotype being defined as the sum of items with corresponding content in baseline HAMD-17, QIDS-SR, QIDS-C and the research outcome assessor-rated 30-item Inventory for Depression Symptomatology[17], with genomic profile including 7405247 SNPs after quality control and imputation[38]. Further, a linear model using PLINK 1.09 was chosen with age, gender, years of education, recruitment centre, and the first four population PCs being included as covariates.

Gene-based and pathway analysis

Emerging evidence has suggested that disease- or trait-associated genetic variants identified from GWASs tend to be enriched in genic regions including multiple associated variants at a single locus[39,40]. Therefore, we utilized fastBAT which stands for a fast and flexible set-Based Association Test and the P values from the LMM analysis for gene-set testing[41] to discover genes associated with the interest-activity score based on the aggregated effect of a set of SNPs (e.g., SNPs within or close to a gene) with their generated P values being adjusted using Bonferroni correction (0.05/22484).

Biological interpretation, heritability, and genetic correlation estimates

In order to further understand the resultant signals and their associations with the interest-activity score, we chose loci with an association P value less than 1×10^{-5} and used DEPICT (Data-driven Expression Prioritized Integration for Complex Traits) to accomplish gene prioritization and tissue/cell type enrichment analysis with a false discovery rate (FDR) set as 1%[42]. Recent studies have shown that mutation-intolerant genes which are presumed to hold critical biological functions are enriched in rare variants in psychiatric disorders such as autism and intellectual disability (ID)[43,44]; this pattern also extends to both rare and common variants for schizophrenia[45]. To test whether it also holds for common variants in our MDD-related phenotype, we investigated the enrichment of genes harboring SNPs attaining an association P value $\leq 10^{-5}$ in the set of loss-of-function (LOF) genes characterized by the Exome Aggregation Consortium (ExAC), setting the constraint metric pLI $\geq$ 0.9 (probability of being LoF intolerant) according to their recommendation[46].

Furthermore, aiming to detect phenotypic variance explained by common SNPs (h_g) in our sample and to explore traits which shared a common genetic effect with the interest-activity score, we chose LDSc (LD score regression)[47] from LD Hub—a centralized database of summary-level GWAS results for 177 diseases/traits from different publicly available resources/consortia and a web interface that automates the LD score regression analysis pipeline for detection of h_g and genetic correlation between target phenotype and multiple traits[48].

Association analysis with longitudinal change of anhedonia following treatment with antidepressants

The baseline interest-activity score from the study by Uher el al.[17], chosen as the primary outcome measure in our GWAS, was found to significantly predict treatment response in both GENDEP and STAR*D. In order to investigate the potential association between the SNPs associated with the baseline interest-activity score and longitudinal change in the score, we summed up all the associated SNPs to calculate a unweighted PRS for each individual, then conducted association analysis between

this PRS and the interest-activity score from week 1 to week 10 using a LMM. The fixed effects of the model included our predictor (PRS) and covariates (age, quadratic effect of age, gender, baseline interest-activity score, and centerid) while the random effect included a random intercept and a random time effect (slope). The PRS was generated using PLINK[26] and the above-mentioned association analysis was implemented using the package "nlme" in an R environment[49].

Results

Demographic characteristics and genome-wide association analysis

Demographic characteristics

Of 796 people with genomic data, 759 had a baseline interest-activity score derived from factor analysis (286 males, 473 females). The mean age was 42.05 (11.59), mean years of education 12.31 (3.12), mean baseline MADRS 28.90 (6.77), mean baseline HDRS 21.88 (5.24), and mean baseline BDI was 28.10 (9.76).

Genome-wide association analysis

After imputation and quality control, 1,313,135 SNPs (of which 789,990 were imputed with high-quality imputation, i.e., info > 0.6, LD pruning at $R^2 < 0.5$) in 760 individuals remained in the present analysis and as shown in Fig. 1, all study subjects were of European ancestry with no gross population stratification.
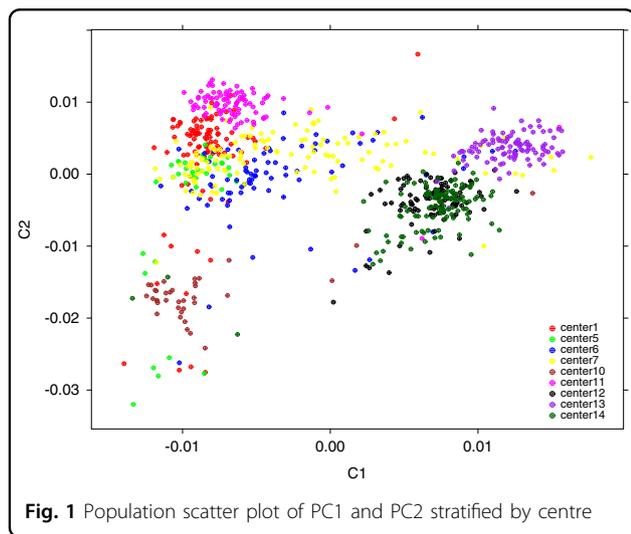
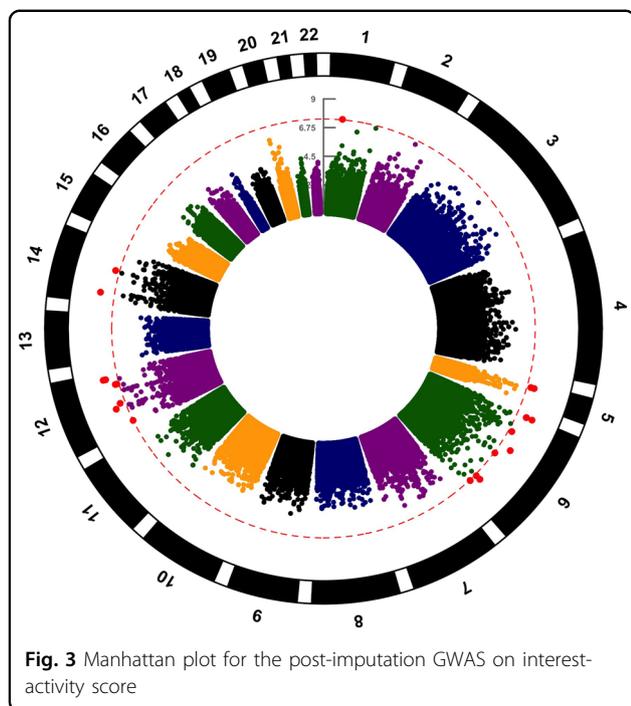**Fig. 1** Population scatter plot of PC1 and PC2 stratified by centre
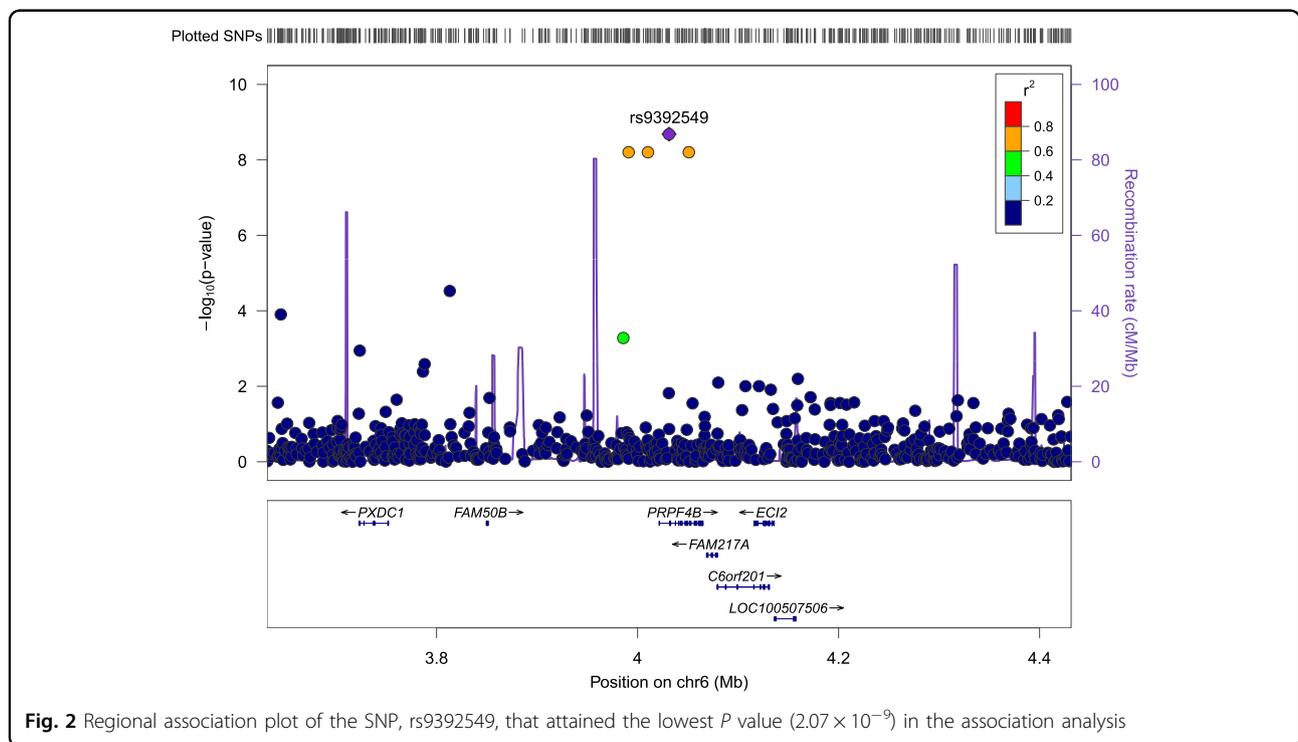
Fig. 2 Regional association plot of the SNP, rs9392549, that attained the lowest P value (2.07×10^{-9}) in the association analysis

Table 1　Significant signals arising from the GWAS of the interest-activity score in GENDEP and corresponding replication in STAR*D

SNP	Chr	ChrPos	MAF	Reference allele	Position	Nearest gene	Impute info	GENDEP (discovery stage)			STAR*D (replication stage)		
								P Value	Beta	SE	P value	Beta	SE
rs9392549	6	4031372	0.01974	A	Intronic	PRPF4B	0.987	2.07E−09	0.19	0.03	0.031	0.22	0.1006
rs650466	12	118646661	0.01581	T	Intronic	TAOK3	0.905	2.83E−09	0.19	0.03	0.576	0.0715	0.1278
rs10498321	14	34194231	0.1303	C	Intronic	NPAS3	1	3.08E−09	0.19	0.03	0.8461	−0.01086	0.05596
rs116367381	6	65987638	0.01008	T	Intronic	EYS	0.641	4.42E−09	0.19	0.03	0.5175	−0.1005	0.1552
rs146142861	12	50821551	0.01546	T	Intronic	LARP4	0.642	5.46E−09	0.19	0.03	NA	NA	NA
rs77266588	6	142944380	0.02905	C	intronic	LOC153910	0.9	1.16E−08	0.18	0.03	0.05122	−0.1913	0.1831
rs147128318	12	61824491	0.01273	T	intergenic	FAM19A2	0.669	1.52E−08	0.18	0.03	NA	NA	NA
rs112593182	14	71425668	0.01184	T	intergenic	COX16	0.827	1.85E−08	0.18	0.03	0.2964	−0.1913	0.1831
rs7732588	5	19767062	0.01598	A	intronic	CDH18	0.9	1.90E−09	0.18	0.03	0.3448	−0.1127	0.1193
rs831431	12	104029996	0.02237	C	intronic	STAB2	0.961	1.92E−08	0.18	0.03	0.046[a]	−0.32	0.1598
rs80074882	6	32184329	0.02078	G	intronc	NOTCH4	0.69	2.09E−08	0.18	0.03	0.8249	0.03	0.1225
rs73187882	12	104013517	0.02656	G	intronic	STAB2	0.9	2.57E−08	0.18	0.03	0.9875	0.001458	0.09282
rs11755163	6	142680533	0.06824	G	intronic	ADGRG6	0.663	2.88E−08	0.18	0.03	0.3793	0.07402	0.08417
rs117555972	6	82161785	0.01123	C	intergenic	FAM46A	0.796	3.30E−08	0.18	0.03	0.6524	0.05853	0.1299
rs78164217	6	155798386	0.01189	G	intergenic	COX16	0.765	3.33E−08	0.18	0.03	0.9439	0.009437	0.1341
12:15021349	12	15021349	0.01126	A	NA	NA	0.751	3.79E−08	0.18	0.03	NA	NA	NA
rs118166313	6	155772619	0.01583	T	intronic	NOX3	0.771	4.18E−08	0.17	0.03	0.9499	0.008294	0.1319
rs138600970	6	23054276	0.01061	A	intronc	LOC105374974	0.651	4.98E−08	0.17	0.03	0.1227	0.2776	0.1798

[a] rs118190482 (in LD with rs831431, $R^2 = 0.5$)

Fig. 3 Manhattan plot for the post-imputation GWAS on interest-activity score

**Fig. 4** Quantile-Quantile (QQ) plot for the post-imputation GWAS of the interest-activity score

Association analysis of interest-activity scores using LMM identified 18 SNPs that passed genome-wide significance (5×10^{-8}) when including gender, age, years of education and 20 PCs of the population structure from EIGENSTRAT as covariates. The top SNP from the analysis, rs9392549, in an intronic region of *PRPF4B* (pre-mRNA processing factor 4B) located on chromosome 6, had a *P* value of 2.07×10^{-9} (Figure 2). Table 1 summarizes the top signals from the association analysis and Fig. 3 displays this as a circularized Manhattan plot. The genomic inflation factor (λ) was calculated as an index of any potential confounding effect in the analysis, and the results were consistent with potential confounding effects having been adequately covered ($\lambda = 0.9958$, Fig. 4).

The replication analysis using the STAR*D dataset indicated that while none of the associated SNPs found in the GENDEP dataset were replicated at a Bonferroni-adjusted significance level ($0.03/18 = 0.0016$); two of them, the top signal (rs9392549) and rs118190482 located in the intronic region of *STAB2* (in LD with rs831431, $R^2 = 0.5$), were nominally significant ($P = 0.03$ and 0.046 respectively, in Table 1).

Gene-based and gene prioritization analysis

Gene-based association analysis indicated no gene was associated at gene-level significance (*P* value $= 2 \times 10^{-6}$). The gene with the strongest signal from the analysis was *KITLG* on chromosome 12 (KIT ligand, *P* value $= 3.09 \times 10^{-5}$).
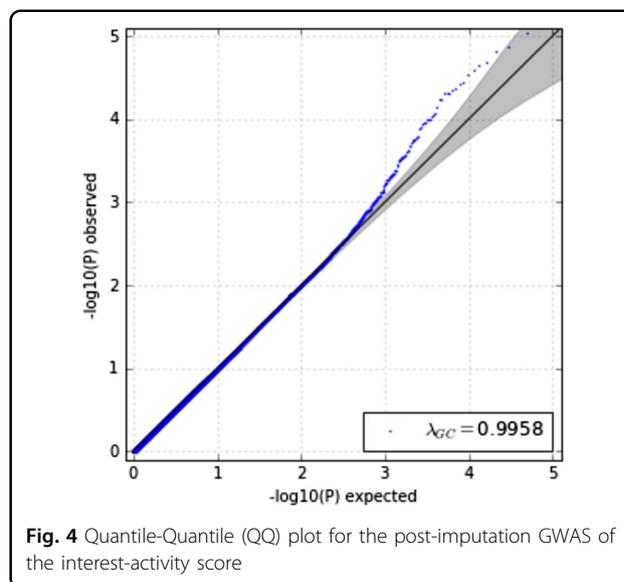
Using DEPICT, one SNP, rs1001415, which is intronic in *EFCAB2* (EF-hand calcium binding domain 2) on chromosome 1, was prioritized owing to sharing more similar biological functions with other associated loci, although the *P* value was only at a trend level (nominal P $= 0.09$). Interestingly, Westra et al. reported that rs1001415 is in high LD with a cis eQTL SNP (rs4658697) in an intronic region of a transcript (NM 001143943.1) of *EFCAB2*[50]. Furthermore, gene-set analysis found one gene ontology item (GO:0008088), axon cargo transport, was over-represented by associated loci from our association analysis with a nominal *P* value being 1.15×10^{-5}. Cell/tissue annotation analysis saw our associated loci were highly annotated in the MeSH first term of "hypothalamus" and the MeSH second term of "nervous system" (nominal *P* = 0.004). Although some results generated from DEPICT showed nominal significance, they failed to reach FDR. Nevertheless, our target genes were shown to be significantly enriched by the gene set (3203 genes) characterized by ExAC as mutation intolerant ($P = 0.001$).

Heritability estimation and genetic correlation analysis

Estimation of h_g showed that 69% of the phenotypic variance of the interest-activity dimension in our sample could be explained by common SNPs ($h_g = 0.69 \pm 0.88$). As shown in Table S1 and Figure S1, the genetics of the interest-activity score was highly positively correlated with Parkinson's disease (PD) (rg $= 0.83$, se $= 1.14$), and with Alzheimer's disease (rg $= 0.43$, se $= 0.32$). Moreover, its genetics was negatively correlated with that of the gray matter volume of nucleus accumbens (rg $= -0.6492$, se $= 0.84$), eczema (rg $= -0.41$, se $= 0.44$) and with subjective well-being (rg $= -0.32$, se $= 0.47$). This is consistent with

a pleiotropic effect. However, the results should be interpreted with caution given that none of the P values generated from our genetic correlation analyses reached the statistical significance of 0.05.

Association analysis between the PRS and longitudinal change of anhedonia up to ten weeks following antidepressant treatment

The association analysis showed that the PRS calculated based on the GWAS of baseline interest-activity score was significantly associated with longitudinal change of anhedonia following antidepressant treatment ($\beta = 1.73$, $P = 0.0023$). In order to evaluate if the top hit (rs9392549) from the GWAS of baseline interest-activity score solely drove the identified association, we conducted a secondary analysis using same model conditioning on rs9392549; the association between the PRS and longitudinal change of anhedonia remained significant ($\beta = 1.64$, $P = 0.0091$).

Discussion

To the best of our knowledge, this is the first genome-wide association analysis of anhedonia in patients with MDD. We used a LMM to conduct the association analysis, which identified 18 SNPs of genome-wide significance, with the most significant being rs9392549 in an intronic region of *PRPF4B* on chromosome 6 ($P = 2.07 \times 10^{-9}$). Although no gene was significant on gene-set testing, gene prioritization analysis found one intronic SNP (rs1001415) in *EFCAB2* to be significant with a trend ($P = 0.09$) and the associated loci showed enrichment for a particular gene ontology locus, axon cargo transport (GO:0008088). Furthermore, using LD regression, we showed that 69% of the variance in our phenotype was explained by common SNPs and the markers associated with anhedonia were positively correlated with PD and with Alzheimer's disease, while being negatively correlated with nucleus accumbens gray matter volume.

The use of a LMM for the genome-wide association analysis is in contrast to the classic general linear model (GLM) in how population stratification or other sample structure issues are addressed. Such confounding factors are detected and addressed in GLM by using genomic control[51], ancestry inference (analysis of population-structure)[52–54] and PCA[29,55]. However, these strategies fail to account for sample features such as family structure or cryptic relatedness; for population stratification owing to ancient population divergence, methods like genomic control are relatively weak[56]. Linear mixed modeling by contrast fits population structure as a fixed effect and a similarity matrix between individuals as the variance-covariance structure of the random effect[57]; such a method has been shown to yield more a conservative λ_{GC} compared to other approaches[57,58]. Using a similar statistical model, the CONVERGE consortium conducted a

genome-wide association analysis in a large cohort of Chinese female patients with severe MDD, with two significant loci being identified and replicated in different samples[59]. These two loci (rs12415800 and rs35936514 on chromosome 10), however, were not replicated in our study given the rarer frequency of these loci in the European population.

One intronic SNP (rs9392549) in *PRPF4B* yielded the lowest P value in association with anhedonia ($P = 2.07 \times 10^{-9}$, replicated $P = 0.03$). *PRPF4B*, pre-mRNA processing factor 4 homolog B, is a kinase involved in mRNA splicing that is involved in biological pathways such as inositol phosphate metabolism[60]. Patients with MDD have been shown to have alterations in mRNA splicing, especially in that of neurotransmitter receptors[61,62]. For instance, in suicide victims with a history of major depression, adenosine-to-inosine RNA editing within the coding sequence of the serotonin 2C receptor (5-HT$_{2C}$) pre-mRNA was significantly decreased and this effect was reversed by treatment with the antidepressant fluoxetine[63]. Additionally, inositol phosphate has been repeatedly implicated in the pathophysiology of affective disorders including MDD, with potential new treatments arising[64–66]. For example, a double-blind, controlled clinical trial in MDD indicated that the overall improvement in scores on the Hamilton Depression Rating Scale was significantly greater for inositol than for placebo after 4 weeks of treatment[67].

One of two associated loci which were replicated with a nominal significance, rs831431 ($P = 1.92 \times 10^{-8}$, replicated $P = 0.046$) is a brain eQTL located in the intronic region of *STAB2*, which encodes stabilin 2. Stabilin 2 plays a critical role in angiogenesis[68]. According to BRAINEAC[69], rs831431 significantly affects the expression of one of *STAB2*'s transcripts (tID = 3429159), especially in the thalamus (eQTL $P = 0.01$). Although the precise role of *STAB2* in the pathogenesis of MDD or anhedonia still remains unclear, it could be hypothesized that deficits in neuroplasticity, potentially mediated by abnormal angiogenesis lead to dysfunction in pleasure-rewarding circuitry. This could be in a temporal-specific manner, analogous to the time-dependent gene expression that is commonly seen in genes related to neurodevelopment[70].

Of the other associated loci, rs10498321 is in an intronic region of *NPAS3*. *NPAS3*, neuronal PAS domain protein 3, is a brain-enriched transcription factor, expression deficits in which can cause deficiency in neurogenesis, especially in the hippocampus[71].

To date, *NPAS3* has been mainly studied in schizophrenia and bipolar disorder[72–74] and schizophrenia, especially with negative symptomatology, is another condition in which anhedonia may be a common feature; to our knowledge, this is the first report of an association between *NPAS3* and a MDD-related phenotype.

Intriguingly, one of the top signals (rs7973260)[75] identified in a GWAS of depressive symptoms in a large cohort from the UK Biobank is in the 18 kb downstream of rs650466, quasi-replicating the current finding and highlighting the potential importance of this genomic region in understanding the biological mechanism of MDD.

Given the modest replication using STAR*D, we carried out a genetic correlation analysis between STAR*D and GENDEP by executing the "sumsum" command in PRSice[76], which takes respective summary statistics as input. The result displayed in Figure S2 indicated that although the two datasets were significantly correlated with each other at multiple P-value thresholds (P_T at 0.04, 0.05, 0.2, 0.3, and 0.5), the variance explained by each other (R^2) was relatively small, which may at least partly explain the relatively weak replication signal in STAR*D. Although it has been widely thought that QTs underpinning the symptomatology of psychiatric disorders could increase the power of the identification of risk variants, the way in which QTs are established has been inconsistent. Of note, the QT of anhedonia was defined in contrasting ways in GENDEP and STAR*D owing to differential measures available. While our study provides an alternative approach for GWAS with limited sample size, it points to the importance of future efforts to validate different measures of QTs along the lines of the RDoC strategy[77].

In our gene prioritization analysis, one intronic SNP (rs1001415) in *EFCAB2* was found to be more similarly associated with other associated loci in terms of biological function. *EFCAB2*, EF-hand calcium binding domain 2, is located in SOR (smallest overlapping region) at 1q44 with three other genes: *HNRNPU*, *FAM36A*, *NCRNA00201*. Patients with microdeletions of this region display ID and seizures[78,79], which implies a role in neurodevelopment and cognitive function. Of note, it is in high LD with one cis eQTL SNP (rs4658697); therefore, we suggest that future studies could use rs1001415 as a proxy for rs4658697 for the expression of *EFCAB2*. In addition, one gene set (GO:0008088, axon cargo transport) was over-represented by our associated markers. It is therefore possible that dysfunctional axon cargo transport affected by our identified genes in brain regions relevant for reward circuitry[80,81] may be associated with impaired neurotransmitter release (dopamine, etc.), putatively leading to anhedonic symptoms.

Although the cross-phenotype LD score regression failed to generate a genetic correlation with a significant P value, it provided a trend worth further elaboration. Specifically, anhedonia in our study was positively correlated with PD (rg = 0.8). In fact, anhedonia independent of clinical diagnosis and PD are both dopamine-dependent processes and anhedonia is one of the most commonly observed non-motor symptoms in PD[82,83]. Moreover, anhedonia was negatively correlated with nucleus accumbens gray matter volume (rg = −0.6). The accumbens is a key structure in the reward circuit; structural and functional changes in the accumbens have been repeatedly implicated in substance abuse-related and MDD-related anhedonia[84,85]. Nonetheless, any inference from our current findings should be made with the caveat that due to the lack of statistical significance, potential type I error (false positive error) cannot be excluded.

Furthermore, the significant association detected between our PRS and the longitudinal change in anhedonia is of interest in that it appears to offer insight not only into the polygenic underpinnings of anhedonia but also into its change during treatment. This preliminary association analysis of the PRS generated by our association findings illustrates the potential of applying such a polygenic profile to better our prediction of treatment response. This could be further tested in the response to treatment of other disorders in which anhedonia is also a feature.

Strengths and limitations

Strengths of our study include the LMM which controls for confounding factors such as population stratification and cryptic relatedness in a perhaps more robust manner than GLM. However, there are limitations. Firstly, the sample size for our study is modest. Generally, the majority of power calculations used for GWAS employ a case-control design; the use of an endophenotype such as anhedonia, a QT of complex disease biologically hypothesized to be closer to underlying genetic variation, should increase the power of association[10]. Many approaches for linear mixed modeling of GWAS are computationally challenging, which makes such methodology less popular for GWAS of large sample sizes. Our study provided another new association strategy for GWAS of modest sample sizes, although replication of significant signals in a larger independent sample is required.

Secondly, 16 out 18 SNPs identified in the association analysis have a MAF lower than 0.03. The MAF distribution of our genomic data indicated that 67% of alleles fall into the interval between 0.01 and 0.05 (Figure S3). Enrichment of signals in the lower bound of the MAF spectrum is methodologically recognized; we are aware that given the sample size, these associations may be false positives (a "winner's curse"), as the number of individuals with a minor allele is very limited.

Thirdly, not all patients were drug-free at the time of recruitment (baseline), some medications such as antidepressants[86,87,88] or benzodiazepines[89] etc. might affect patients' anhedonia level at the baseline.

Conclusion

In summary, this first GWAS of anhedonia in MDD identified a number of SNPs attaining genome-wide significance. The top hits include loci such as *NAPS3* which has been associated with schizophrenia, another condition in which anhedonia may be a prominent feature. It is therefore possible that our findings are relevant not only for anhedonia in MDD, but also for anhedonia in other neuropsychiatric conditions. Consistent with this, cross-phenotype correlation analysis gave suggestive signals for PD and nucleus accumbens size. We suggest that further genetic exploration of anhedonia in MDD and other disorders could be a new and productive avenue that could lead to new treatments for this disabling feature of many neuropsychiatric conditions.

Author details
[1]Psychiatry and Medical Genetics, University of Alberta, Edmonton, AB, Canada. [2]MRC Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [3]Psychiatry Department, Dalhousie University, Halifax, NS, Canada. [4]Division of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Mannheim, Germany. [5]Clinical Medicine, Aarhus University, Aarhus, Denmark. [6]Croatian Institute for Brain Research, University of Zagreb, Zagreb, Croatia. [7]Psychiatry Department, University of Poznan, Poznan, Poland. [8]Psychiatry Department, University of Bonn, Bonn, Germany. [9]University Psychiatric Clinic, University of Ljubljana, Ljubljana, Slovenia. [10]Psychological Medicine, Free University of Brussels, Brussels, Belgium. [11]IRCCS, University of Brescia, Brescia, Italy

References
1. Bierut, L. J. et al. Major depressive disorder in a community-based twin sample: are there different genetic and environmental contributions for men and women? *Arch. Gen. Psychiatry* **56**, 557–563 (1999).
2. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).
3. Belmaker, R. H. & Agam, G. Major depressive disorder. *New Engl. J. Med.* **358**, 55–68 (2008).
4. Sullivan, P. F. et al. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol. Psychiatry* **14**, 359 (2009).
5. Wray, N. R. et al. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol. Psychiatry* **17**, 36 (2012).
6. Cross-Disorder, Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
7. Lee, S. H. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984 (2013).
8. McCarthy, M. I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**, 356 (2008).
9. Whalley, H. C. et al. Dissection of major depressive disorder using polygenic risk scores for schizophrenia in two independent cohorts. *Transl. Psychiatry* **6**, e938 (2016).
10. Potkin, S. G. et al. Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. *Cogn. NeuroPsychiatry* **14**, 391–418 (2009).
11. Potkin, S. G. et al. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. *PLoS ONE* **4**, e6501 (2009).
12. Pucilowski, O., Overstreet, D. H., Rezvani, A. H. & Janowsky, D. S. Chronic mild stress-induced anhedonia: greater effect in a genetic rat model of depression. *Physiol. Behav.* **54**, 1215–1220 (1993).
13. Romeas, T., Morissette, M. C., Mnie-Filali, O., Piñeyro, G. & Boye, S. M. Simultaneous anhedonia and exaggerated locomotor activation in an animal model of depression. *Psychopharmacol. (Berl.)* **205**, 293–303 (2009).
14. Fawcett, J., Clark, D. C., Scheftner, W. A. & Gibbons, R. D. Assessing anhedonia in psychiatric patients: The Pleasure Scale. *Arch. Gen. Psychiatry* **40**, 79–84 (1983).
15. American Psychiatric Association, American Psychiatric Association. *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, Text Revision* 75 (American Psychiatric Association, Washington, DC, 2000; 78–85.
16. McMakin, D. L. et al. Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment-resistant depression. *J. Am. Acad. Child & Adolesc. Psychiatry* **51**, 404–411 (2012).
17. Uher, R. et al. Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. *Psychol. Med.* **42**, 967–980 (2012).
18. Liu, W. H. et al. Anhedonia is associated with blunted reward sensitivity in first-degree relatives of patients with major depression. *J. Affect Disord.* **190**, 640–648 (2016).
19. Bogdan, R. & Pizzagalli, D. A. The heritability of hedonic capacity and perceived stress: a twin study evaluation of candidate depressive phenotypes. *Psychol. Med.* **39**, 211–218 (2009).
20. Wing, J. K., Sartorius, N., Üstün, T. B. (eds.) *Diagnosis and Clinical Measurement in Psychiatry: A Reference Manual for SCAN* (Cambridge University Press, Cambridge, 1998).
21. Uher, R. et al. Differential efficacy of escitalopram and nortriptyline on dimensional measures of depression. *Br. J. Psychiatry* **194**, 252–259 (2009).
22. Uher, R. et al. Genetic predictors of response to antidepressants in the GENDEP project. *Pharm. J.* **9**, 225 (2009).

23. Uher, R. et al. Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol. Med.* **38**, 289–300 (2008).

24. Freeman, B. et al. DNA from buccal swabs recruited by mail: evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behav. Genet.* **33**, 67–72 (2003).

25. Uher, R. et al. Genome-wide pharmacogenetics of antidepressant response in the GENDEP project. *Am. J. Psychiatry* **167**, 555–564 (2010).

26. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Human. Genet.* **81**, 559–575 (2007).

27. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

28. Simon-Sanchez, J. et al. Genome-wide association study reveals genetic risk underlying Parkinson's Disease. *Nat. Genet.* **41**, 1308 (2009).

29. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904 (2006).

30. Price, A. L. et al. Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132 (2008).

31. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS. Genet.* **5**, e1000529 (2009).

32. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5 (2013).

33. Dreszer, T. R. et al. The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res.* **40**(D1), D918–D923 (2011).

34. Lippert, C. et al. FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833 (2011).

35. Trivedi, M. H. et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR* D: implications for clinical practice. *Am. J. Psychiatry* **163**, 28–40 (2006).

36. Rush, A. J. et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR* D report. *Am. J. Psychiatry* **163**, 1905–1917 (2006).

37. Shyn, S. I. et al. Novel loci for major depression identified by genome-wide association study of sequenced treatment alternatives to relieve depression and meta-analysis of three studies. *Mol. Psychiatry* **16**, 202 (2011).

38. Fabbri, C. et al. New insights into the pharmacogenomics of antidepressant response from the GENDEP and STAR* D studies: rare variant analysis and high-density imputation. *Pharmacogenomics J.* **18**, 413 (2018).

39. Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519 (2011).

40. Schork, A. J. et al. All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS. Genet.* **9**, e1003449 (2013).

41. Bakshi, A. et al. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Sci. Rep.* **6**, 32894 (2016).

42. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).

43. Kosmicki, J. A. et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49**, 504 (2017).

44. Samocha, K. E. et al. A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944 (2014).

45. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. 068593. Preprint at http://biorxiv.org/content/early/2016/08/09/068593 (2016).

46. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285 (2016).

47. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291 (2015).

48. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).

49. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., RCore, T. E. NLME: Linear and Nonlinear Mixed Effects Models. R package version 3.1-120, URL http://CRAN. R-project.org/package=nlme (2015).

50. Westra, H. J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238 (2013).

51. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

52. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).

53. Rosenberg, N. A. et al. Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).

54. Pritchard, J. K., Stephens, M., Rosenberg, N. A. & Donnelly, P. Association mapping in structured populations. *Am. J. Human. Genet.* **67**, 170–181 (2000).

55. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Human. Genet.* **88**, 76–82 (2011).

56. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459 (2010).

57. Zhang, Z. et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355 (2010).

58. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348 (2010).

59. Cai, N. et al. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588 (2015).

60. Shi, Q. et al. Gene expression profiling in the developing rat brain exposed to ketamine. *Neuroscience* **166**, 852–863 (2010).

61. Klok, M. D. et al. Decreased expression of mineralocorticoid receptor mRNA and its splice variants in postmortem brain regions of patients with major depressive disorder. *J. Psychiatr. Res.* **45**, 871–878 (2011).

62. McCullumsmith, R. E. & Meador-Woodruff, J. H. Striatal excitatory amino acid transporter transcript expression in schizophrenia, bipolar disorder, and major depressive disorder. *Neuropsychopharmacology* **26**, 368 (2002).

63. Gurevich, I. et al. Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims. *Neuron* **34**, 349–356 (2002).

64. Alvarez, J. C. et al. Decreased platelet serotonin transporter sites and increased platelet inositol triphosphate levels in patients with unipolar depression: effects of clomipramine and fluoxetine. *Clin. Pharmacol. Ther.* **66**, 617–624 (1999).

65. Pacheco, M. A. et al. Alterations in phosphoinositide signaling and G-protein levels in depressed suicide brain. *Brain Res.* **723**, 37–45 (1996).

66. Kofman, O. & Belmaker, R. H. Biochemical, behavioral, and clinical studies of the role of inositol in lithium treatment and depression. *Biol. Psychiatry* **34**, 839–852 (1993).

67. Levine, J., Barak, Y. & Gonzalves, M. Szor H. Double-blind, controlled trial of inositol treatment of depression. *Am. J. Psychiatry* **152**, 792 (1995).

68. Abdulkadir, Ö. et al. Temporal expression analysis of angiogenesis-related genes in brain development. *Vascular Cell.* **4**, 16 (2012).

69. Ramasamy, A. et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418 (2014).

70. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483 (2011).

71. Pieper, A. A. et al. The neuronal PAS domain protein 3 transcription factor controls FGF-mediated adult hippocampal neurogenesis in mice. *Proc. Natl Acad. Sci. U. S. A.* **102**, 14052–14057 (2005).

72. Pickard, B. S. et al. Interacting haplotypes at the NPAS3 locus alter risk of schizophrenia and bipolar disorder. *Mol. Psychiatry* **14**, 874 (2009).

73. Lavedan, C. et al. Association of the NPAS3 gene and five other loci with response to the antipsychotic iloperidone identified in a whole genome association study. *Mol. Psychiatry* **14**, 804 (2009).

74. Macintyre, G. et al. Association of NPAS3 exonic variation with schizophrenia. *Schizophr. Res.* **120**, 143–149 (2010).

75. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624 (2016).

76. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2014).

77. Research domain criteria RDoC. https://www.nimh.nih.gov/researchpriorities/rdoc/constructs/rdoc-matrix.shtml. Accessed: 2016.

78. Thierry, G. et al. Molecular characterization of 1q44 microdeletion in 11 patients reveals three candidate genes for intellectual disability and seizures. *Am. J. Med. Genet. A.* **158**, 1633–1640 (2012).

79. Nagamani, S. C. et al. Delineation of a deletion region critical for corpus callosal abnormalities in chromosome 1q43–q44. *Eur. J. Hum. Genet.* **20**, 176 (2012).

80. Keedwell, P. A., Andrew, C., Williams, S. C., Brammer, M. J. & Phillips, M. L. The neural correlates of anhedonia in major depressive disorder. *Biol. Psychiatry* **58**, 843–853 (2005).

81. De Vos, K. J., Grierson, A. J., Ackerley, S. & Miller, C. C. Role of axonal transport in neurodegenerative diseases. *Annu. Rev. Neurosci.* **31**, 151–173 (2008).

82. Houeto, J. L., Magnard, R., Dalley, J. W., Belin, D. & Carnicella, S. trait impulsivity and anhedonia: two Gateways for the development of impulse Control disorders in Parkinson's Disease? *Front. Psychiatry* **7**, 91 (2016).

83. Nagayama, H. et al. Anhedonia and its correlation with clinical aspects in Parkinson's disease. *J. Neurol. Sci.* **372**, 403–407 (2017).

84. Salamone, J. D., Cousins, M. S. & Snyder, B. J. Behavioral functions of nucleus accumbens dopamine: empirical and conceptual problems with the anhedonia hypothesis. *Neurosci. Biobehav. Rev.* **21**, 341–359 (1997).

85. 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).

86. Lally, N. et al. Anti-anhedonic effect of ketamine and its neural correlates in treatment-resistant bipolar depression. *Transl. Psychiatry* **4**, e469 (2014).

87. Gargoloff, P.D. et al. Effectiveness of agomelatine on anhedonia in depressed patients: an outpatient, open-label, real-world study. *Human Psychopharmacol.: Clinical Exp.* **31**, 412–418 (2016).

88. Papp, M. et al. Attenuation of anhedonia by cariprazine in the chronic mild stress model of depression. *Behav. Pharmacol.* **25**(5 and 6), 567–574 (2014).

89. Rizvi, S. J., Sproule, B. A., Gallaugher, L., McIntyre, R. S. & Kennedy, S. H. Correlates of benzodiazepine use in major depressive disorder: the effect of anhedonia. *J. Affect Disord.* **187**, 101–105 (2015).