Analysis of 200,000 exome-sequenced UK Biobank subjects fails to identify genes influencing probability of developing a mood disorder resulting in psychiatric referral

David Curtis

Academic affiliations:

UCL Genetics Institute, University College London

Centre for Psychiatry, Queen Mary University of London

Correspondence:

David Curtis MD PhD, UCL Genetics Institute, University College London, Darwin Building, Gower Street, London WC1E 6BT.
d.curtis@ucl.ac.uk
+ 44 20 7679 2212

Abstract

Background

Depression is moderately heritable but there is no common genetic variant which has a major effect on susceptibility. A previous analysis of 50,000 exome-sequenced subjects failed to implicate any genes or sets of genes in which rare variants were associated with risk of affective disorder requiring specialist treatment. A much larger exome-sequenced dataset is now available.

Methods

Data from 200,632 exome-sequenced UK Biobank participants was analysed. Subjects were treated as cases if they had reported having seen a psychiatrist for "nerves, anxiety, tension or depression". Gene-wise weighted burden analysis was performed to see if there were any genes or sets of genes for which there was an excess of rare, functional variants in cases.

Results

There were 22,886 cases and 176,486 controls. There were 22,642 informative genes but no gene or gene set produced a statistically significant result after correction for multiple testing. None of the genes or gene sets with the lowest p values appeared to be an obvious biological candidate.

Conclusions

The results conform exactly with the expectation under the null hypothesis. It seems unlikely that the use of common, poorly defined phenotypes will produce useful advances in understanding genetic contributions to affective disorder and it might be preferable to focus instead on obtaining large exome-sequenced samples of conditions such as bipolar 1 disorder and severe, recurrent depression.

This research has been conducted using the UK Biobank Resource.

Keywords

Depression; exome; UK Biobank; gene.

Introduction

Genetic investigations of diseases such as schizophrenia and Alzheimer's disease have resulted in the identification of individual genes in which coding variants can have substantial effects on susceptibility and such results provide helpful information which may eventually result in advances in therapy (Curtis and Bandyopadhyay, 2021; Mukai et al., 2019). Genome wide association studies (GWAS) of less severe mental illnesses such as depression do not reveal any individual variant which has a substantial effect on risk. A large genome-wide association meta-analysis of major depressive disorder (MDD) reported 44 independent and significant genetic loci but, although 153 genes were listed in a supplementary table as being significant in gene-wise analyses after controlling for multiple comparisons, the authors did not claim that any individual gene was definitely implicated although they said that the presence of two independent genetic associations strongly suggested *RBFOX1* to be a relevant gene (Wray et al., 2018). A recent review of the genetic contribution to major depressive disorder reports that over 100 genetic loci with small effect sizes have been implicated but states that "the identification of MDD risk variants has been more challenging than for schizophrenia because of its higher prevalence, lower heritability and more heterogeneous samples" (Kendall et al., 2021). The review states that the loci with most support map to *NEGR1* and *OLFM4* and have OR of 1.04. Exome-sequencing studies have reported association of MDD symptoms with rare, nonsynonymous variants in *LIPG*, *RCL1* and *NKPD1* although these findings await replication (Amin et al., 2018, 2017b, 2017a). A study of rare, coding variants in 50,000 exome-sequenced UK Biobank subjects was completely unsuccessful in highlighting any genes influencing the probability of being referred to a specialist for treatment of a mood disorder (Curtis, 2021a). The latter study concluded that it was unlikely that specific genes would be implicated until far larger samples became available.

As explained previously, the UK Biobank is a resource of 500,000 British volunteers for whom variable amounts of phenotypic information is available (http://www.ukbiobank.ac.uk/about-biobank-uk/) (Curtis, 2021a). 94.6% are of white ethnicity and for a standard GWAS it would be necessary to discard data from other participants (Fry et al., 2017). However weighted burden analysis is much less sensitive to population stratification than simple association analyses because instead of relying on detecting differences in allele frequencies between cases and controls it relies on detecting differences in the total variant load within a gene. In a study of BMI we showed that, although population stratification did produce inflation of the test statistic, this was remarkably well controlled for by including the first 20 population principal components as covariates and doing this allowed subjects of all ancestries to be included in the analysis (Curtis, 2020). Subsequently we recognised that there was a bias to call X chromosome variants more frequently in females than males and that this could be adequately controlled for by also including sex as a covariate (Curtis, 2021b).

Although some phenotypic measures are only available for a proportion of participants, one which was completed by almost all of them was an item at the initial assessment, consisting of a touchscreen question asking, "Have you ever seen a psychiatrist for nerves, anxiety, tension or depression?" to which 497,000 answered either Yes or No. This item was chosen as the measure of interest because most subjects had responded and because requiring referral to a specialist might reasonably be taken as a measure of having some significant degree of psychiatric morbidity. Weighted burden analysis of rare variants within genes was carried out on the 50,000 subjects for whom exome sequence data was then available, with variants which were more likely to have functional effects and/or which were rarer receiving higher weights. No gene was identified for which there was a statistically significant association between the weighted burden score and probability of referral for specialist psychiatric treatment. UK Biobank has now increased the number

of subjects for whom exome sequence is available to 200,000 and we report here the results of applying the same analytic approach to this much expanded dataset.

Methods

The same approach was followed as previously fully described (Curtis, 2021a). The UK Biobank dataset was downloaded along with the variant call files for 200,632 subjects who had undergone exome-sequencing and genotyping by the UK Biobank Exome Sequencing Consortium using the GRCh38 assembly with coverage 20X at 95.6% of sites on average (Szustakowski et al., 2020). UK Biobank had obtained ethics approval from the North West Multi-centre Research Ethics Committee which covers the UK (approval number: 11/NW/0382) and had obtained informed consent from all participants. The UK Biobank approved an application for use of the data (ID 51119) and ethics approval for the analyses was obtained from the UCL Research Ethics Committee (11527/001). All variants were annotated using the standard software packages VEP, PolyPhen and SIFT (Adzhubei et al., 2013; Kumar et al., 2009; McLaren et al., 2016). To obtain population principal components reflecting ancestry, version 2.0 of *plink* (https://www.cog-genomics.org/plink/2.0/) was run with the options *--maf 0.1 --pca 20 approx* (Chang et al., 2015; Galinsky et al., 2016).

The phenotype was determined according to how participants had responded in their initial assessment to the touchscreen question: "Have you ever seen a psychiatrist for nerves, anxiety, tension or depression?" Those answering "Yes" were taken to be cases and those answering "No" controls. No attempt was made to screen out controls who might have had some other psychiatric diagnosis.

SCOREASSOC was used to carry out a weighted burden analysis to test whether, in each gene, sequence variants which were rarer and/or predicted to have more severe functional effects occurred more commonly in cases than controls. Attention was restricted to rare variants with minor allele frequency (MAF) <= 0.01 in both cases and controls. As previously described, variants were weighted by MAF so that variants with MAF=0.01 were given a weight of 1 while very rare variants with MAF close to zero were given a weight of 10 (Curtis, 2012). Variants were also weighted according to their functional annotation, with weights ranging from 1 for intergenic variants to 40 for variants predicted to cause complete loss of function in the gene. Additionally, 10 was added to the weight if the PolyPhen annotation was possibly or probably damaging and also if the SIFT annotation was deleterious. The full range of annotations and their weights is shown in Table 1. The weight due to MAF and the weight due to functional annotation were then multiplied together to provide an overall weight for each variant. For each subject a gene-wise weighted burden score was derived as the sum of the variant-wise weights, each multiplied by the number of alleles of the variant which the given subject possessed. For variants on the X chromosome, hemizygous males were treated as homozygotes. Variants were excluded if there were more than 10% of genotypes missing in the cases or controls or if the heterozygote count was smaller than both homozygote counts in the cases or controls. If a subject was not genotyped for a variant then they were assigned the subject-wise average score for that variant.

For each gene, a logistic regression analysis was carried out to test whether the gene-wise variant burden score was associated with the psychiatric phenotype, including the first 20 population principal components and sex as covariates. The statistical significance is summarised as a signed log p value (SLP) which is the log base 10 of the associated p value, given a positive sign if the score is higher in cases and negative if it is higher in controls.

Gene set analyses were carried out using the 1454 "all GO gene sets, gene symbols" pathways as listed in the file *c5.all.v5.0.symbols.gmt* downloaded from the Molecular Signatures Database at

(Subramanian et al., 2005). For each set of genes, the natural logs of the gene-wise p values were summed according to Fisher's method to produce a chi-squared statistic with degrees of freedom equal to twice the number of genes in the set. The p value associated with this chi-squared statistic was expressed as a minus log10 p (MLP) as a test of association of the set with the mental illness phenotype. The weighted burden analysis makes the assumption that rare, functional variants within a gene are expected on average to more likely impair the function of the gene than enhance it. However, within a gene set it might be that impairing the function of some genes increases susceptibility to the phenotype whereas impaired function of other genes in the set might be protective and hence, as described previously, Fisher's method is used so as to be agnostic as to the direction of effect of each gene (Curtis, 2020).

Results

There were 22,886 cases who answered positively to the question about having seen a psychiatrist for "nerves, anxiety, tension or depression" and 176,486 controls. There were 22,642 genes for which there were qualifying variants and the QQ plot for the SLPs obtained for each gene is shown in Figure 1. This shows that the test is well-behaved and that the SLPs conform exactly to what one would expect under the null hypothesis that there are no genes for which an increased burden of rare, functional variants affects the risk of mental illness.

Table 2 shows the results for all genes with an absolute value of SLP >= 3 (equivalent to p<=0.001). By chance, from 22,642 genes one would expect 11 to have SLP greater than 3 and 11 to have SLP less than -3, whereas the actual numbers are 7 and 13. Applying a Bonferroni correction to test for genome-wide statistical significance would yield a threshold of log10(22,028/0.05) = 5.6 for the absolute value of the SLP and no gene achieves this. The genes listed in the table are involved in a wide variety of different functions but none appears to be an obvious biological candidate for affecting susceptibility to depression or anxiety.

There is no gene set which would be statistically significant after a Bonferroni correction for the number of sets tested. The most significant set was TRANSFORMING GROWTH FACTOR BETA RECEPTOR SIGNALING PATHWAY (MLP = 4.07) which contains 36 genes.

The SLPs for all genes and the MLPs for all gene sets are provided in supplementary tables S1 and S2.

Discussion

Even with the considerably larger sample, the results obtained from this study are as completely negative as were the results from the previous analysis of 50,000 exomes. None of the results is formally significant and even the genes which are ranked highest and lowest do not include any which would be regarded as being obvious candidates from a biological point of view. This contrasts with results obtained when applying similar methods to considerably smaller sample sizes of other complex phenotypes, comprising schizophrenia, late onset Alzheimer's disease and BMI (Curtis, 2020; Curtis et al., 2019, 2018). For these phenotypes it was possible to identify gene sets which were statistically significant and a small number of biologically plausible genes which were either genome-wide significant or which at least had uncorrected p values less than 0.001. In the current dataset, a similar analysis of type 2 diabetes successfully identified genes at exome-wide significance levels which were previously known to affect susceptibility, as well as highlighting other genes of interest (Curtis, 2021c). However with the psychiatric phenotype studied here there is no suggestion of such a signal and the results follow a completely random distribution.

As discussed previously, the fact that the results are negative may arise from a number of factors. Obviously, the answer to a single touch screen question results in a poorly defined phenotype. Nevertheless, referral to a specialist might be taken as an indicator of significant psychiatric

morbidity, more so than a simple self-report of anxiety or depression and/or being prescribed medication. One might expect that the bulk of the cases might have a mix of anxiety and depression related diagnoses and studies using common genetic variants suggest a considerable overlap for genetic contributions to these conditions (Purves et al., 2019). Likewise, it is not expected that the question would pick up everybody with psychiatric illness and it is expected that a proportion of controls would also have depression, anxiety and other diagnoses. The overall heritability of depression is lower than for schizophrenia and it is possible that rare genetic variants with major effect sizes make a smaller overall contribution to liability. It may be that some variants do have major effects but that they are spread across so many different genes that the methodology applied here would fail to detect them.

As for the smaller sample, this study fails to identify genes which might influence the risk of mental disorder of sufficient severity to warrant referral for psychiatric treatment. Given the complete lack of signal, there is no expectation that simply increasing the sample size would yield useful results. These results raise serious questions about the utility of carrying out genetic studies using common mental disorder as a phenotype. Instead, progress in understanding the genetics of affective disorders might depend on obtaining sequence data from large samples of more severe and narrowly defined phenotypes, such as bipolar 1 disorder or severe depression requiring repeated hospitalisation.

Conflicts of interest

The author declares he has no conflict of interest.

Data availability

The raw data is available on application to UK Biobank. Detailed results with variant counts cannot be made available because they might be used for subject identification. Software and scripts used to perform the analyses is available at https://github.com/davenomiddlenamecurtis.

References

Adzhubei, I., Jordan, D.M., Sunyaev, S.R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. 7 Unit7.20.

Amin, N., Belonogova, N., Jovanova, O., Brouwer, R., van Rooij, J., van den Hout, M.C.G.N., Svishcheva, G.R., Kraaij, R., Zorkoltseva, I. V., Kirichenko, A. V., Hofman, A., Uitterlinden, A.G., van IJcken, W.F.J., Tiemeier, H., Axenovich, T.I., van Duijn, C.M. (2017a) Nonsynonymous Variation in NKPD1 Increases Depressive Symptoms in European Populations. Biol. Psychiatry 81, 702–707.

Amin, N., De Vrij, F., Baghdadi, M., Brouwer, R., Van Rooij, J.G.J., Jovanova, O., Uitterlinden, A.G., Hofman, A., Janssen, H.L.A., Darwish Murad, S., Kraaij, R., Stedehouder, J., Van Den Hout, M.C.G.N., Kros, J.M., Van Ijcken, W.F.J., Tiemeier, H., Kushner, S.A., Van Duijn, C.M. (2018) A rare missense variant in RCL1 segregates with depression in extended families. Mol. Psychiatry 23, 1120–1126.

Amin, N., Jovanova, O., Adams, H.H., Dehghan, A., Kavousi, M., Vernooij, M.W., Peeters, R.P., De Vrij,

F.M.S., Van Der Lee, S.J., Van Rooij, J.G.J., Van Leeuwen, E.M., Chaker, L., Demirkan, A., Hofman, A., Brouwer, R.W.W., Kraaij, R., Willems Van DIjk, K., Hankemeier, T., Van Ijcken, W.F.J., Uitterlinden, A.G., Niessen, W.J., Franco, O.H., Kushner, S.A., Ikram, M.A., Tiemeier, H., Van Duijn, C.M. (2017b) Exome-sequencing in a large population-based study reveals a rare Asn396Ser variant in the LIPG gene associated with depressive symptoms. Mol. Psychiatry 22, 537–543.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7.

Curtis, D. (2012) A rapid method for combined analysis of common and rare variants at the level of a region, gene, or pathway. Adv Appl Bioinform Chem 5, 1–9.

Curtis, D. (2020) Multiple Linear Regression Allows Weighted Burden Analysis of Rare Coding Variants in an Ethnically Heterogeneous Population. Hum. Hered. 1–10.

Curtis, D. (2021a) Analysis of 50,000 exome-sequenced UK Biobank subjects fails to identify genes influencing probability of developing a mood disorder resulting in psychiatric referral. J. Affect. Disord. 281, 216–219.

Curtis, D. (2021b) Analysis of 200 000 exome-sequenced UK Biobank subjects illustrates the contribution of rare genetic variants to hyperlipidaemia. J. Med. Genet. jmedgenet-2021-107752.

Curtis, D. (2021c) Weighted burden analysis in 200,000 exome-sequenced subjects characterises rare variant effects on risk of type 2 diabetes. medRxiv 2021.01.08.21249453.

Curtis, D., Bakaya, K., Sharma, L., Bandyopadhay, S. (2019) Weighted burden analysis of exome-sequenced late onset Alzheimer's cases and controls provides further evidence for involvement of PSEN1 and demonstrates protective role for variants in tyrosine phosphatase genes. Ann Hum Genet 84, 291–302.

Curtis, D., Bandyopadhyay, S. (2021) Mini-review: Role of the PI3K/Akt pathway and tyrosine phosphatases in Alzheimer's disease susceptibility. Ann. Hum. Genet. 85, 1–6.

Curtis, D., Coelewij, L., Liu, S.-H., Humphrey, J., Mott, R. (2018) Weighted Burden Analysis of Exome-Sequenced Case-Control Sample Implicates Synaptic Genes in Schizophrenia Aetiology. Behav. Genet. 43, 198–208.

Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., Allen, N.E. (2017) Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am. J. Epidemiol. 186, 1026–1034.

Galinsky, K.J., Bhatia, G., Loh, P.R., Georgiev, S., Mukherjee, S., Patterson, N.J., Price, A.L. (2016) Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. Am. J. Hum. Genet. 98, 456–472.

Kendall, K.M., Van Assche, E., Andlauer, T.F.M., Choi, K.W., Luykx, J.J., Schulte, E.C., Lu, Y. (2021) The genetic basis of major depression. Psychol. Med.

Kumar, P., Henikoff, S., Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F. (2016) The Ensembl Variant Effect Predictor. Genome Biol. 17, 122.

Mukai, J., Cannavo, E., Sun, Z., Crabtree, G., Diamantopoulou, A., Thakur, P., Chang, C.-Y., Cai, Y., Lomvardas, S., Takata, A., Xu, B., Gogos, J. (2019) Recapitulation and reversal of schizophrenia-related phenotypes in Setd1a-deficient mice. bioRxiv 529701.

Purves, K.L., Coleman, J.R.I., Meier, S.M., Rayner, C., Davis, K.A.S., Cheesman, R., Bækvad-Hansen, M., Børglum, A.D., Wan Cho, S., Jürgen Deckert, J., Gaspar, H.A., Bybjerg-Grauholm, J., Hettema, J.M., Hotopf, M., Hougaard, D., Hübel, C., Kan, C., McIntosh, A.M., Mors, O., Bo Mortensen, P., Nordentoft, M., Werge, T., Nicodemus, K.K., Mattheisen, M., Breen, G., Eley, T.C. (2019) A major role

for common genetic variation in anxiety disorders. Mol. Psychiatry 25.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550.

Szustakowski, J.D., Balasubramanian, S., Sasson, A., Khalid, S., Bronson, P.G., Kvikstad, E., Wong, E., Liu, D., Davis, J.W., Haefliger, C., Loomis, A.K., Mikkilineni, R., Noh, H.J., Wadhawan, S., Bai, X., Hawes, A., Krasheninina, O., Ulloa, R., Lopez, A., Smith, E.N., Waring, J., Whelan, C.D., Tsai, E.A., Overton, J., Salerno, W., Jacob, H., Szalma, S., Runz, H., Hinkle, G., Nioi, P., Petrovski, S., Miller, M.R., Baras, A., Mitnaul, L., Reid, J.G. (2020) Advancing Human Genetics Research and Drug Discovery through Exome Sequencing of the UK Biobank. medRxiv 2020.11.02.20222232.

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., Bækvad-Hansen, M., Beekman, A.F.T., Bigdeli, T.B., Binder, E.B., Blackwood, D.R.H., Bryois, J., Buttenschøn, 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-Rodríguez, 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, Yihan, Li, Yun, 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, Di., 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., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J., Painter, J.N., Pedersen, C.B., Pedersen, M.G., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pistis, G., Posthuma, D., Purcell, S.M., Quiroz, J.A., Qvist, P., Rice, J.P., Riley, B.P., Rivera, M., Saeed Mirza, S., Saxena, R., Schoevers, R., Schulte, E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.B.C., Smit, J.H., Smith, D.J., Stefansson, H., Steinberg, S., Stockmeier, C.A., Streit, F., Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Tian, C., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Van Der Auwera, S., Van Hemert, A.M., Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F., Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., De Geus, E.C.J., Depaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J., Hamilton, S.P., Hayward, C., Heath, A.C., Hinds, D.A., Kendler, K.S., Kloiber, S., Lewis, G., Li, Q.S., Lucae, S., Madden, P.F.A., Magnusson, P.K., Martin, N.G., McIntosh, A.M., Metspalu, A., Mors, O., Mortensen, P.B., Müller-Myhsok, B., Nordentoft, M., Nöthen, M.M., O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G., Smoller, J.W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M.M., Werge, T., Winslow, A.R., Lewis, C.M., Levinson, D.F., Breen, G., Børglum, A.D., Sullivan, P.F. (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–681.

Table 1 The table shows the weight which was assigned to each type of variant as annotated by VEP, Polyphen and SIFT (Adzhubei et al., 2013; Kumar et al., 2009; McLaren et al., 2016).

VEP / SIFT / Polyphen annotation	Weight
intergenic_variant	1
feature_truncation	3
regulatory_region_variant	3
feature_elongation	3
regulatory_region_amplification	3
regulatory_region_ablation	3
TF_binding_site_variant	3
TFBS_amplification	3
TFBS_ablation	3
downstream_gene_variant	3
upstream_gene_variant	3
non_coding_transcript_variant	3
NMD_transcript_variant	3
intron_variant	3
non_coding_transcript_exon_variant	3
3_prime_UTR_variant	10
5_prime_UTR_variant	5
mature_miRNA_variant	5
coding_sequence_variant	5
synonymous_variant	5
stop_retained_variant	5
incomplete_terminal_codon_variant	5
splice_region_variant	5
protein_altering_variant	10
missense_variant	10
inframe_deletion	15
inframe_insertion	15
transcript_amplification	15
start_lost	30
stop_lost	30
frameshift_variant	40
stop_gained	40
splice_donor_variant	40
splice_acceptor_variant	40
transcript_ablation	20
SIFT deleterious	10
PolyPhen possibly damaging	10
PolyPhen probably damaging	10

Table 2 Genes with absolute value of SLP exceeding 3 or more (equivalent to p<0.001) for test of association of weighted burden score with referral for psychiatric treatment.

Gene symbol	SLP	Gene name
CLDND1	3.95	Claudin Domain Containing 1
DUSP16	3.48	Dual Specificity Phosphatase 16
LOC101928847	3.43	Uncharacterized LOC101928847
EFTUD2	3.31	Elongation Factor Tu GTP Binding Domain Containing 2
LOC105377103	3.19	Uncharacterized LOC105377103
EED	3.17	Embryonic Ectoderm Development
HIST2H3D	3.07	H3 Clustered Histone 13
JADE3	-3.05	Jade Family PHD Finger 3
APBA2	-3.06	Amyloid Beta Precursor Protein Binding Family A Member 2
LOC105370163	-3.27	Uncharacterized LOC105370163
CCDC160	-3.28	Coiled-Coil Domain Containing 160
ADHFE1	-3.31	Alcohol Dehydrogenase Iron Containing 1
MIR6720	-3.37	MicroRNA 6720
IFRD2	-3.62	Interferon Related Developmental Regulator 2
MAGEA8-AS1	-3.65	MAGEA8 Antisense RNA 1
LOC112267907	-3.75	Uncharacterized LOC112267907
COX18	-4.08	Cytochrome C Oxidase Assembly Factor COX18
MIR4695	-4.83	MicroRNA 4695

Figure 1

QQ plot of SLPs obtained for weighted burden analysis of 22,642 genes for association with referral for psychiatric treatment showing observed against expected SLP for each gene. Also shown is the SLP=eSLP line, marking the distribution of SLPs expected under the null hypothesis that no genes are associated.