Genome-wide association study identifies new locus associated with OCD [preprint]

Nora I. Strom
Humboldt University of Berlin
Et al.

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/faculty_pubs

Part of the Genetics and Genomics Commons, Medical Genetics Commons, Mental and Social Health Commons, Neuroscience and Neurobiology Commons, Psychiatry Commons, and the Psychiatry and Psychology Commons

Repository Citation

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License. This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in University of Massachusetts Medical School Faculty Publications by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
Genome-wide association study identifies new locus associated with OCD


1-181 Affiliations appear at the end of the document.

*Authors contributed equally.

Obsessive-compulsive disorder (OCD) is a heritable disorder, but no definitive, replicated OCD susceptibility loci have yet been identified by any genome-wide association study (GWAS). Here, we report results from a GWAS in the largest OCD case-control sample (**N = 14,140** OCD cases and **N = 562,117** controls) to date. We explored the genetic architecture of OCD, including its genetic relationships to other psychiatric and non-psychiatric phenotypes. In the GWAS analysis, we identified one SNP associated with OCD at a genome-wide significant level. Subsequent gene-based analyses identified additional two genes as potentially implicated in OCD pathogenesis. All SNPs combined explained 16% of the heritability of OCD. We show substantial positive genetic correlations between OCD and a range of psychiatric disorders, including anxiety disorders, anorexia nervosa, and major depression. We thus for the first time provide evidence of a genome-wide locus implicated in OCD and strengthen previous literature suggesting a polygenic nature of this disorder.

**Obsessive-compulsive disorder | OCD | GWAS**

Correspondence: Manuel Mattheisen, Department of Psychiatry, Dalhousie University, Halifax, Canada, manuel.mattheisen@dal.ca

**Introduction**

Obsessive-compulsive disorder (OCD) is a chronic mental illness that affects approximately 2-3% of the general population. It is characterized by obsessions and compulsions that vary in type and severity across patients, with most showing a waxing and waning illness course. Age at onset is usually in early to late adolescence but diagnosis and initiation of treatment are often delayed by several years. OCD frequently co-occurs with other disorders, particularly tic disorders, eating disorders, depressive disorders, and anxi-
ity disorders. Although anxiety is a common symptom of OCD, OCD is currently considered to be distinct from anxiety disorders in terms of its epidemiological, clinical and pathophysiological presentation. This uniqueness is recognized in DSM-5 and ICD11, which separate OCD, along with several related disorders, into a separate disorder category, Obsessive Compulsive and Related Disorders (OCRDs)².

Although its pathogenesis is yet to be fully elucidated, it is now clear that complex genetic factors play a role in the susceptibility to and/or development of OCD. The heritability of obsessive-compulsive (OC) symptoms as estimated by twin studies is between 27 and 47% in adults and between 45 and 65% in children⁹⁻¹¹. SNP-based heritability of OCD, estimated using genome complex trait analysis (GCTA), is between 28-37%, with higher estimates for childhood-onset OCD, in line with heritability estimates from twin studies.¹²⁻¹³. GCTA analyses further indicate that much of the genetic risk for OCD arises from common variation in multiple loci, each with a small effect, acting in concert. Although rare variants exerting larger effects may also play a role in OCD development, less work has been done in this area.

Two genome-wide association studies (GWAS) of OCD as well as a combined meta-analysis of 2,688 cases and 7,037 controls have been published to date. However, all three studies were significantly under-powered based on estimates that suggest a need for tens to hundreds of thousands of individuals (cases and controls) for definitive gene identification in psychiatric disorders with complex inheritance. Accordingly, no definitive, replicated OCD susceptibility loci have yet been identified by any study. Nevertheless, while small, these prior GWASs did demonstrate significant SNP-based heritability, indicating that a well-powered OCD GWAS with a much larger sample size would likely identify genome-wide significant loci and more accurate SNP effects, similar to what has been demonstrated for other major psychiatric disorders. Such findings will provide insight into the biological underpinnings of OCD and may improve follow-up analyses and contribute to (multimodal) models for risk prediction. In addition, the frequent co-occurrence seen between OCD and other psychiatric illnesses appears to be due, at least in part, to shared genetic susceptibility. Cross-disorder studies demonstrate that OCD is genetically correlated with a range of other psychiatric disorders, most notably Tourette syndrome (TS), anorexia nervosa (AN), and major depressive disorder (MDD), all of which are highly comorbid with OCD.

Here, we performed a GWAS in a substantially expanded case-control sample (N = 14,140 OCD cases and N = 562,117 controls). We explored the genetic architecture of OCD, including its genetic relationships to other psychiatric and non-psychiatric phenotypes, such as TS, AN, and MDD. In the GWAS analysis, we identified one SNP associated with OCD at a genome-wide significant level. Subsequent gene-based analyses nominated two additional genes as potentially implicated in OCD pathogenesis. All SNPs combined explained 16% of the heritability of OCD. We also show substantial positive genetic correlations between OCD and a range of psychiatric disorders, including those with anxiety disorders and MDD.

**Methods**

**Samples**

We analyzed genomic data from 17 OCD case-control cohorts (11,312 OCD cases and 557,230 controls in total) not included in any previous OCD GWAS publications. In addition, three previously published GWAS datasets (two from the International OCD Foundation-Genetics Consortium (IOCDF-GC) and one from The OCD Collaborative Genetic Association Study (OCGAS)) were re-analyzed using newly matched control samples that were genotyped with the same microarrays as the cases (2,828 cases and 4,887 controls). In total, 20 cohorts comprising 14,140 OCD cases and 562,117 controls of European ancestry were included in the analyses. Among these, 323 cases were part of a parent-proband trio; in these cases, parents were used as pseudo controls. A total of 12,607 cases met DSM-IV or ICD10 criteria for OCD, while the remaining 1,533 cases were based on self-reported OCD diagnosis. Supplementary Table S1 provides an overview of the individual cohorts, cohort-specific sample and analytic details can be found in the Supplementary data. Data collections were approved by the relevant institutional review boards at all participating sites, and all participants provided written informed consent.

**Genetic data formatting, cleaning, alignment, and individual GWASs**

First, the data of each participating cohort were analyzed individually (see Supplementary Methods for details). Genetic data were imputed using either the Haplotype Reference Consortium (HRC) or 1000 Genomes Project Phase 3 reference panels. The resulting GWAS summary statistics were then harmonized before a conjoint meta-analysis was conducted. Each summary statistic data set was transformed to ‘damer’ file format following ricopi using the following criteria for inclusion: minor allele frequency (MAF) > 1% in cases and controls, imputation-quality (INFO) score > 0.8 and < 1.2. If the effect measure, p-value or standard error (SE) was missing or was out of bounds (infinite), the SNP was removed. Only biallelic SNPs were retained in the data. Once cleaned summary statistics were produced, all datasets were aligned to the HRC reference panel. If variants were reported on different strands, they were flipped to the orientation in the HRC-reference panel. Furthermore, single-allelic T/C SNPs were removed if their MAF was > 0.4. In case A/T and C/G SNPs showed a MAF < 0.4 allele frequencies were compared to frequencies in the HRC-reference. If an allele frequency
match was found, i.e., the respective allele was also the minor allele in the HRC reference, same strand orientation was assumed. If an allele mismatch was found, i.e. the allele had a frequency > 0.5 in HRC, it was assumed that alleles were reported on different strands and alleles were flipped subsequently. Marker-names were uniformly switched to those present in the HRC reference. If a variant did not overlap with the variants in the HRC reference, it was removed.

**Genome-wide association meta-analysis**

In reverse variance weighted meta-analysis was conducted on 20 European cohorts using METAL. Heterogeneity was assessed with Cochran’s I² statistic. The genomic control factor (λ₁₀₀₀) was calculated for each individual GWAS and for the overall meta-analysis to identify residual population stratification or systematic technical artifact. GWAS summary statistics were subjected to linkage disequilibrium (LD) score regression (LDSC) analyses on high-quality common SNPs (INFO score > 0.9) to examine the LDSC intercept as an alternative evaluation of test statistic inflation due to residual artifact or population stratification, and to estimate the genetic heritability from the meta-analysis and genetic correlations between cohorts. The genome-wide significance threshold for the GWAS was set at a p-value of 5.0×10⁻⁸. Sign tests on the top SNPs (inclusion threshold of p = 0.0001, p = 0.00001, p = 0.000001, and p = 5.000×10⁻⁸) were performed between each individual cohort as well as 'leave-one-out' meta-analyses to identify any cohort in which the summary statistics significantly deviated from the rest of the cohorts.

**Polygenic risk scoring**

OCD polygenic risk scores (PRS) were calculated to assess the predictive value of the new meta-analysis. For this purpose, we conducted three 'leave-one-out' GWAS meta-analyses that each excluded one of three samples (IOCDF-GC, OCGAS, and Psych_Broad) and tested how well the PRS predicted case-control status in the excluded sample. We selected these three datasets based on their previous involvement in a recent meta-analysis (IOCDF-GC and OCGAS; see above) and their availability at the site that conducted the analysis (Psych_Broad). PRS were generated at eight p-value thresholds (0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 1) as a weighted sum of the risk allele dosages. The contribution of PRS was measured by comparing Nagelkerke’s pseudo-R² of the full model (including principal components (PCs), sex, and PRS) and of the reduced model (adjusted for PCs and sex).

**Gene-based analyses**

**Conventional MAGMA (C-MAGMA)** We used conventional MAGMA (C-MAGMA v1.07) to test genetic associations at the gene level for the combined effect of SNPs in or near protein-coding genes while adjusting for LD within the SNPs and the size of the gene. We used the 1000 Genomes reference panel (Phase 3) to control for LD. SNPs were mapped to genes if they were located within 10 kb of the gene coding region. The significance threshold was set using Bonferroni correction (i.e. 0.05/20,031 tests performed: p < 2.50×10⁻⁶).

**Expression informed analysis (E-MAGMA, S-PrediXcan, S-MultiXcan)** We conducted gene-based tests informed by expression quantitative trait loci (eQTL) to identify genes that are enriched for association signal. We used a novel eQTL-informed gene-based test (E-MAGMA) that assigns SNPs to genes based on significant associations (FDR-corrected p < 0.05) with tissue-specific eQTLs in 48 GTEx tissues (v8). E-MAGMA is conducted in MAGMA (v1.07) by modifying gene annotation files by integration of eQTL information and uses the 1000 Genomes reference panel (Phase 3) to model LD. The significance threshold was set using Bonferroni correction across all tissues at p < 2.51×10⁻⁷ (i.e. 0.05/199,421 tests performed).

We used S-PrediXcan to integrate expression quantitative trait loci (eQTL) information with our GWAS summary statistics and identify genes whose genetically predicted expression levels were associated with OCD. S-PrediXcan estimates gene expression weights by training a linear prediction model in samples with both gene expression and SNP genotype data. The weights are then used to predict gene expression from GWAS summary statistics, while incorporating the variance and covariance of SNPs from an LD reference panel. We used expression weights for 48 tissues from the GTEx Project (v8) and LD information from the 1000 Genomes Project (Phase 3) (see web resources). These data were integrated with β values and standard errors from the OCD GWAS meta-analysis to test the association between imputed levels of gene expression and OCD risk. The significance threshold was set at p < 1.67×10⁻⁷ using Bonferroni correction (i.e. 0.05/299,996). Next, we used S-MultiXcan, which integrates the Predixcan associations across the 48 GTEx tissues into a single test, thus maximizing statistical power. We applied a transcriptome-wide significance threshold of p < 2.31×10⁻⁶ (i.e. 0.05/21,601 tests).

**Chromatin informed (H-MAGMA)** We further conducted gene-based tests that integrated genetic results with Hi-C information using H-MAGMA. SNPs were assigned to genes by leveraging regulatory (chromatin interaction) relationships in human brain tissue. H-MAGMA is conducted in MAGMA (v1.07) using the 1000 Genomes reference panel (Phase 3) to model LD. The significance threshold was set using Bonferroni correction (i.e. 0.05/24,358: p < 2.05×10⁻⁶).

**Functional annotation of GWAS findings** We also performed an extensive functional annotation for all SNPs in genomic areas identified by our lead SNPs using FUMA (see web-resources) to identify the most likely causal variants. In addition, the list of genes significantly associated with OCD, as indicated by at least one of the four gene-based methods (Table 2; except LOC101928274 as it is neither contained in the gene catalog of FUMA nor in the ex-
pression data) was used to create expression heatmaps to visualize the average expression value of the associated genes. Further, a differential gene expression analysis (DEG) was performed to test the expression of associated genes against all other genes for each expression data set (tissue and developmental stage) using GTEx v8 and BrainSpan data.

**Tissue and cell-type enrichment analysis**

To determine if there are specific tissues or cell-types whose gene expression profiles are enriched for OCD risk variation, we used an analysis protocol recently described in Bryois et al. Consistent with this we utilized their code-base for this analysis, and in particular the sets of genes that mark different highlighted tissue and cell-type datasets (see web-resources for data source).

We selected 3 datasets that had been preprocessed by Bryois et al. for inclusion in this analysis. The first features tissue-specific gene expression data derived from GTEx, with a total of 37 tissues represented. The second and third datasets are derived from Zeisel et al., and represent 1) broad cell-type groups across the entirety of the mouse nervous system, and 2) a high-resolution single cell-type map of the same data. In Zeisel et al., a total of 39 broad cell-type groups and 265 individual cell-types are represented.

We followed the analysis protocol described in Bryois et al. for the analyses of 37 tissues from GTEx and 39 broad cell-type groups from Zeisel et al., and utilized a simplified approach for the analysis of each of the 265 individual cell-types from Zeisel et al. For the tissue and broad cell-type group datasets, we conducted the full protocol from Bryois et al. which included analyzing tissue/cell-type enrichment using both LDSC and MAGMA. For each dataset, p-values were adjusted for a false discovery rate of 5%. We only considered a tissue or cell-type significantly enriched if the FDR-adjusted p-value was less than 0.05 in both the LDSC and MAGMA-based tests. Due to the high computational demands of analysing 265 individual cell-types across the mouse nervous system in LDSC, we limited the assessment protocol to using MAGMA only, and considered a cell-type as significant if it had an FDR-adjusted p-value of less than 0.05. All statistical analyses downstream of LDSC and MAGMA (namely, p-value adjustment) were done using R v3.6.1, and all plotting was done inside of R v3.6.1 using the package ggplot2 v3.2.1.

**Overlap of the genome-wide significant locus with high-confidence chromatin interactions**

To derive chromatin interactions found in brain tissue that overlap with the genome-wide significant region(s) from the OCD GWAS, we submitted a corresponding analysis job to FUMA on May 3rd 2020 using the OCD GWAS sumstats as input. We selected the following chromatin interaction datasets for inclusion: 1) Promoter anchored loops from the PsychENCODE project, 2) loops from fetal cortex tissue, from Giusti-Rodriguez et al. 2019, and 3) loops from adult cortex tissue, also from Giusti-Rodriguez et al. 2019. All other parameters were set to default settings (SNP2GENE: a maximum p-value of lead SNPs < 5x10^{-8}, maximum p-value cutoff < 0.05, r^2 threshold to define independent significant SNPs ≥ 0.6, second r^2 threshold to define lead SNPs ≥ 0.1, reference population: 1000G phase 3 EUR, maximum distance between LD blocks to merge into a locus < 250kB. b) Positional gene mapping: distance to genes or functional consequences of SNPs on genes to map: 50 kb, SNP filtering: minimum CADD score ≥ 12.37, maximum RegulomeDB score = 7, annotation datasets eQTL catalogue: PsychENCODE, FANTOM5, Brain Open Chromatin Atlas. c) Gene eQTL mapping: Tissue types from BrainSeq Brain, BRAINEAC, GTEx v8 Brain, eQTL: eQTL p-value threshold: only significant snp-gene pairs with FDR < 0.05. GENE2FUNC: all background genes used, gene expression datasets: GTEx v8:54 tissue types, 30 general tissue types, 29 different ages of brain samples, and 11 general developmental stages of brain samples, Benjamin-Hochberg (FDR) correction was used to test for multiple testing in gene-enrichment analyses, maximum adjusted p-value for gene set association < 0.05, minimum overlapping genes with gene sets ≥ 2).

**SNP-based heritability and genetic correlation with other traits**

The proportion of variance in liability to OCD that could be explained by the aggregated effect of all included SNPs (SNP-based heritability) was estimated using LDSC. The SNP heritability is based on the estimated slope from the regression of the SNP effect sizes from the GWAS on the LD score. Genetic correlations between OCD and 82 other disorders/traits of potential relevance to OCD were estimated in cross-trait LDSC, a method that computes genetic correlations between GWASs without bias from ancestry differences or sample overlap. The genetic correlation between traits is based on the estimated slope from the regression of the product of Z-scores from two GWASs on the LD score and represents the genetic covariation between two traits based on all polygenic effects captured by the included SNPs. The source studies of the GWAS summary statistics used are summarized in Supplementary Table S9. The genome-wide LD information used by these methods was based on European populations from the HapMap 3 reference panel (see web resources). The GWAS summary statistics were filtered to only include SNPs that were part of the 1,290,028 million genome-wide HapMap 3 SNPs used in the original LD score regression studies. As a follow-up, we repeated the same analyses separately for three OCD subsets: 1) including only clinical samples, 2) including only (large-scale) biobank samples, and 3) including only data sets that were not primarily ascertainment for OCD (AUS and iPSYCH).

**Results**

**Genome-wide association results**

Prior to conducting the GWAS meta-analysis, QQ plots and genomic inflation factors (reported in Supplementary Table
SNP is located in the intergenic region between SFMBT1 and RFT1. This SNP (rs2581789, \( p = 2.1 \times 10^{-5} \), OR = 0.92) exceeded the genome-wide threshold for significance (Figure 1 for Manhattan plot, Figure 2 for regional plot of rs2581789, and Figure 3 for a forest plot showing the effect sizes in each individual cohort for rs2581789). This SNP is located in the intergenic region between SFMBT1 and RFT1 on chromosome 3p21.1. The LD block tagged by rs2581789 spans 325.8 kb (LD \( r^2 > 0.6 \)) and encompasses 12 genes, including NEK4, ITIH1, ITIH3, ITIH4, ITIH4-AS1, MUSTN1, STIMATE-MUSTN1, STIMATE, MIR8064, SFMBT1, RFT1, and PRKCD. Eight additional independent GWAS loci with \( p \)-values < 1.0 \times 10^{-5} were identified (corresponding association data, genomic regions, and genes in LD with the lead SNP are reported in Table 1, see Supplementary Figure S2-S9 for regional plots and Supplementary Figures S10-S17 for forest plots of each SNP).

This SNP (rs2581789) has previously been reported to be associated with schizophrenia (SCZ)\(^{47}\) and a combined SCZ/bipolar disorder (BP) phenotype\(^{48}\), as well as with anthropomorphic traits such as body mass index (BMI)\(^{49}\) and waist-hip-ratio\(^{49}\), and with psychological traits such as worry\(^{50-52}\) and well-being\(^{53}\). Other top-ranked SNPs identified in this analysis have also been previously reported to be significantly associated with psychological traits such as anxiety\(^{54}\), depressive symptoms\(^{55}\), neuroticism\(^{50,52,53,55}\), harm\(^{50}\), worry too long after embarrassment\(^{50,52}\), nervous feelings\(^{50}\) and worry/vulnerability\(^{54}\), as well as with cognitive performance and educational attainment\(^{56}\) (see Supplementary Table S2 for a full list of previously reported significant associations of the top nine SNPs identified in the OCD GWAS).

### Polygenic risk scoring

In the 'leave-one-out' PRS analyses we observed significantly higher OCD PRS among OCD cases compared to the controls in each target sample, explaining 3.9%, 3.5%, and 3.3% of overall phenotypic variance of IOCDF-GC, OCGAS, and Psych_Broad, respectively (Supplementary Figure S18).

### Gene-based analyses

We identified 18 genes that were significantly associated with OCD in at least one of the four gene-based methods (C-MAGMA, E-MAGMA, H-MAGMA, and S-PrediXcan) following correction for multiple hypothesis testing at the experiment-wide or transcriptome-wide level (see Table 2 for a list of all significant genes; full results can be found in Supplementary Tables S3, S4, S5, and S6). Genes identified by the four methods partially overlapped (Venn-diagram, Supplementary Figure S19). C-MAGMA revealed five genes significantly associated with OCD: PER1, SFMBT1, CTC1, TMEM107, and NEFH at a Bonferroni-corrected significance threshold of \( p < 2.50 \times 10^{-6} \) (Supplementary Table S3). H-MAGMA identified 14 risk genes across 7 genomic regions associated with OCD (Bonferroni-corrected \( p < 2.05 \times 10^{-6} \); Supplementary Table S4). Using E-MAGMA, we identified one gene that was significantly associated with OCD (CTC1 in whole blood, \( p < 2.51 \times 10^{-7} \), Supplementary Table S5). CTC1 was also the most significantly associated gene using S-PrediXcan (whole blood, \( p = 3.04 \times 10^{-7} \)), but did not survive multiple testing correction in this analysis (\( p < 1.78 \times 10^{-7} \) (Supplementary Table S6). We therefore meta-analyzed the tissue-specific S-PrediXcan associations (i.e. combining association signals across 48 GTEx tissues) using S-MultiXcan to generate a single test statistic (see Supplementary Table S7) for each gene. This approach yielded two significant \( (p < 2.31 \times 10^{-6}) \) genes: TMEM107 \( (p = 2.44 \times 10^{-7}) \) and RPL35 \( (p = 1.01 \times 10^{-7}) \).

### Functional annotation of GWAS findings

The functional annotation of individual GWAS findings (FUMA) is presented in Supplementary Table S8 and in the regional annotation plot for the genome-wide significant chromosome 3 locus (see Supplementary Figure S20). We explored tissue-specific expression patterns for 17 of the 18 genes resulting from the gene-based tests (excluding LOC101928274 for which no information was available). Results are visualized for the 54 GTEx tissues (see Supplementary Figure S21) and for Brainspan gene expression data (Supplementary Figures S22 and S23). Deferentially Expressed Gene (DEG) set analyses were conducted to test whether the set of 17 genes implicated by one or more gene-based tests are significantly up-regulated or down-regulated in any of the tissues. No significant differences were found (Supplementary Figures S24, S25, and S26).

### Tissue and cell-type enrichment analyses

We performed association tests for enrichment between OCD GWAS summary statistics and three gene expression profiles from human tissues and mouse cell-types. No tissue or cell-type was significantly associated with OCD after FDR correction for both the LDSC and MAGMA methods (Supplementary Figure S27A), thus failing to reach our criterion for significance. Nevertheless, tissues derived from brain regions clustered preferentially at the top of the results distribution. Also, no broad cell-type category reached significance for both LDSC and MAGMA (Supplementary Figure S27B). Here, the top-ranking result was 'telencephalon projecting excitatory neurons'. A further eight types included three types of di- and mesencephalon inhibitory neurons from the midbrain, two types of di- and mesencephalon excitatory neurons (one from thalamus and one from midbrain), and two types of telencephalon inhibitory interneurons from the hippocampus/cortex. In the high-resolution single cell-type analysis using MAGMA, nine specific cell-types were significantly enriched for OCD after FDR-correction (Sup-
Overlap of the genome-wide significant locus with high-confidence chromatin interactions

Genome-wide significant hits in other psychiatric GWAS studies have been shown to be enriched in overlap with brain-specific chromatin interactions\(^{44,57}\), and we sought to determine if this overlap (and potential subsequent mechanism of effect) existed for the genome-wide significant locus. We utilized FUMA\(^{36}\) to determine if the genome-wide significant locus from the GWAS (lead SNP located at hg19 chr3:53096230) overlapped with any high-confidence chromatin interactions derived from either PsychENCODE\(^{57}\) or fetal and adult cortex from Giusti-Rodriguez et al.\(^{44}\). While several loops connected the locus with a distal gene on the same chromosome (see Supplementary Figure S29), follow-up analyses did not reveal any genes with an obvious dosage sensitivity.

SNP-based heritability and genetic correlations with other traits

The SNP-based heritability of OCD as obtained from LD-score regression was estimated to be 0.164 (SE = 0.012). In genetic correlation analyses between OCD and summary statistics from 82 behavioral, cognitive, psychiatric, neurological, allergic/immunologic, metabolic, and anthropomorphic traits 31 of the 82 traits investigated had significant genetic correlations after correction for multiple testing (Figure 4 and Supplementary Table S9). In particular, OCD showed significant positive genetic correlations with all psychiatric disorders, with especially high correlations with anxiety disorder (\(r_g = 0.627, SE = 0.057, FDR\)-corrected \(p = 6.11 \times 10^{-27}\)), AN (\(r_g = 0.588, SE = 0.04, FDR\)-corrected \(p = 1.91 \times 10^{-48}\)), and MDD (\(r_g = 0.54, SE = 0.042, FDR\)-corrected \(p = 6.58 \times 10^{-36}\)). OCD was also positively correlated with alcohol dependence, while cannabis use disorder and other substance-use traits did not show significant correlations. Moreover, OCD was significantly genetically correlated with several cognitive/SES related traits, including a positive correlation with memory and negative correlations with intelligence, income, and job satisfaction. Also, some of the anthropomorphic traits (BMI, waist-hip-circumference, and hip-circumference) and auto-immune disorders (Crohn’s disease, ulcerative colitis, and inflammatory bowel disease) showed a significant negative correlation with OCD. In addition, significant correlations of substantial effect (\(r > 0.25\)) were found with neuroticism, suicidality, tiredness (all positive), and with subjective well-being (negative). Somewhat less prominent, but also significant was the positive correlation with childhood maltreatment. Genetic correlations with neurological, substance use, cardiovascular, and fertility traits were weak and non-significant.

As follow-up sensitivity analyses, we repeated the genetic correlation analyses for clinically-ascertained samples and biobank samples separately, as well as for the full OCD GWAS sample excluding the two datasets that were not primarily ascertained for OCD (AUS and DOTS). The resulting correlations were similar, though not identical, across these three OCD sub-samples (see Supplementary Figure S30). As an example for differences in the sub-samples, only the clinically-ascertained OCD samples had significant genetic correlations with ADHD and PTSD. Generally, the biobank-derived OCD samples showed less pronounced correlations than the other two subsets, which might be expected due to their smaller case sample sizes.

Discussion

In this GWAS, containing approximately five times the number of individuals diagnosed with OCD than previous studies\(^{13,18,19}\), we report the first definitive genome-wide significant locus for OCD (rs2581789 on chromosome 3p21.1). All SNPs combined in our analysis explained 16% of the heritability of OCD. We also showed substantial positive genetic correlations between OCD and a range of psychiatric disorders, including anxiety disorders, AN, and MDD. Our top SNP rs2581789 resides in a gene-rich genomic region (see Figure 2 for a regional association plot and a list of genes in the region). This region has previously been associated with a broad range of other psychiatric disorders and related traits, including SCZ\(^{47}\), well-being\(^{52}\), and the worry-subcluster of neuroticism\(^{30}\). Of note, this SNP was also identified in a recent cross-disorder meta-analysis of eight psychiatric disorders from the Psychiatric Genomics Consortium (PGC)\(^{24}\) (\(p = 6.51 \times 10^{-14}\) without the 23andMe data). Although the PGC cross-disorder analysis used a subset of the samples included in the current study, these previously published OCD cases\(^{13}\) represent only 20% of the cases in the current analysis, and did not have the power to identify any individual locus with genome-wide significance for OCD alone\(^{13}\). Despite being under-powered in the cross-disorder analysis\(^{24}\), OCD was among only three of the eight phenotypes that were shown to contribute to the rs2581789 association results with \(m\)-values > 0.9 (SCZ and BP being the other two). These findings indicate that the genome-wide significant SNP in this study is not solely associated with OCD, but rather with multiple psychiatric phenotypes, reflecting a pleiotropic effect, possibly contributing to a shared underlying neurobiological susceptibility across a subset of psychiatric disorders and other neurobehavioral traits. These findings add to the confidence that this SNP, and the encompassing 3p21.1 region, is robustly, but not uniquely associated with OCD. Notably, another recent manuscript\(^{58}\), reported one genome-wide significant region for a compulsive disorder factor (rs9821797 on chromosome 3p21.31, \(P = 3.61 \times 10^{-9}, r^2\) with rs2581789 < 0.01 and D < 0.2 in European ancestry individuals), using structural equation modeling and summary statistics from 11 major psychiatric disorders. This locus reached a \(p\)-value of 4.32 \times 10^{-4}\) in the present OCD GWAS, although it should be noted that there was partial overlap between samples using the same sample as the PGC CDG study (see above).

Gene-based analyses, incorporating multiple types of functional annotations, including eQTLs, fetal and adult brain
Hi-C information, and enrichment of tissue and single-cell RNAseq data, revealed no consistent associations. This is likely due to the absence of multiple, independent genome-wide association loci from the primary GWAS, as these are typically necessary to have enough statistical power for such studies. However, two genes were identified by three of the four gene-based methods: the Transmembrane Protein 107 gene (TMEM107) and the CST Telomere Replication Complex Component 1 gene (CTC1), both located in the same locus on chromosome 17 (see Table 2 for the respective results of the four different gene-based methods). Given that SNPs in these genes were not identified as genome-wide significant in the primary GWAS meta-analysis, there is not yet sufficient evidence to consider the potential role(s) of genes or SNPs in these loci in the etiology of OCD.

Our analyses of the shared genetic risk between OCD and other psychiatric disorders provides deeper insights into the etiology of OCD. Previous work, including the cross-disorder analyses discussed above\textsuperscript{24,58}, have confirmed the shared genetic risk between OCD and two neuropsychiatric disorders that are highly co-morbid with OCD - TS\textsuperscript{12,59} and AN\textsuperscript{60,61}. The above mentioned cross-disorder analyses also identified other psychiatric disorders as being correlated with OCD, albeit less strongly (SCZ, BP, and MDD). The present study, with its substantially increased sample size, confirms the genetic relationships between OCD and all of these disorders (TS, AN, SZ, BP, and MDD), and for the first time also shows significant genetic correlations between OCD and anxiety disorders, ADHD, as well as ASD (see Figure 4, Supplementary Figure S30 and Supplementary Tables S9 and S10).

In addition, our analyses of genome-wide genetic correlations between OCD and a wide range of additional brain- and non-brain-based traits identified a number of significant results that warrant a more in-depth study in the future (see also below for potential limitations of our study). The significant negative association between genome-wide genetic risk for OCD and BMI, body-fat and hip circumference may be reflective of the shared genetic risk between OCD and AN, while the positive genetic correlation between OCD and tiredness, loneliness, neuroticism, and suicidality may represent the known clinical overlap of OCD, anxiety and MDD at a genetic level. Less clear regarding its interpretation is the significant genetic correlation between OCD and childhood maltreatment. Prior clinical studies have reported increased rates of childhood maltreatment in individuals with OCD\textsuperscript{52,63}, which may correlate with co-occurring depression and anxiety. However, further studies are needed to shed light on the genetic relationship and the overlap in genetic contributions to their etiology.

One potential study limitation is the fact that, due to potentially higher than normal rates of psychiatric co-morbidity in many of our OCD datasets (e.g. through inclusion of individuals diagnosed with more severe OCD), our results may be biased towards identifying genes that contribute to multiple comorbid psychiatric disorders, or alternatively to an underlying trait related to general psychopathology, rather than to OCD specifically. A substantial proportion of new samples in the current study were derived from large-scale biobanks and national registries, two of which were drawn from studies that originally aimed to study other psychiatric disorders, such as depression (AUS), or that included a higher-than-average number of cases with depression, ADHD, and ASD (iPSYCH). As such, there is a larger degree of heterogeneity in our meta-analysis which likely influenced the genetic correlation results. We conducted sensitivity analyses to address and to mitigate these potential biases by dividing the OCD meta-analysis into three subsets including a) only samples from large-scale biobanks, b) only clinical samples, and c) excluding the two studies with comorbid ascertainment (AUS and iPSYCH) from the analysis. We observed in part substantial changes in the point estimates of the genetic correlations with some disorders and traits (e.g. depression, anxiety disorders, and ADHD).

We are currently conducting additional analyses to further address these limitations and will update the current version of this manuscript as soon as these results become available. Our observations echo similar results in other disorders (e.g. MDD) that call for more sophisticated analytical approaches (e.g. through inclusion of sensitivity meta-analytic methods based on structural equation modelling and other techniques) in order to combine samples with heterogeneous ascertainment. As with these other disorders the limitations inherent in our study design and in the available samples should be addressed in future genetic studies of OCD. In light of the above highlighted results, it is of note that we also performed leave-one-out polygenic risk score analyses using, among others, the previously published OCD meta-analysis of clinically-ascertained cases\textsuperscript{13}, which was used in all previous cross-disorder studies mentioned above. We found significant overlap in genetic liability between the prior OCD GWAS meta-analysis and leave-one-out GWAS data using predominantly samples that were newly added (mostly from large-scale biobanks) to our current meta-analysis (see Supplementary Figure S18). Once sample sizes sufficiently large to identify and replicate clear OCD susceptibility genes are collected, further analyses such as functional characterization, Mendelian randomization, and additional genetic correlation analyses may prove to be useful with the goal of further elucidating the underlying genetic etiology of OCD.

In sum, the present study provides insights into the current state of the largest GWAS for OCD to date. Our study identified the first definite genome-wide significant association of a locus with OCD (on chromosome 3p21.1). We are confident that with inclusion of additional samples (in the near future) we will be able to add new genome-wide significant regions to our current findings and further our understanding of OCD genetics. With these new meta-analysis results we will then also be able to use approaches (e.g., Mendelian Randomization and other) that will allow us to study the causal rela-
We are deeply grateful for the participation of all subjects contributing to this research. In Sweden we would further like to thank the local collection team: Anders Jureus, Jessica Pege, Malin Rådström, Radja Satgunanthan-Dawood, Miika Kremers, and Birgita Ohlander for all data management. We also wish to thank the National Quality Registry for Eating Disorders (RIKSÄT) for help with recruiting patients. We finally wish to thank the BBMRI.se and KI Biobank at Karolinska Institutet for professional biobank service. Data analysis for ESTB was carried out in the High-Performance Computing Center of the University of Tartu. AGDS would like to thank all the people who helped in the conception, implementation, beta testing, media campaign and data cleaning. They would specifically like to thank Knowledge Dales for support and guidance, Anna Hedberg, Per K. Kender, Patrick Sullivan, Andrew McIntosh and Cathryn Lewis for input on the questionnaire; Lorraine Nunn, Mary Ferguson, Lucy Winklter and Natalie Garden for data and sample collection; Natalia Zniczewska, Alissa Nichlces and Candace Brennan for participant recruitment support. Jonathan Davies, Luke Lowerd and Valerie Antonini for support with IT aspects; Vera Morgan and Ken Kirkby for help with the media campaign. AGDS would like to thank VIVAI Communications for their effort in promoting the study. They also acknowledge David Whitman and Catherine Olsen from Otskin. The authors thank the Psychiatric Genomics Consortium (PGC) for the use of their servers for data integration and analysis, the many families who have participated in the study, as well as the clinicians, study managers, and clinical interviewers at the respective study sites for their efforts in participant recruitment and data collection and assessment. The views expressed here do not reflect the views of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

CONFLICTS OF INTEREST

Ole Andreassen is a consultant to Healthlytx and has received speakers honorarium from Lundbeck and Sanovion in the past. CM Bulk reports: Shire (grant recipient), Scientific Advisory Board member); Idorsia (consultant); Lundbeckfonden (grant recipient); Pearson (author, royalty recipient). David Mataix-Cols receives royalties for contributing articles to UpToDate. Wolters Kluwer Health, and personal fees for editorial work from Elsevier, all unrelated to the current work. Hans-Joerg Grabe has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm, Servier and Janssen Cilag, as well as research funding from Fresenius Medical Care. Jan Haavik has received lecture honoraria as part of continuing medical education programs sponsored by Shire, Takeda and Medice. Professor Ian Hickie was an inaugural Commissioner on Australia’s National Mental Health Commission (2012-18). He is the Co-Director, Health and Policy at the Brain and Mind Centre (BMC) University of Sydney, Australia. The BMC operates an alcohol intervention youth service and a residential drug and alcohol treatment centre. Professor Hickie has previously led community-based and pharmaceutical industry-supported (Wyeth, Eli Lily, Servier, Pfizer, Astazeneca) projects focused on the identification and better management of anxiety and depression. He was a member of the Medical Advisory Panel for Medibank Private until October 2017, a Board Member of Psychosis Australia Trust and a member of Veterans Mental Health Clinical Reference Group. He is the Chief Scientific Advisor to, and a 5% equity shareholder in, InnoWell Pty Ltd. InnoWell was formed by the University of Sydney and Medibank Private to develop the $30 M Australian Government-funded Project Synergy (2017-20; a three-year program for the transformation of mental health services) and to lead transformation of mental health services through the innovative use of technologies. Erika Nummi is currently leading a project for Scientific Advisor in the field of the mental health industry and浙江大学, and is a member of the National Social-Economic and Medical Administration Board for Tea Pharmaceuticals and the Tourette Association of America.

WEB RESOURCES

ggplot2: www.ggplot2.tidyverse.org
Single-cell RNAseq data: www.github.com/jbr肴is/scRNA_disease

AFFILIATIONS

1Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany
2Departments of Psychiatry, Psychosomatics, and Psychotherapy, University Hospital Würzburg, Würzburg, Germany
3Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
4Department of Biomedical Research in Psychiatry and Clinical Genetics, University of Aarhus, Aarhus, Denmark
5Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
6Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
7Department of Mental Health, Gillette Children’s Specialty Health Care, Minneapolis, MN, USA
8Department of Biomedical Research in Psychiatry and Clinical Genetics, University of North Carolina At Chapel Hill, Chapel Hill, NC, USA
9Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
10CIMUS (centre for Research In Molecular Medicine and Chronic Diseases), University of Santiago De Compostela, Santiago De Compostela, A Coruña, Spain
11Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA
12Departments of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, USA
13Institute of Mental Health, RIKSÄT (National Quality Registry for Eating Disorders), Karolinska Institutet, Stockholm, Sweden
14Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
15Department of Psychological Medicine, University of Oxford, Oxford, UK
16National Institute of Public Health and the Environment, Bilthoven, The Netherlands
17Swedish Institute for Social Research, Stockholm, Sweden
18Mental Health Research Center, Karolinska Institutet, Stockholm, Sweden
19Institute of Biomedical Research in Psychiatry and Clinical Genetics, University of Aarhus, Aarhus, Denmark
20Department of Biomedical Research in Psychiatry and Clinical Genetics, University of Munich, Munich, Germany
21National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA
22Department of Psychiatry, Semmelweis University, Budapest, Hungary
23Biological Psychiatry Group, Broad Institute of MIT and Harvard, Cambridge, MA, USA
24Mood Disorders Laboratory, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA
25Department of Health Research Methods, Knowledge Translation, and Prevention, McMaster University, Hamilton, ON, Canada
26Division of Public Health Nutrition, School of Human Sciences, University of Auckland, Auckland, New Zealand
27Department of Psychology, Stanford University, Stanford, CA, USA
28Psychiatry Department, Karolinska Institutet, Stockholm, Sweden
29Department of Psychiatry, Karolinska Institutet, Stockholm, Sweden
30Department of Psychological Medicine, University of Oxford, Oxford, UK
31Institute of Mental Health, RIKSÄT (National Quality Registry for Eating Disorders), Karolinska Institutet, Stockholm, Sweden
32National Institute of Health, the Department of Health and Human Services, or the United States government

8 | medRxiv
The OCD Working Group of the Psychiatric Genomics Consortium et al. | PGC OCD2

medRxiv preprint doi: https://doi.org/10.1101/2021.10.13.21261078; this version posted October 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.
Medicine, Chicago, IL, USA
Department of Cardiology, University of Milan, Milan, Italy
Department of Mental Health Services, Copenhagen University Hospital, Copenhagen, Denmark
GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
Institute of Clinical Medicine, University of Oslo, Oslo, Norway
Departments of Child and Adolescent Psychiatry and Psychotherapy, University of Zurich, Zürich, Switzerland
Neuroscience Center Zurich, University of Zurich and The ETH Zurich, Zurich, Switzerland
Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland
Department of Psychology, University of Bergen, Bergen, Norway
Psychosis Research Unit, Aarhus University Hospital - Psychiatry, 8200 Aarhus N, Denmark
Psychiatric Genetics Unit, Vall d’Hebron Research Institute (VHIR), Universitat Autònoma De Barcelona, Barcelona, Spain
Biomedical Network Research Centre On Mental Health (CIBERSAM), Barcelona, Spain
Center for Basics In Neuromodulation, University of Freiburg, Freiburg, Germany
Department of Medicine, University of Oslo, Oslo, Norway
Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland
Department of Psychiatry, Cummning School of Medicine, University of Calgary, Calgary, AB, Canada
Program In Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada
Department of Cell Biology, SUNY Downstate Health Sciences University, Brooklyn, NY, USA
Department of Psychiatry, Amsterdam UMC Location AMC, Amsterdam, The Netherlands
Psychiatry and Genetics Institute, Center for OCD, Anxiety and Related Disorders, University of Florida, Gainesville, FL, USA
Department of Psychiatry, Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
Departments of Psychiatry and Community Health and Epidemiology, Dalhousie University, Halifax, NS, Canada
Institute of Psychiatric Phenomics and Genomics (IPPPG), University Hospital of Munich, Munich, Germany
Contributed equally

Contributed equally
References


Figures and Tables
<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>BP</th>
<th>P</th>
<th>OR</th>
<th>SE</th>
<th>A1/A2</th>
<th>FRQ_A</th>
<th>FRQ_U</th>
<th>INFO</th>
<th>genes.6.50kb(dist2index)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2581789</td>
<td>3</td>
<td>53096230</td>
<td>2.138x10^-8</td>
<td>0.91677</td>
<td>0.0155</td>
<td>0.328</td>
<td>0.353</td>
<td>0.996</td>
<td></td>
<td>NEK4(-241.3), ITIH1(-220.2), ITIH3(-203.2), ITIH4(-181.5), ITIH4-AS1(-186.9), MUSTN1(-177.2), STIMATE-MUSTN1(-114.7), STIMATE(-114.7), MIR8064(-165.7), SFMBT1(0.0), RFT1(0.0), PRKCD(49.0)</td>
</tr>
<tr>
<td>rs13262595</td>
<td>8</td>
<td>143316970</td>
<td>1.585x10^-7</td>
<td>1.07573</td>
<td>0.0139</td>
<td>0.454</td>
<td>0.424</td>
<td>0.993</td>
<td></td>
<td>LINC00051(0.0), TSNARE1(0.0), SERPING1(-233.0), MIR130A(-206.6), YPEL4(-197.9), CL1P(-186.0), ZDHHC5(-146.7), MED19(-135.5), TMX2(-106.9), TMX2-CTNND1(-28.7), SELENOH(-104.3), BTBD18(-96.1), CTNND1(-28.3)</td>
</tr>
<tr>
<td>rs674094</td>
<td>11</td>
<td>57665336</td>
<td>1.844x10^-7</td>
<td>1.08622</td>
<td>0.0159</td>
<td>0.319</td>
<td>0.332</td>
<td>0.99</td>
<td></td>
<td>SERPING1(-233.0), MIR130A(-206.6), YPEL4(-197.9), CL1P(-186.0), ZDHHC5(-146.7), MED19(-135.5), TMX2(-106.9), TMX2-CTNND1(-28.7), SELENOH(-104.3), BTBD18(-96.1), CTNND1(-28.3)</td>
</tr>
<tr>
<td>rs79712033</td>
<td>2</td>
<td>147846501</td>
<td>1.999x10^-7</td>
<td>1.17975</td>
<td>0.0318</td>
<td>0.945</td>
<td>0.937</td>
<td>0.971</td>
<td></td>
<td>LOC646029(-24.2), AQP11(0.0), CLNS1A(0.0), RHF17(0.0), AAMDC(121.2), INTS4(178.8), KCTD14(315.8), NDUC2-KCTD14(315.8), AP1B1(-29.4), REFL1S(0.0), REFL1(0.0), NEFH(0.0), THOC5(0.0), NIPSAP1(36.9), NF2(85.6)</td>
</tr>
<tr>
<td>rs7219489</td>
<td>17</td>
<td>8121622</td>
<td>2.79x10^-7</td>
<td>1.08405</td>
<td>0.0157</td>
<td>0.732</td>
<td>0.727</td>
<td>0.994</td>
<td></td>
<td>LOC646029(-24.2), AQP11(0.0), CLNS1A(0.0), RHF17(0.0), AAMDC(121.2), INTS4(178.8), KCTD14(315.8), NDUC2-KCTD14(315.8), AP1B1(-29.4), REFL1S(0.0), REFL1(0.0), NEFH(0.0), THOC5(0.0), NIPSAP1(36.9), NF2(85.6)</td>
</tr>
<tr>
<td>rs9535127</td>
<td>13</td>
<td>31606961</td>
<td>5.725x10^-7</td>
<td>1.07659</td>
<td>0.0148</td>
<td>0.487</td>
<td>0.49</td>
<td>0.993</td>
<td></td>
<td>RNF145(19.5), RANRF15(41.9), ODF4(71.5)</td>
</tr>
<tr>
<td>rs7128224</td>
<td>11</td>
<td>77360928</td>
<td>5.817x10^-7</td>
<td>1.0835</td>
<td>0.016</td>
<td>0.316</td>
<td>0.308</td>
<td>0.989</td>
<td></td>
<td>LOC646029(-24.2), AQP11(0.0), CLNS1A(0.0), RHF17(0.0), AAMDC(121.2), INTS4(178.8), KCTD14(315.8), NDUC2-KCTD14(315.8), AP1B1(-29.4), REFL1S(0.0), REFL1(0.0), NEFH(0.0), THOC5(0.0), NIPSAP1(36.9), NF2(85.6)</td>
</tr>
<tr>
<td>rs424541</td>
<td>22</td>
<td>29863922</td>
<td>8.089x10^-7</td>
<td>0.91439</td>
<td>0.0181</td>
<td>0.818</td>
<td>0.789</td>
<td>0.997</td>
<td></td>
<td>LOC646029(-24.2), AQP11(0.0), CLNS1A(0.0), RHF17(0.0), AAMDC(121.2), INTS4(178.8), KCTD14(315.8), NDUC2-KCTD14(315.8), AP1B1(-29.4), REFL1S(0.0), REFL1(0.0), NEFH(0.0), THOC5(0.0), NIPSAP1(36.9), NF2(85.6)</td>
</tr>
<tr>
<td>rs34289388</td>
<td>2</td>
<td>208280861</td>
<td>9.928x10^-7</td>
<td>1.08741</td>
<td>0.0171</td>
<td>0.8</td>
<td>0.788</td>
<td>0.998</td>
<td></td>
<td>LOC646029(-24.2), AQP11(0.0), CLNS1A(0.0), RHF17(0.0), AAMDC(121.2), INTS4(178.8), KCTD14(315.8), NDUC2-KCTD14(315.8), AP1B1(-29.4), REFL1S(0.0), REFL1(0.0), NEFH(0.0), THOC5(0.0), NIPSAP1(36.9), NF2(85.6)</td>
</tr>
</tbody>
</table>

Notes: SNP: single-nucleotide polymorphism; CHR: chromosome; BP: base pair position (hg19); P: p value; OR: odds ratio for allele1; A1/A2: effect allele/ reference allele; FRQ_A: frequency of allele1 in affected cases; FRQ_U: frequency of allele1 unaffected controls; INFO: imputation INFO score; genes.6.50kb(dist2index): list of genes within the region of LD-friends.6 (variants with LD-r2 > 0.6 to index SNP) ± 50 kb, in brackets distance to index SNP in kb.
# Table 2. Genes significantly associated with OCD, as indicated by 4 gene-based methods

<table>
<thead>
<tr>
<th>Locus</th>
<th>Top genes</th>
<th>BP start</th>
<th>BP stop</th>
<th>N SNPs</th>
<th>p value</th>
<th>Gene-based method*</th>
</tr>
</thead>
<tbody>
<tr>
<td>3p21.1</td>
<td>SPCS1</td>
<td>52738971</td>
<td>52745162</td>
<td>59</td>
<td>2.57x10^{-7}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NEK4</td>
<td>52742465</td>
<td>52804965</td>
<td>43</td>
<td>1.47x10^{-6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TMEM110</td>
<td>52870772</td>
<td>52931612</td>
<td>34</td>
<td>1.04x10^{-6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SFMBT1</td>
<td>52933221</td>
<td>53080089</td>
<td>274</td>
<td>2.83x10^{-7}</td>
<td></td>
</tr>
<tr>
<td>9q33.3</td>
<td>RPL35</td>
<td>127620158</td>
<td>127624260</td>
<td>N/A</td>
<td>1.01x10^{-6}</td>
<td></td>
</tr>
<tr>
<td>9q33.3</td>
<td>HSPA5</td>
<td>127997132</td>
<td>128003609</td>
<td>72</td>
<td>5.49x10^{-7}</td>
<td></td>
</tr>
<tr>
<td>11q12.1</td>
<td>TIMM10</td>
<td>57295936</td>
<td>57298276</td>
<td>27</td>
<td>3.73x10^{-7}</td>
<td></td>
</tr>
<tr>
<td>12q24.23</td>
<td>LOC101928274</td>
<td>118501398</td>
<td>118541753</td>
<td>209</td>
<td>1.22x10^{-6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VSIG10</td>
<td>118501398</td>
<td>118541753</td>
<td>209</td>
<td>1.22x10^{-6}</td>
<td></td>
</tr>
<tr>
<td>13q12.3</td>
<td>MEDAG</td>
<td>31480328</td>
<td>31499709</td>
<td>84</td>
<td>1.07x10^{-7}</td>
<td></td>
</tr>
<tr>
<td>13q13.3</td>
<td>POSTN</td>
<td>38136720</td>
<td>38172981</td>
<td>333</td>
<td>1.52x10^{-6}</td>
<td></td>
</tr>
<tr>
<td>17p13.1</td>
<td>ACAP1</td>
<td>7239848</td>
<td>7254797</td>
<td>105</td>
<td>2.47x10^{-8}</td>
<td></td>
</tr>
<tr>
<td>17p13.1</td>
<td>PER1</td>
<td>8043788</td>
<td>8059723</td>
<td>30</td>
<td>8.89x10^{-8}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TMEM107</td>
<td>8076555</td>
<td>8079717</td>
<td>13</td>
<td>1.39x10^{-7}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNORD118</td>
<td>8076772</td>
<td>8076905</td>
<td>7</td>
<td>2.38x10^{-7}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BORCS6</td>
<td>8091651</td>
<td>8094225</td>
<td>14</td>
<td>6.63x10^{-8}</td>
<td></td>
</tr>
<tr>
<td>17p13.1</td>
<td>CTC1</td>
<td>8128139</td>
<td>8151413</td>
<td>134</td>
<td>2.30x10^{-7}</td>
<td></td>
</tr>
<tr>
<td>22q12.2</td>
<td>NEFH</td>
<td>29876181</td>
<td>29887279</td>
<td>27</td>
<td>4.74x10^{-7}</td>
<td></td>
</tr>
</tbody>
</table>

Notes: P values thresholds for each gene-based method was calculated using Bonferroni correction, for the total number of tests performed for a given method. The P value for the most significant gene-based association is reported in the Table. C-MAGMA: 2.50x10^{-6} (N=20,031 tests); E-MAGMA: P<2.51x10^{-7} (N=199,421 tests); H-MAGMA: P<2.05x10^{-6} (N=24,358 tests); S-MultiXcan: P<2.31x10^{-6} (N=21,601 tests). The table describes the locus, top gene-based associations within each locus (top genes), location of the associated gene in base pairs (hg19) (BP start and BP stop, respectively), the number of SNPs included in the model (N SNPs), e.g., . * Black shading indicates the method with the most significant gene-based P value; grey shading indicates a significant association for the same gene using a different gene-based method.
Fig. 1. Manhattan plot of the results from the GWAS meta-analysis of OCD: The y-axis represents $-\log_{10} p$ values for association of variants with OCD, from meta-analysis using an inverse-variance weighted fixed effects model. The x-axis represents chromosomes 1 to 22. The horizontal red line represents the threshold for genome-wide significance. The index variant of the one genome-wide significant locus is highlighted as a green diamond.
Fig. 2. Regional plot of the genome-wide significant locus rs2581789. The $-\log_{10}(p)$-value of SNPs in the OCD meta-analysis GWAS is shown on the left y axis. The recombination rates expressed in centimorgans (cM) per Mb (Megabase) (blue line) are shown on the right y axis. Position in Mb is on the x axis. Only the SNPs with association p-value less than 0.1 were plotted. The most associated SNP is shown as a purple diamond.
Fig. 3. Forest plot of SNP rs13262595. The plot shows the effect estimate with 95%-confidence interval for each cohort contributing to the meta-analysis and for the inverse variance weighted meta analysis. The table lists INFO (imputation score), p-value, f_ca(n) (frequency cases), f_co(n) (frequency controls), ln(OR), and STDerr (standard error) for each of the contributing cohorts and for the meta-analysis. At the top, + indicates a positive direction of effect, - a negative direction of effect while ? indicates that the SNP was not contained in the respective cohort.

<table>
<thead>
<tr>
<th>rs2581789</th>
<th>C/T</th>
<th>3:53096230</th>
<th>info</th>
<th>p_value</th>
<th>f_ca(n)</th>
<th>f_co(n)</th>
<th>ln(OR)</th>
<th>STDerr</th>
</tr>
</thead>
<tbody>
<tr>
<td>IOCDF-GC TR</td>
<td>1.01</td>
<td>0.111</td>
<td>0.317(323)</td>
<td>0.360(323)</td>
<td>-0.187</td>
<td>0.117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUS</td>
<td>0.99</td>
<td>0.9</td>
<td>0.338(757)</td>
<td>0.338(5368)</td>
<td>-0.00747</td>
<td>0.0593</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOTS</td>
<td>0.98</td>
<td>0.0826</td>
<td>0.332(2678)</td>
<td>0.344(10410)</td>
<td>-0.0571</td>
<td>0.0329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPOC</td>
<td>0.99</td>
<td>0.426</td>
<td>0.343(195)</td>
<td>0.371(204)</td>
<td>-0.122</td>
<td>0.153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estonia</td>
<td>1.00</td>
<td>0.227</td>
<td>0.385(493)</td>
<td>0.403(138015)</td>
<td>-0.079</td>
<td>0.0654</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finngen</td>
<td>1.00</td>
<td>0.33</td>
<td>0.313(780)</td>
<td>0.335(161390)</td>
<td>-0.0539</td>
<td>0.0553</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUNT</td>
<td>0.99</td>
<td>0.25</td>
<td>0.353(284)</td>
<td>0.325(66192)</td>
<td>0.103</td>
<td>0.0895</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MoBa</td>
<td>1.00</td>
<td>0.352</td>
<td>0.341(104)</td>
<td>0.311(2193)</td>
<td>0.145</td>
<td>0.156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NORDiC-nor</td>
<td>1.01</td>
<td>0.0702</td>
<td>0.301(365)</td>
<td>0.355(315)</td>
<td>-0.233</td>
<td>0.129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NORDiC-swe_EGOS</td>
<td>0.99</td>
<td>0.126</td>
<td>0.328(1997)</td>
<td>0.340(3943)</td>
<td>-0.0648</td>
<td>0.0423</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKBB</td>
<td>0.99</td>
<td>0.0141</td>
<td>0.307(776)</td>
<td>0.337(125729)</td>
<td>-0.134</td>
<td>0.0545</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IOCDF-GC CC</td>
<td>0.98</td>
<td>0.0019</td>
<td>0.310(1519)</td>
<td>0.344(3541)</td>
<td>-0.15</td>
<td>0.0483</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCGAS</td>
<td>1.01</td>
<td>0.00637</td>
<td>0.318(986)</td>
<td>0.359(1023)</td>
<td>-0.183</td>
<td>0.0671</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psych_Broad</td>
<td>0.99</td>
<td>0.0062</td>
<td>0.347(1396)</td>
<td>0.382(4009)</td>
<td>-0.129</td>
<td>0.0471</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWF</td>
<td>1.01</td>
<td>0.0879</td>
<td>0.289(140)</td>
<td>0.344(526)</td>
<td>-0.252</td>
<td>0.147</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Meta-analysis | 1.00 | 2.14e−08 | 0.328(14140) | 0.353(562117) | -0.0869 | 0.0155 |

ln(OR), 95% CI
Fig. 4. Genetic correlations \((r_g)\) between OCD and a broad range \((N=82)\) of other phenotypes. Error bars represent 95\% confidence intervals and asterisks indicate significant associations after FDR correction for multiple testing. Non-significant correlations with \(SE > 0.5\) are excluded from display.