Supplemental Online Content

Wainberg M, Merico D, Huguet G, et al. Deletion of loss-of-function–intolerant genes and risk of 5 psychiatric disorders. *JAMA Psychiatry*. Published online December 1, 2021. doi:10.1001/jamapsychiatry.2021.3211

This supplemental material has been provided by the authors to give readers additional information about their work.

© 2021 American Medical Association. All rights reserved.

eMethods

Cohort

The UK Biobank[1] is a population-based prospective cohort of 502,617 individuals drawn from across the UK. Of 451,992 genotyped self-reported white participants (as defined by the "Ethnic Background" field, https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21000), 973 were excluded due to evidence of neuro- developmental disorders (ICD-10 codes F70-F89) from linked inpatient, primary care or death records. (This exclusion criterion does not rule out the possibility of undiagnosed neurodevelopmental disorders.) An additional 19,873 participants either lacked CNV calls or failed previously described CNV-related quality control metrics[2] (number of CNVs >30, waviness factor >0.03 in magnitude, call rate <96%, or standard deviation of log R ratio >0.35). Thus, the final cohort comprised 431,146 self-reported white participants (234,544 female, 196,602 male).

Phenotype definitions

Cases for five psychiatric disorders were defined based on the presence of particular ICD-10 codes in participants' linked inpatient, primary care or death records: anxiety (F40 or F41), bipolar disorder (F31), MDD (F32 or F33), OCD (F42), and schizophrenia (F20). To guard against control group contamination, we defined controls for each disorder based on an elaborate set of exclusion criteria:

1. Participants self-reporting being diagnosed with the disorder, via the UK Biobank's "Source of report" Data-Fields, e.g. "Source of report of F31 (bipolar affective disorder)".
2. Participants self-reporting seeking help for the disorder from a medical professional. For MDD, this comprised patients answering yes to "Ever been offered/sought treatment for depression" or "Professional informed about depression". For anxiety, this comprised patients answering yes to "Ever been offered/sought treatment for anxiety" or "Professional informed about anxiety".
3. Participants with inpatient, primary care, death record-based, or self-reported diagnoses of ICD-10 codes for related disorders, according to the above-mentioned "Source of report" Data-Fields. For MDD, we excluded participants with F30-F31 or F34-F39 codes; for bipolar disorder and schizophrenia, we excluded those with F21-F30 or F34-F39 codes.
4. Participants self-reporting being diagnosed with related disorders, via the UK Biobank's "Mental health problems ever diagnosed by a professional" or "Non-cancer illness code, self-reported" Data-Fields. For MDD, we excluded participants self-reporting "mania, hypomania, bipolar, or manic-depression" or "depression" for the former, or "mania/bipolar disorder/manic depression" or "depression" for the latter. For anxiety, we excluded participants self-reporting "Social anxiety or social phobia", "Any other phobia (eg disabling fear of heights or spiders)", "Panic attacks", "Anxiety, nerves or generalized anxiety disorder" or "Agoraphobia" for the former, or "anxiety/panic attacks" or "nervous breakdown" for the latter. For OCD, we excluded participants self-reporting "Obsessive compulsive disorder (OCD)" for the former, or "obsessive compulsive disorder (ocd)" for the latter. For bipolar disorder and schizophrenia, we excluded participants self-reporting "Schizophrenia", "Any other type of psychosis or psychotic illness" or "Mania, hypomania, bipolar or manic-depression" for the former, or "schizophrenia" or "mania/bipolar disorder/manic depression" for the latter.
5. Participants who screened positive for clinical or subclinical symptoms of the disorder, according to self-reported screening tools. For MDD, we excluded participants with a PHQ-9 score ≥10 according to an online Mental Health Questionnaire, or PHQ-2 scores ≥3 at any in-person assessment (each participant had 1 to 4 in-person assessments, and the PHQ-2 was conducted at each one). For anxiety, we excluded participants with GAD-7 scores ≥10 according to the online Mental Health Questionnaire, or feelings of tenseness/restlessness or nervousness/anxiety on most or all days during the two weeks prior to any in-person assessment. For bipolar disorder and schizophrenia, we excluded participants self-reporting "Ever manic/hyper for 2 days" at any in-person assessment visit, or "Ever had period of mania / excitability", "Ever

© 2021 American Medical Association. All rights reserved.

heard an un-real voice", "Ever seen an un-real vision", "Ever believed in an un-real conspiracy against self" or "Ever believed in un-real communications or signs" on the online Mental Health Questionnaire.

6. Participants meeting expert-defined criteria[3] for "Probable Recurrent major depression (severe)", "Probable Recurrent major depression (moderate)" or "Single Probable major depression episode" (for MDD) or for "Bipolar I Disorder" or "Bipolar II Disorder" (for bipolar disorder and schizophrenia), according to the "Bipolar and major depression status" Data-Field.

Note that these exclusions were only applied to the controls for each disorder; participants simultaneously meeting the case criteria and exclusion criteria were still deemed to be cases.

CNV calls

To guard against false positives, we analyzed autosomal CNV calls from two different studies[2,4], obtained from two different Hidden Markov Model-based CNV callers (PennCNV[5] and QuantiSNP[6]) and two different quality control pipelines, and required them to agree that a gene was deleted in its entirety. For each set of CNV calls, we used the same CNV-level quality control metrics as the study that generated the calls. Specifically, for the PennCNV calls[2], CNVs were excluded if covered by <10 microarray probes or <1 probe per 20 kilobases, while for the QuantiSNP calls[4], CNVs were excluded if less than 1 kilobase in length, covered by <3 microarray probes (with no probe density filter), or had likelihood score < 15. Both sets of calls were obtained from the UK Biobank's returns catalog (biobank.ctsu.ox.ac.uk/crystal/dset.cgi?id=1701 for PennCNV; biobank.ctsu.ox.ac.uk/crystal/dset.cgi?id=3104 for QuantiSNP). For additional robustness, we additionally applied a minimum length threshold of 10 kb and maximum length threshold of 10 MB to both sets of CNV calls, and excluded CNV calls where more than 50% of the CNV overlapped segmental duplications from genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=genomicSuperDups.

Association testing

Association testing between the eight gene categories and five psychiatric disorders was performed using version 2.0.2 of REGENIE[7], a genome-wide association study toolkit that accounts for sample relatedness and population structure by computing a polygenic risk score (PRS) for the phenotype being associated (Step 1), then including this PRS as a covariate when performing the actual association testing (Step 2).

While by default REGENIE uses a leave-one-chromosome-out scheme to construct the PRS (e.g. when performing association tests on chr17, a PRS derived from all chromosomes except chr17 is used as a covariate), this was not suitable for our application since our eight gene categories included genes spread across multiple chromosomes. Thus, we used all chromosomes to construct the PRS, via REGENIE's "--print-prs" and "--use-prs" options.

Step 1 of REGENIE, the PRS construction, was run on a quality controlled subset of the UK Biobank's microarray data, namely single-nucleotide variants with minor allele frequency >1%, <5% missingness and Hardy-Weinberg equilibrium $p > 1 \times 10^{-15}$. For computational efficiency, REGENIE constructs the PRS using stacked ridge regression, by partitioning the genome into non-overlapping blocks of N markers (where N is specified using the "--bsize" option), running ridge regression within each block, and then running a second level of ridge regression to aggregate across blocks. Here we set "--bsize" to 1000, as recommended for REGENIE analyses of the UK Biobank (rgcgithub.github.io/regenie/recommendations).

Step 2 of REGENIE, the association testing, was performed using logistic regression, with REGENIE's fast approximate Firth regression as a fallback when the logistic regression p-value was less than 0.01, using the options "--firth --approx --firth-se --pThresh 0.01".

Both steps of REGENIE were corrected for age, sex, genotyping array (Axiom versus BiLEVE), and the top 10 genotype principal components (according to the "Genetic principal components" Data-Field). In step 2, we also corrected for the total number of genes called as deleted by CNVs in each person.

Neurodevelopmental disorder genes

Genes were characterized as neurodevelopmental disorder genes if they were among the Simons Foundation Autism

© 2021 American Medical Association. All rights reserved.

Research Initiative autism database's 194 high-confidence (category 1) autism risk genes (https://gene.sfari.org/database/human-gene), 102 candidate autism genes from Satterstrom *et al.* 2020[8], or 253 candidate neurodevelopmental disease genes from Coe *et al.* 2019[9].

eTable. The 32 Recurrent Deletion CNV Regions From Kendall et al. 2019.

Syndrome	Locus	hg19 coordinates	Overlapping genes with pLI > 0.99 [or highest-pLI gene, in square brackets]
1p36 deletion	1p36.32-p36.33	chr1:0-2,500,000	*GNB1* (0.999), *SKI* (0.999), *GABRD* (0.993)
Thrombocytopenia-absent radius (TAR)	1q21.1	chr1:145,394,955-145,807,817	[*PIAS3* (0.985)]
1q21.1 microdeletion	1q21.1-q21.2	chr1:146,527,987-147,394,444	[*PRKAB2* (0.726)]
2p16.3 deletion	2p16.3	chr2:50,145,643-51,259,674	*NRXN1* (1.0)
2q11.2 deletion syndrome	2q11.2	chr2:96,742,409-97,677,516	*SEMA4C* (1.0), *SNRNP200* (1.0), *KANSL3* (1.0)
2q13 deletion	2q13	chr2:111,394,040-112,012,649	[*BCL2L11* (0.889)]
2q37 deletion	2q37.3	chr2:239,716,679-243,199,373	*HDLBP* (1.0), *HDAC4* (1.0), *KIF1A* (1.0), *PPP1R7* (0.995)
3q29 microdeletion	3q29	chr3:195,720,167-197,354,826	*UBXN7* (1.0), *SENP5* (1.0), *DLG1* (0.994)
Wolf-Hirschhorn	4p16.3	chr4:1,552,030-2,091,303	*WHSC1* (1.0)
Sotos	5q35.2-q35.3	chr5:175,720,924-177,052,594	*NSD1* (1.0), *UNC5A* (1.0), *FAF2* (0.999), *DBN1* (0.997), *FAM193B* (0.993)
Williams	7q11.23	chr7:72,744,915-74,142,892	*BAZ1B* (1.0), *CLIP2* (1.0), *LIMK1* (0.999), *GTF2I* (0.996)
8p23.1 microdeletion	8p23.1	chr8:8,098,990-11,872,558	*XKR6* (0.999)
10q22.3q23.2 microdeletion	10q22.3-q23.2	chr10:82,045,472-8,931,651	*WAPAL* (1.0), *GRID1* (0.999)
Potocki-Shaffer	11p11.2	chr11:43,940,000-46,020,000	*PHF21A* (1.0), *MAPK8IP1* (1.0)
15q11.2 microdeletion	15q11.2	chr15:22,805,313-23,094,530	[*CYFIP1* (0.968)]
Prader-Willi, Angelman	15q11-q13	chr15:22,805,313-28,390,339	*HERC2* (1.0), *UBE3A* (1.0)
15q13.3 microdeletion	15q13.3	chr15:31,080,645-32,462,776	[*OTUD7A* (0.952)]
15q24 microdeletion	15q24	chr15:72,900,171-78,151,253	*SIN3A* (1.0), *CSK* (1.0), *HCN4* (0.999), *LINGO1* (0.994), *NPTN* (0.993)
15q25.2 deletion	15q25.2	chr15:83,219,735-85,722,039	*ZNF592* (1.0), *CPEB1* (0.999), *RP11-152F13.10* (0.996)
16p13.11 microdeletion	16p13.11	chr16:15,511,655-16,293,689	*MARF1* (1.0)

© 2021 American Medical Association. All rights reserved.

16p12.2 microdeletion	16p12.2	chr16:21,950,135-22,431,889	[*MOSMO* (0.259)]
Distal 16p11.2 deletion	16p11.2	chr16:28,823,196-29,046,783	*ATXN2L* (1.0)
16p11.2 deletion	16p11.2	chr16:29,650,840-30,200,773	*TAOK2* (1.0)
Miller-Dieker	17p13.3	chr17:1,247,834-1,303,556	[*YWHAE* (0.985)]
Miller-Dieker	17p13.3	chr17:2,496,923-2,588,909	*PAFAH1B1* (1.0)
Smith-Magenis	17p11.2	chr17:16,812,771-20,211,017	*RAI1* (1.0), *COPS3* (0.999)
17q11.2 deletion	17q11.2	chr17:29,107,491-30,265,075	*ATAD5* (1.0), *SUZ12* (1.0), *RAB11FIP4* (0.992)
Renal cysts and diabetes	17q12	chr17:34,815,904-36,217,432	*GGNBP2* (1.0), *ACACA* (1.0), *HNF1B* (1.0)
Koolen de Vries	17q21.31	chr17:43,705,356-44,164,691	*KANSL1* (1.0)
22q11.2 deletion	22q11.2	chr22:19,037,332-21,466,726	*HIRA* (1.0), *SCARF2* (1.0), *MED15* (1.0), *DGCR8* (1.0), *UFD1L* (0.996)
22q11.2 distal deletion	22q11.2	chr22:21,920,127-23,653,646	*MAPK1* (0.997)
Phelan-Mcdermid	22q13.33	chr22:51,113,070-51,171,640	*SHANK3* (1.0)

eReferences

1. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.

2. Crawford K, Bracher-Smith M, Owen D, et al. Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *J Med Genet*. 2019;56(3):131-138.

3. Smith DJ, Nicholl BI, Cullen B, 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.

4. Moreau C, Huguet G, Urchs S, et al. The general impact of haploinsufficiency on brain connectivity underlies the pleiotropic effect of neuropsychiatric CNVs. *medRxiv*. Published online March 23, 2020.

5. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007;17(11):1665-1674.

6. Colella S, Yau C, Taylor JM, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res*. 2007;35(6):2013-2025.

7. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*.

8. Satterstrom FK, Kosmicki JA, Wang J, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell. 2020;180(3):568-584.e23.

9. Coe BP, Stessman HAF, Sulovari A, et al. Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. Nat Genet. 2019;51(1):106-116.

© 2021 American Medical Association. All rights reserved.