

UCLA

UCLA Previously Published Works

Title

Genome-wide association study of pediatric obsessive-compulsive traits: shared genetic risk between traits and disorder.

Permalink

<https://escholarship.org/uc/item/48s7t244>

Journal

Translational psychiatry, 11(1)

ISSN

2158-3188

Authors

Burton, Christie L

Lemire, Mathieu

Xiao, Bowei

et al.

Publication Date

2021-02-01

DOI

10.1038/s41398-020-01121-9

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

ARTICLE

Open Access

Genome-wide association study of pediatric obsessive-compulsive traits: shared genetic risk between traits and disorder

Christie L. Burton¹, Mathieu Lemire¹, Bowei Xiao², Elizabeth C. Corfield¹, Lauren Erdman², Janita Bralten^{3,4}, Geert Poelmans³, Dongmei Yu^{5,6}, S.-M. Shaheen^{7,8}, Tara Goodale¹, Vanessa M. Sinopoli^{2,9}, OCD Working Group of the Psychiatric Genomics Consortium, Noam Soreni¹⁰, Gregory L. Hanna¹¹, Kate D. Fitzgerald¹¹, David Rosenberg¹², Gerald Nestadt¹³, Andrew D. Paterson^{2,14}, Lisa J. Strug^{2,15}, Russell J. Schachar^{1,16}, Jennifer Crosbie^{1,16} and Paul D. Arnold^{2,7,8,16}

Abstract

Using a novel trait-based measure, we examined genetic variants associated with obsessive-compulsive (OC) traits and tested whether OC traits and obsessive-compulsive disorder (OCD) shared genetic risk. We conducted a genome-wide association analysis (GWAS) of OC traits using the Toronto Obsessive-Compulsive Scale (TOCS) in 5018 unrelated Caucasian children and adolescents from the community (Spit for Science sample). We tested the hypothesis that genetic variants associated with OC traits from the community would be associated with clinical OCD using a meta-analysis of all currently available OCD cases. Shared genetic risk was examined between OC traits and OCD in the respective samples using polygenic risk score and genetic correlation analyses. A locus tagged by rs7856850 in an intron of *PTPRD* (protein tyrosine phosphatase δ) was significantly associated with OC traits at the genome-wide significance level ($p = 2.48 \times 10^{-8}$). rs7856850 was also associated with OCD in a meta-analysis of OCD case/control genome-wide datasets ($p = 0.0069$). The direction of effect was the same as in the community sample. Polygenic risk scores from OC traits were significantly associated with OCD in case/control datasets and vice versa (p 's < 0.01). OC traits were highly, but not significantly, genetically correlated with OCD ($r_g = 0.71$, $p = 0.062$). We report the first validated genome-wide significant variant for OC traits in *PTPRD*, downstream of the most significant locus in a previous OCD GWAS. OC traits measured in the community sample shared genetic risk with OCD case/control status. Our results demonstrate the feasibility and power of using trait-based approaches in community samples for genetic discovery.

Introduction

Obsessive-compulsive disorder (OCD) is a common (1–2% prevalence)¹ psychiatric disorder characterized by intrusive, recurrent thoughts and repeated, ritualized behaviors. Up to 50% of OCD cases have a childhood-onset (before the age of 18)², which is more heritable than

adult-onset OCD³. Two genome-wide association studies (GWAS) in clinical samples with mixed ages of OCD-onset and a meta-analysis of these studies did not identify genome-wide significant loci^{4–6}. The most significant loci from previous GWAS include SNPs within *DLGAP1*, *BTBD3*, *GRID2*, and one close to *PTPRD*. Using obsessive-compulsive (OC) symptoms rather than a clinical diagnosis, a study of adult twins identified a genome-wide significant SNP in *MEF2B* (rs8100480)⁷. However, this SNP was not replicated in an independent sample⁵.

Correspondence: Christie L. Burton (christie.burton@sickkids.ca)

¹Neurosciences and Mental Health, Toronto, Canada

²Genetics and Genome Biology Hospital for Sick Children, Toronto, Canada

Full list of author information is available at the end of the article

© The Author(s) 2021



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

We conducted a GWAS of quantitative OC traits in a large pediatric, community-based sample: Spit for Science^{8,9}. We measured OC traits using the Toronto Obsessive-Compulsive Scale (TOCS; <https://lab.research.sickkids.ca/schachar/resources-and-tools/>)⁸. This heritable measure¹⁰ includes negative scores that represent ‘strengths’ (e.g., never upset when their belongings are rearranged) and positive scores that represent ‘weaknesses’ (e.g., very upset when their belongings are rearranged). We reasoned that a strength-to-weakness format would generate scores with a more normal distribution in a community sample⁸ than those observed with typical OCD scales and would therefore boost the power of genetic discovery¹¹. Typical OCD trait measures generate J-shaped distributions because their format calls for ratings of symptoms from absence to presence (score of zero to a positive integer). A j-shaped distribution is especially likely when using typical OCD measures in a community sample where the prevalence of OC symptoms is low and most people would get scores of zero¹². This j-shaped distribution can be replicated with the TOCS by collapsing the ‘strengths’ (i.e., negative scores) into scores of zero (Fig. 1). We tested the hypothesis that the distribution of TOCS scores would boost the power of genetic discovery¹¹ by running a GWAS with the collapsed TOCS measure as well as the full distribution. We characterized the genetic associations for TOCS by conducting gene-based analyses, examining brain expression quantitative trait loci (eQTLs) of the most significant loci, estimating SNP-based heritability and genetic correlations of total OC trait scores with other medical/mental health disorders and traits. We also examined if the most significant loci from the previous GWAS of OC symptoms⁷ replicated in our study. Finally, we tested the hypothesis that OC traits in the community share genetic risk with OCD by examining individual genetic variants, genetic correlations, and polygenic risk between OC traits in Spit for Science and three independent OCD case/control samples.

Subjects and methods

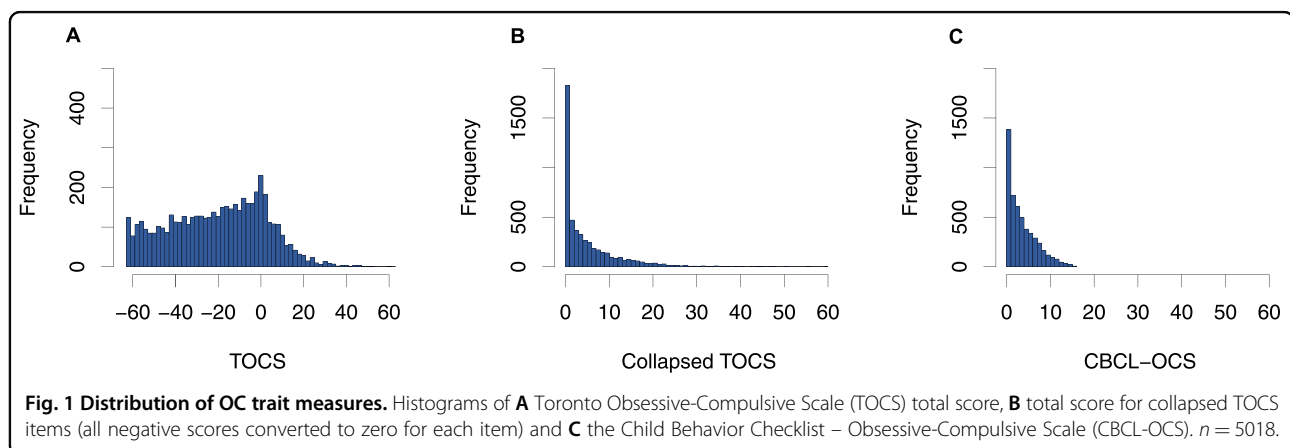
OC traits

Participants

The Spit for Science sample is described in detail elsewhere⁹. Briefly, the sample included 15,880 participants with complete demographic, questionnaire, and family information (mean age = 11.1 years [SD 2.8]; 49.4% female) from the 17,263 youth (6–18 years of age) recruited at the Ontario Science Centre over 16 months. Informed consent, and assent where applicable, were obtained using a protocol approved by the local Research Ethics Board at the Hospital for Sick Children. Participants provided a saliva sample in Oragene saliva kits (OG-500; DNA Genotek, Ottawa, Canada) for genetic analyses. See the supplement for details.

OC trait measure

We measured parent- and self-reported OC traits within the last 6 months using the TOCS, a 21-item questionnaire described previously^{8,10}. Each item was scored on a 7-point Likert scale ranging from –3 (‘far less often than others of the same age’) to +3 (‘far more often than others of the same age’). A score of zero was designated as an average amount of time compared to same-age peers. The TOCS total score was standardized into a z-score to account for age, sex, and questionnaire respondent (parent or self). Details of z-score creation are described in the supplement. We tested the impact of the strength/weakness structure of the TOCS by re-scoring the TOCS to convert all negative scores for individual items to zero before summing scores (i.e., no scores less than 0, which collapsed the left side of the distribution). We also compared the TOCS to an additional OCD symptom measure with a j-shaped distribution: The Obsessive-Compulsive Scale of the Child Behavior Checklist (CBCL-OCS)¹³. Each of the eight CBCL-OCS items was scored on a scale of 0–2 (0 = not true; 1 = somewhat/sometimes true; and 2 = very/often true) and was summed to generate a total score



(range: 0–16). This ‘collapsed’ TOCS total score, with a cluster of scores at zero, created a distribution similar to the CBCL-OCS (Fig. 1).

Genetic data

DNA was extracted manually from saliva using standard methods (see the supplement for additional details). We excluded any samples with concentrations <60 ng/μl and insufficient quality based on agarose gels. We genotyped 5645 samples on the Illumina HumanCoreExome-12v1.0_B (HumanCore) and 192 samples on the Illumina HumanOmni1-Quad V1.0_B (Omni) bead chip arrays (Illumina, San Diego, CA, USA) at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, CA). There were 538,448 markers on the HumanCore and 1,140,419 markers on the Omni array.

Quality control (QC) was conducted separately for each array using standard methods with PLINK v1.90¹⁴. Sample exclusion and selection criteria are described in the supplemental methods and Supplemental Figure S1. Imputation was performed separately for all platforms and sample sets, using Beagle v4.1 using the data from phase 3, version 5 of the 1000 Genomes project for reference (http://bochet.gcc.biostat.washington.edu/beagle/1000_Genomes_phase3_v5a/). We excluded individuals who were non-Caucasian based on principal component (PC) analysis and included only one participant from each family (inferred sibs or half-sibs, see supplement and Supplemental Figure S2). Genetic data will be available through the SickKids Healthy Kids Biobank.

Analyses

GWAS was conducted using R (v3.5.1). Our primary analysis tested if imputed dosage and standardized TOCS total score were associated using a linear regression model that included the top three PCs and genotyping array as covariates. We included SNPs with a minor allele frequency (MAF) > 1%, allelic R^2 imputation quality ($AR2 > 0.6$) and used the standard genome-wide threshold of $p \leq 5 \times 10^{-8}$. We also tested if any genome-wide significant variants from the analysis with the standardized scores were still significant using a non-standardized TOCS score. For these analyses, age, sex, respondent, and their 2- and 3-way interactions were used as covariates in addition to the above (interactions were included to mimic the construction of the Z scores, which were calculated independently in age-, sex- and respondent-defined bins; see supplement).

In secondary analyses, we evaluated the association between SNPs and the collapsed TOCS score, and between SNPs and CBCL-OCS, using zero-inflated negative binomial likelihood ratio tests, using the function *zeroinfl* from the R package *pscl* (v1.5.2). This model was chosen because of the high proportion of zero scores that created a j-shaped distribution. The test is a mixture

of two models: a negative binomial model, which contributes to zero and positive scores, and a logit model, which contributes to possible inflation of zero scores (point mass at 0) compared to what a negative binomial model predicts. These analyses used non-standardized scores for the collapsed TOCS and CBCL-OCS so the model adjusted for the covariates and the association of SNP allele dosage with the OC trait scores is tested against the null of having no effects on both the logit part and the negative binomial part using likelihood ratio tests.

We subsequently used FUMA to conduct a gene-based GWAS of the TOCS standardized total score with MAGMA using a Bonferonni correction for the number of protein-coding genes included¹⁵ (fuma.ctglab.nl).

We tested each genome-wide significant variant for colocalization with brain eQTLs using LocusFocus¹⁶ (<https://locusfocus.research.sickkids.ca/>). We examined the 14 GTEx sets from brain tissue types and examined SNPs within ± 1 Mbp of each SNP.

We estimated SNP heritability using both GCTA¹⁷ v1.91.2-beta (<http://cns.genomics.com/software/gcta/>) with further exclusion of cousins and SNPs with $AR2 > 0.9$ and LDSC¹⁸ (v1.0.0, <https://github.com/bulik/ldsc>) calculated from SNPs in HapMap3. We used LDSC¹⁹ to examine the genetic correlation of TOCS total scores with the 850 phenotypes available on LD Hub (<http://ldsc.broadinstitute.org/ldhub/>).

Finally, we examined the p -values and effect sizes of the top variants from the present study in the only previous GWAS of OC symptoms⁷ of 6931 twins and sibs from the Netherlands Twin Registry (only 20 loci reported in the results from the previous paper were also in SNP set from the present study).

OCD case/control

Participants

For validation analyses, we investigated three independent OCD case/control cohorts: (1) the International OCD Foundation Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OC GAS) meta-analysis⁶, (2) the Philadelphia Neurodevelopmental Cohort (PNC) from the Children’s Hospital of Philadelphia (CHOP)²⁰, and (3) the Michigan/Toronto OCD Imaging Genomics Study²¹. See the supplement and Table 1 for sample sizes.

Analyses

To validate findings from the GWAS of the TOCS total score, we combined the GWAS summary statistics from each OCD cohort using a fixed-effect inverse variance meta-analysis. For completeness, we also conducted a meta-analysis of the summary statistics from the GWAS of the TOCS total score with the OCD samples using a modified sample size-based weighted meta-analysis

Table 1 Overview of samples.

OC trait sample	Phenotype	Age	Total samples	Cases	Controls
Spit for Science	TOCS, CBCL-OCS	6–18 years	5018	n/a ^a	n/a
All replication	Phenotype	Age of onset			
IOCDF/OCGAS	Clinician-diagnosed (DSM-IV)	Child and adult	9725	2688	7037
CHOP	Case/control status based on GO-ASSESS symptoms	<18 years of age	1775	406	1369
Michigan/Toronto	Clinician-diagnosed (DSM-IV)	<18 years of age	480	275	205
Total meta-analysis			11,980	3369	8611

^aThere were 62 individuals that had a parent- or self-reported diagnosis of OCD. Number of samples reflects samples included in analyses (i.e., after quality control analyses). IOCDF/OCGAS = International OCD Foundation Collaborative and OCD Collaborative Genetics Association Studies meta-analysis sample, CHOP = the Philadelphia Neurodevelopmental Cohort (PNC) from the Children's Hospital of Philadelphia, TOCS = Toronto Obsessive-Compulsive Scale, CBCL-OCS = Child Behavior Checklist - Obsessive-Compulsive Scale, DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th edition.

method for combining continuous and categorical variables²² (see supplemental methods for details). In brief, this approach weights the two sets of results based on their SNP heritability and genetic correlation. Polygenic risk score (PRS) analyses were performed using LDpred v1.06²³ (see supplement). First, we derived PRS for TOCS from the Spit for Science sample and tested their association with case/control status in the combined OCD cohorts (target sample: CHOP, Michigan/Toronto and a subset of the IOCDF-GC/OCGAS—see supplement). Second, we derived PRS from the combined OCD cohorts and tested their association with the standardized TOCS total score in the Spit for Science sample (target sample). We examined the potential shared genetic risk between the Spit for Science and the meta-analyzed OCD samples using genetic correlations estimated with LDSC¹⁹.

Results

OC traits

We used 5018 participants for GWAS analyses after sample exclusion and selection (see supplement and Supplemental Figures S1/S2). In the primary analysis, rs7856850 in *PTPRD* was significantly associated with TOCS total scores at the genome-wide level ($p = 2.48 \times 10^{-8}$, $\beta = 0.14$, s.e. = 0.025, $R^2 = 0.618\%$; Fig. 2A, most significant loci listed in Supplemental Table S1). Several variants in this region that approached genome-wide significance were in linkage disequilibrium (LD) with rs7856850, which was genotyped on both the HumanCore and OMNI arrays (Fig. 2B). The inflation factor λ was 1.008 while the intercept of LD score regression was 1.003 and not significantly different from 1 (s.e. = 0.007, $p = 0.66$; Fig. 2C). rs7856850 was still associated with TOCS total scores using raw instead of standardized scores ($p = 2.75 \times 10^{-8}$, $R^2 = 0.615\%$; data not shown). There was no eQTL data for the SNP in *PTPRD*, rs7856850, in LocusFocus¹⁶, or in the most recent version of GTEX v8²⁴.

When we analyzed the collapsed TOCS total score and the CBCL-OCS, the genome-wide significant locus for the

TOCS total score rs7856850 was no longer genome-wide significant, although the remaining effect was in the same direction and had the same direction of effect ($p = 0.00045$ and $p = 0.025$, respectively; see supplement for details). For both collapsed TOCS and the CBCL-OCS, the A allele was associated with both higher scores (collapsed TOCS: $\beta = 0.0735$, s.e. = 0.0292, CBCL-OCS: $\beta = 0.0465$, s.e. = 0.020) and lower proportion of zero scores (collapsed TOCS: $\beta = -0.231$, s.e. = 0.104, CBCL-OCS: $\beta = -0.126$, s.e. = 0.179).

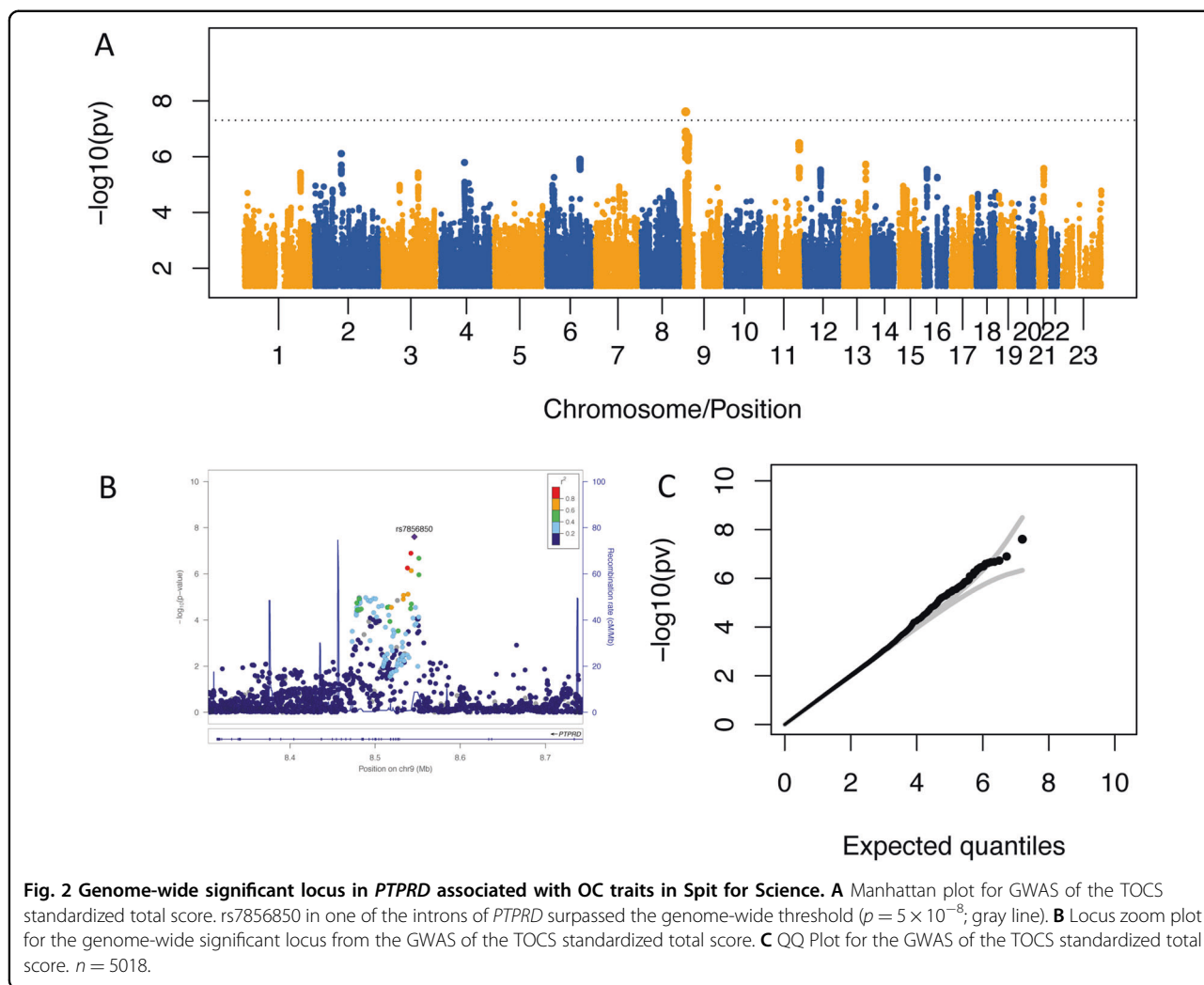
A gene-based GWAS of the TOCS standardized total score using MAGMA on the FUMA platform did not identify any genome-wide significant genes (at a Bonferroni-corrected level $p = 0.05/19369$ protein-coding genes = 2.58×10^{-6}). The most significant genes were *SH3GL2* ($p = 4.21 \times 10^{-6}$, $z = 4.45$); *RRN3* ($p = 6.23 \times 10^{-6}$, $z = 4.37$), and *PDXDC1* ($p = 1.10 \times 10^{-5}$, $z = 4.24$; Supplemental Figure S3). *PDXDC1* and *RRN3* have overlapping coding regions.

The heritability of the TOCS total score was $h^2 = 0.068$ (s.e. = 0.052, $p = 0.19$) using GCTA and $h^2 = 0.071$ (s.e. = 0.060; $p = 0.24$) using LDSC when the intercept was constrained to 1. TOCS total score was not significantly associated with any phenotypes on LD Hub (see supplement).

One of the top-ranked SNPs from a previous GWAS of OC symptoms⁷ was nominally associated with TOCS total scores in the Spit for Science sample with the same direction of effect (rs60588302, $p = 0.025$). This SNP is in the same region as our most significant locus (9p24.1), but not in LD ($r^2 = 0.004$, $D' = 0.517$). Another 16 of the reported most significant loci in den Braber⁷, including a variant in *MEF2BNB* (rs8100480) that was genome-wide significant in their sample, had effects in the same direction but were not significantly associated in the current sample (Supplemental Table S2).

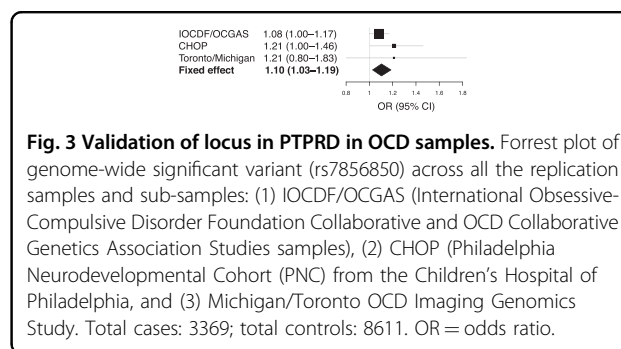
OCD case/control

Following standard QC and sample exclusion where applicable (see supplement), we had a total of 3369 cases

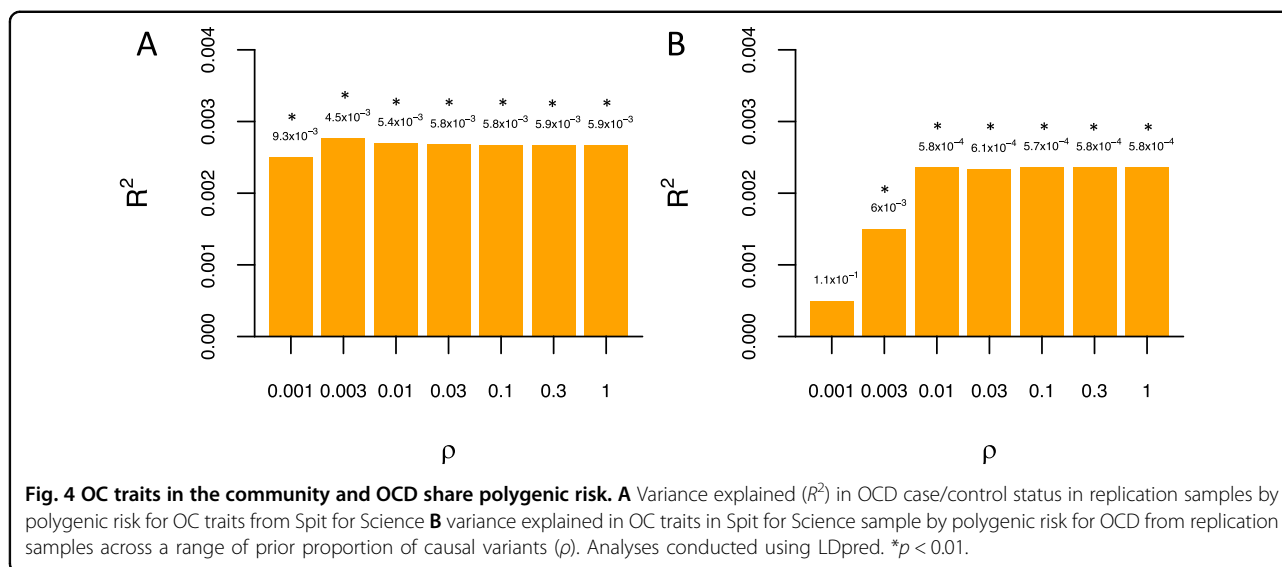


and 8611 controls in our validation samples (Table 1). We tested if the genome-wide SNP associated with TOCS total scores in Spitz for Science were also associated with OCD in the meta-analysis of case/control cohorts. *rs7856850* was associated with increased odds of being an OCD case ($p = 0.0069$, $OR = 1.104$ per A allele [95% confidence limit 1.03–1.19], Fig. 3, Supplemental Figure S4). When the summary statistics of the TOCS total score were meta-analyzed with the OCD cohorts, there were no genome-wide significant variants (Supplemental Figure 5). *rs7856850* approached genome-wide significance $p = 1.2 \times 10^{-7}$ when using conventional sample size-weighted meta-analysis but fell to $p = 0.00054$ when the sample sizes were adjusted for SNP heritability and genetic correlation (see supplemental results for details).

The genetic correlation between standardized TOCS total scores and OCD meta-analysis was $r_g = 0.71$ (s.e. = 0.382; $p = 0.062$; 95% CI: [-0.04,1]) when intercepts are constrained to 1. Figure 4A shows that PRS calculated for



TOCS total scores was significantly associated with increased odds of being a case in the meta-analyzed OCD samples (Nagelkerke’s pseudo $R^2 = 0.277\%$, $p = 0.0045$ at $\rho = 0.003$). Figure 4B shows that PRS constructed from the OCD sample were significantly associated with TOCS total scores in Spitz for Science ($R^2 = 0.24\%$; $p = 0.00057$ at $\rho = 0.1$).



Discussion

Using a trait-based approach in a community sample, we identified a genome-wide significant variant associated with OC traits (rs7856850) that was also associated with OCD case/control status. Polygenic risk and genetic correlation findings showed sharing of genetic risks between OC traits in the community and OCD case/control status in independent samples.

The genome-wide significant variant (rs7856850) associated with OC traits is in an intron of the consensus transcript of *PTPRD* that codes for protein tyrosine phosphatase δ . No eQTLs have been calculated yet for rs7856850 (GTEx V8)²⁴. To validate this finding, we tested if this SNP was also associated with OCD in a meta-analysis of three independent cohorts. The significant association of rs7856850 with OCD case/control makes it the first variant associated with OC traits and OCD. For completeness, we presented genome-wide results for the meta-analysis of the OCD cohorts as well as a meta-analysis of TOCS total score with the OCD cohorts, where no genome-wide significant findings were revealed. However, the direction of effect for rs7856850 was in the same direction in all samples. The small size of the OCD cohorts likely precluded finding genome-wide significant SNPs. In the meta-analysis of OC traits and OCD, summary statistics were combined using sample size-based weights that were modified and calibrated to account for SNP heritability to reflect differences in power and ascertainment between continuous (OC traits) and categorical (OCD case/control) designs²⁵. The low SNP heritability of TOCS severely down-weighted the OC trait sample size while up-weighting the already underpowered OCD case/control cohorts. Larger samples will be helpful to confirm the results from the present study.

Previous GWAS of OCD symptoms or diagnosis identified variants that approached significance in the region around *PTPRD*. However, those variants were independent of the locus found in our study^{4,7}. These observations support a possible role of the 9p24.1 region in OCD. The 9p region is also the location of one of the strongest linkage peaks in earlier genome-wide linkage studies of pediatric OCD^{26,27}. Rare CNVs in *PTPRD* have been identified in cases with OCD²¹ and ADHD²⁸. SNPs in *PTPRD* were genome-wide significantly associated with ASD²⁹, restless legs syndrome³⁰, and self-reported mood instability³¹. *Ptprd*-deficient mice show learning deficits and altered long-term potentiation magnitudes in hippocampal synapses³². *PTPRD* is expressed highly in the brain compared to non-brain tissues, especially in myelinating axons and growth cones^{33,34} in the prenatal cerebellum³⁵. The presynaptically located *PTPRD* is involved in axon outgrowth and guidance³⁶ and interacts with postsynaptic proteins such as Slitrk-2, interleukin-1 receptor, and Trk to mediate synapse adhesion and organization in mice^{37,38} and the development of excitatory and inhibitory synapses³⁹. Members of the Slitrk and interleukin protein families have been associated with OC behaviors in humans and mice^{40,41}.

Our results show that OC traits in the community share genetic risk with OCD. Polygenic risk for OC traits was associated with OCD case/control status and vice versa. OC traits and OCD case/control status also were substantially, but not significantly, genetically correlated. This estimate is higher than reported in a recent study ($r_g = 0.42$, $p = 0.095$; 50)⁴³. Lack of power is the most likely explanation for the absence of a significant result. Previous studies of other psychiatric disorders reported shared genetic risk between traits and diagnoses, with

polygenic risk and genetic correlations similar to what we report for OC traits and OCD case/control status^{25,31,44}. The shared genetic risk between OC traits and OCD supports the hypothesis that an OCD diagnosis could represent the high extreme of OC traits that are widely distributed in the general population. One implication of this finding is that population-based samples with quantitative trait measures can serve as a powerful complementary approach to case/control studies to accelerate gene discovery in psychiatric genetics.

SNP-based heritability for OC traits in the current sample was not significant in line with previous studies. Previous research reports lower SNP-based heritability for self-reported OC symptoms (0.058)⁴² than for clinical OCD (0.28–0.37)^{6,45}. A similar trend for lower SNP heritability in traits vs. diagnosis has been observed for ADHD^{25,46}. The reason for the disparity in SNP heritability between traits and diagnosis is unclear as there are several differences that may play a role including informant (parent/self vs. teacher or clinician)⁴⁶, type of measurement (categorical vs. quantitative), consideration of impairment, and timing (cross-sectional vs. lifetime symptoms). Regardless of a non-significant SNP heritability for OC traits from our sample, we still identified and validated a genome-wide significant variant.

The TOCS scale is similar to existing OC trait/symptom measures in item content but is unlike existing scales in that it measures OC traits from ‘strengths’ to ‘weaknesses’. As a result, the distribution of the total score is closer to a normal distribution than the j-shaped distributions typically observed with most symptom-based scales that rate behaviors from zero to a positive integer¹⁰ (e.g., not at all to quite a lot). Our results indicate that the distribution of the OC trait measure impacts power to identify genome-wide significant associations. A ‘strengths’ to ‘weaknesses’ measure identified a genome-wide significant association. However, when we collapsed the ‘strength’ end of the TOCS distribution to zero, the significance of this variant was substantially reduced to below genome-wide significance, although the effect was in the same direction. The same effect was observed using another OC measure that generates a j-shaped distribution: CBCL-OCS. One implication of our results is that there is genetic information in the ‘strengths’ end of the distribution captured by the TOCS. This information would be lost in scales that only measure ‘weaknesses’, particularly in community samples where the prevalence of clinically significant OC symptoms is relatively low. Trait-based scales that capture ‘strengths’ and ‘weaknesses’ and have a less skewed distribution could improve power to identify genome-wide hits and variants associated with disorders, especially in population samples.

The results of this study should be considered in light of its limitations. Although our sample was large enough to

detect a genome-wide significant locus that was also significant in meta-analyzed OCD case/control cohorts, substantially larger samples will be needed to identify most of the contributing common variants. The current version of the TOCS measures OC traits cross-sectionally, which does not account for symptom waxing and waning and does not measure impairment directly. However, our polygenic risk and genetic correlation analyses show that OC traits and OCD share genetic risk, suggesting that the TOCS is capturing traits that are likely to be on a continuum with OCD.

Conclusions

We identified the first genome-wide significant variant for OC traits that was also associated with OCD case status. Power to detect a genome-wide association was impacted by the distribution of the OC trait measure. OC traits and OCD share genetic risks supporting the hypothesis that OCD represents the extreme end of widely distributed OC traits in the population. Trait-based approaches in community samples using measures that capture the whole distribution of traits is a powerful and rapid complement to case/control GWAS designs to help drive genetic discovery in psychiatry.

Acknowledgements

We thank the families and children that participated in this study and the Ontario Science Centre for their collaboration and support. This study was funded by the Canadian Institutes of Health Research (P.D.A.: MOP-106573; R.J.S.: MOP-93696). P.D.A. is supported by the Alberta Innovates Translational Health Chair in Child and Youth Mental Health. The OCD Collaborative Genetics Association Study (OC GAS) was funded by the following NIMH Grant Numbers: MH071507 (G.N.), MH079489 (D.A.G.), MH079487 (J.M.), MH079488 (A.F.), and MH079494 (J.K.). Y.Y.S. and W.G. were also supported by the Intramural Research Program of the NIMH (MH002930-06). The IOCDF-GC was supported by a grant from the David Judah Foundation (a private, nonindustry-related foundation established by a family affected by OCD), MH079489 (D.L.P.), MH073250 (D.L.P.), S40024 (J.M.S.), MH 085057 (J.M.S.), and MH087748 (C.A.M.). Support for the collection of the data for Philadelphia Neurodevelopment Cohort (PNC) was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. All subjects were recruited through the Center for Applied Genomics at The Children’s Hospital in Philadelphia. The Michigan/Toronto sample was supported by the Ontario Brain Institute – Province of Ontario Neurodevelopmental Disorders (POND) Network (grant number: IDP-PND-2018; awarded to R.J.S., J.C., and P.D.A.) and the National Institutes of Mental Health (MH101493 awarded to GH and PDA; MH085321 awarded to G.H., D.R., and P.D.A.). The authors wish to thank Tara Paton, Chao Lu, and Jo-Anne Herbrick of The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada for assistance with genotyping. We gratefully acknowledge all the studies and databases that made GWAS summary data available: ADIPOGen (Adiponectin genetics consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDIOGRAM (Coronary ARtery Disease Genome wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), dbGAP (database of Genotypes and Phenotypes), DIAGRAM (DIAbetes Genetics Replication And Meta-analysis), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), EAGLE (EARly Genetics & Lifecourse Epidemiology Eczema Consortium, excluding 23andMe), EGG (Early Growth Genetics Consortium), GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOG (Genetic Factors for Osteoporosis Consortium), GIANT (Genetic Investigation of Anthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics

Consortium), GPC (Genetics of Personality Consortium), GUGC (Global Urate and Gout consortium), HaemGen (haematological and platelet traits genetics consortium), HRgene (Heart Rate consortium), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetic Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MESA (Multi-Ethnic Study of Atherosclerosis), PGC (Psychiatric Genomics Consortium), Project MinE consortium, ReproGen (Reproductive Genetics Consortium), SSGAC (Social Science Genetics Association Consortium), TAG (Tobacco and Genetics Consortium), TRICL (Transdisciplinary Research in Cancer of the Lung consortium), and UK Biobank. We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

Author details

¹Neurosciences and Mental Health, Toronto, Canada. ²Genetics and Genome Biology Hospital for Sick Children, Toronto, Canada. ³Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands. ⁴Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands. ⁵The Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. ⁶The Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁷The Mathison Centre for Mental Health Research and Education, Hotchkiss Brain Institute, Calgary, Canada. ⁸Departments of Psychiatry and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Canada. ⁹Institute of Medical Science, University of Toronto, Toronto, Canada. ¹⁰Department of Psychiatry and Behavioural Neurosciences, McMaster University, Ontario, Canada. ¹¹Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA. ¹²Department of Psychiatry and Behavioural Neurosciences, Wayne State University, Detroit, MI, USA. ¹³Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, USA. ¹⁴Divisions of Epidemiology and Biostatistics, Dalla Lana School of Public Health, Toronto, Canada. ¹⁵Department of Statistical Sciences, Faculty of Arts and Science, Toronto, Canada. ¹⁶Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, Canada

Conflict of interest

R.J.S. is the TD Bank Group Chair in Child and Adolescent Psychiatry, owns equity in Ehave, and has consulted for Highland Therapeutics, Purdue Pharma, E Lilly Corp, and Ehave. All other authors declare that they have no conflict of interest.

Publisher's note

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

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41398-020-01121-9>).

Received: 12 May 2020 Revised: 6 November 2020 Accepted: 24 November 2020

Published online: 02 February 2021

References

- Ruscio, A. M., Stein, D. J., Chiu, W. T. & Kessler, R. C. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Mol. Psychiatry* **15**, 53–63 (2010).
- Zohar, A. The epidemiology of obsessive-compulsive disorder in children and adolescents. *Child Adolesc. Psychiatr. Clin. N. Am.* **8**, 445–460 (1999).
- Van Grootheest, D. S., Cath, D. C., Beekman, A. T. & Boomsma, D. I. Twin studies on obsessive-compulsive disorder: a review. *Twin Res Hum. Genet.* **8**, 450–458 (2005).
- Mattheisen, M. et al. Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. *Mol. Psychiatry*. <https://doi.org/10.1038/mp.2014.43> (2014).
- Stewart, S. E. et al. Genome-wide association study of obsessive-compulsive disorder. *Mol. Psychiatry* **18**, 788–798 (2013).
- International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
- Den Braber, A. et al. Obsessive-compulsive symptoms in a large population-based twin-family sample are predicted by clinically based polygenic scores and by genome-wide SNPs. *Transl. Psychiatry* **6**, 1–7 (2016).
- Park, L. S. et al. The Toronto Obsessive-Compulsive Scale: psychometrics of a dimensional measure of obsessive-compulsive traits. *J. Am. Acad. Child Adolesc. Psychiatry* **55**, 310–318 (2016).
- Crosbie, J. et al. Response inhibition and ADHD traits: correlates and heritability in a community sample. *J. Abnorm Child Psychol.* **41**, 497–507 (2013).
- Burton, C. L. et al. Heritability of obsessive-compulsive trait dimensions in youth from the general population. *Transl Psychiatry* **8**, <https://doi.org/10.1038/s41398-018-0249-9> (2018).
- van der Sluis, S., Posthuma, D., Nivard, M. G., Verhage, M. & Dolan, C. V. Power in GWAS: lifting the curse of the clinical cut-off. *Mol. Psychiatry* **18**, 2–3 (2013).
- Burns, G. L., Fornea, G. M., Keortge, S. & Sternberger, L. G. The utilization of nonpatient samples in the study of obsessive compulsive disorder. *Behav. Res. Ther.* **33**, 133–144 (1995).
- Hudziak, J. J. et al. The Obsessive Compulsive Scale of the Child Behavior Checklist predicts obsessive-compulsive disorder: a receiver operating characteristic curve analysis. *J. Child Psychol. Psychiatry Allied Discip.* **47**, 160–166 (2006).
- Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1–10 (2017).
- Gong, J. et al. Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. *PLoS Genet.* **15**, <https://doi.org/10.1371/journal.pgen.1008007> (2019).
- Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- Bulik-Sullivan, B. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- Bulik-sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2016).
- Calkins, M. E. et al. The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *J. Child Psychol. Psychiatry* **56**, 1356–1369 (2015).
- Gazzellone, M. J. et al. Uncovering obsessive-compulsive disorder risk genes in a pediatric cohort by high-resolution analysis of copy number variation. *J. Neurodev. Disord.* **8**, <https://doi.org/10.1186/s11689-016-9170-9> (2016).
- Demontis, D. et al. Discovery of the first genome-wide significant risk loci for ADHD. <https://doi.org/10.1101/145581> (2017).
- Vilhjálmsdóttir, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
- Lonsdale, J. et al. The Genotype-Tissue Expression (GTEx) project. <https://doi.org/10.1038/ng.2653> (2013).
- Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
- Stewart, S. E. et al. Meta-analysis of association between obsessive-compulsive disorder and the 3' region of neuronal glutamate transporter gene SLC1A1. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **162B**, 367–379 (2013).
- Willour, V. L. et al. Replication study supports evidence for linkage to 9p24 in obsessive-compulsive disorder. *Am. J. Hum. Genet.* **75**, 508–513 (2004).
- Elia, J. et al. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol. Psychiatry* **15**, 637–646 (2010).
- Liu, X. et al. Genome-wide association study of autism spectrum disorder in the East Asian populations. *Autism Res.* **9**, 340–349 (2016).
- Schormair, B. et al. Identification of novel risk loci for restless legs syndrome in genome-wide association studies in individuals of European ancestry: a meta-analysis. *Lancet Neurol.* **16**, 898–907 (2017).
- Ward, J. et al. The genomic basis of mood instability: identification of 46 loci in 363,705 UK Biobank participants, genetic correlation with psychiatric disorders,

- and association with gene expression and function. *Mol. Psychiatry*, <https://doi.org/10.1038/s41380-019-0439-8> (2019).
32. Uetani, N. et al. Impaired learning with enhanced hippocampal long-term potentiation in PTP δ -deficient mice. *EMBO J.* **19**, 2775–2785 (2000).
 33. Zhu, Q. et al. Developmental expression and function analysis of protein tyrosine phosphatase receptor type D in oligodendrocyte myelination. *Neuroscience* **308**, 106–114 (2015).
 34. Gershon, T. R., Baker, M. W., Nitabach, M. & Macagno, E. R. The leech receptor protein tyrosine phosphatase HmlAR2 is concentrated in growth cones and is involved in process outgrowth. *Development* **125**, 1183–1190 (1998).
 35. Ohi, K. et al. Spatial and temporal expression patterns of genes around nine neuroticism-associated loci. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **77**, 164–171 (2017).
 36. Johnson, K. G., Van & Vactor, D. Receptor protein tyrosine phosphatases in nervous system development. *Physiol. Rev.* **83**, 1–24 (2003).
 37. Yoshida, T. et al. IL-1 receptor accessory protein-like 1 associated with mental retardation and autism mediates synapse formation by trans-synaptic interaction with protein tyrosine phosphatase δ . *J. Neurosci.* **31**, 13485–13499 (2011).
 38. Yamagata, A. et al. Structure of Slitrk2-PTP δ complex reveals mechanisms for splicing-dependent trans-synaptic adhesion. *Sci. Rep.* **5**, 9686 (2015).
 39. Takahashi, H. & Craig, A. M. Protein tyrosine phosphatases PTP δ , PTPC, and LAR: presynaptic hubs for synapse organization. *Trends Neurosci.* **36**, 522–534 (2013).
 40. Gray, S. M. & Bloch, M. H. Systematic review of proinflammatory cytokines in obsessive-compulsive disorder. *Curr. Psychiatry Rep.* **14**, 220–228 (2012).
 41. Ozomaro, U. et al. Characterization of SLITRK1 variation in obsessive-compulsive disorder. *PLoS ONE* **8**, <https://doi.org/10.1371/journal.pone.0070376> (2013).
 42. Smit, D. J. A. et al. Genetic meta-analysis of obsessive-compulsive disorder and self-report compulsive symptoms. 2019.
 43. Smit, D. J. A. et al. Genetic meta-analysis of obsessive-compulsive disorder and self-report compulsive symptoms. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet* **183**, 208–216 (2020).
 44. Robinson, E. B. et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
 45. Davis, L. K. et al. Partitioning the heritability of tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. *PLoS Genet.* **9**, <https://doi.org/10.1371/journal.pgen.1003864> (2013).
 46. Middeldorp, C. M. et al. A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *J. Am. Acad. Child Adolesc. Psychiatry* **55**, 896–905 (2016).

Author details

OCD Working Group of the Psychiatric Genomics Consortium

Paul D. Arnold^{2,7,8,16}, Kathleen D. Askland^{16,17}, Cristina Barlassina¹⁸, O. Joseph Bienvenu¹³, Donald Black¹⁹, Michael Bloch²⁰, Helena Brentani²¹, Christie L. Burton¹, Beatriz Camarena²², Carolina Cappi²³, Danielle Cath²⁴, M. Cristina Cavallini²⁵, Valentina Ciullo²⁶, David Conti²⁷, Edwin H. Cook²⁸, Vladimir Coric²⁹, Bernadette A. Cullen¹³, Danielle Cusi³⁰, Lea K. Davis³¹, Richard Delorme³², Damiaan Denys³³, Eske Derks³⁴, Valsamma Eapen³⁵, Christopher Edlund²⁷, Lauren Erdman², Peter Falkai³⁶, Abby J. Fyer³⁷, Daniel A. Geller^{5,38}, Fernando S. Goes¹³, Hans J. Grabe³⁹, Marco A. Grados¹³, Benjamin D. Greenberg⁴⁰, Edna Grünblatt⁴¹, Wei Guo⁴², Gregory L. Hanna¹¹, Ana G. Hounie²¹, Michael Jenike^{38,43,44}, Clare L. Keenan⁴⁵, James L. Kennedy^{9,16,46}, Ekaterina A. Khramtsova⁴⁷, James A. Knowles⁴⁸, Janice Krasnow¹³, Cristoph Lange⁴⁹, Nuria Lanzagorta⁵⁰, Marion Leboyer⁵¹, Kung-Yee Liang¹³, Christine Lochner⁵², Fabio Macciardi⁵³, Brion Maher⁵⁴, Carol A. Mathews⁵⁵, Manuel Mattheisen^{56,57,58}, James T. McCracken⁵⁹, Nathaniel McGregor⁵², Nicole C. R. McLaughlin⁴⁰, Euripedes c. Miguel²¹, Benjamin Neale^{5,6,38}, Gerald Nestadt¹³, Paul S. Nestadt¹³, Humberto Nicolini^{50,60}, Erika L. Nurmi⁵⁹, Lisa Osiecki^{38,43}, John Piacentini⁵⁹, Christopher Pittenger²⁹, Danielle Posthuma⁶¹, Ann E. Pulver¹³, Steven A. Rasmussen⁴⁰, Scott Rauch⁴⁴, Margaret A. Richter^{9,16,62}, Mark A. Riddle¹³, Stephan Ripke^{6,43,63}, Stephan Ruhrmann⁶⁴, Aline S. Sampaio⁶⁵, Jack F. Samuels¹³, Jeremiah M. Scharf^{5,6,38}, Yin Yao Shugart⁴², Jan H. Smit⁶⁶, Dan J. Stein⁶⁷, S. Evelyn Stewart⁶⁸, Maurizio Turiel¹⁸, Homero Vallada²¹, Jeremy Veenstra-VanderWeele^{37,69}, Nienke Vulink³³, Michael Wagner⁷⁰, Susanne Walitza⁴¹, Ying Wang⁵⁴, Jens Wendland⁷¹ and Dongmei Yu^{5,6} and Gwyneth Zai^{16,46}

¹⁷Waypoint Centre for Mental Health Care, Penetanguishene, ON, Canada. ¹⁸University of Milan, Milano, Italy. ¹⁹Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA, USA. ²⁰Yale Child Study Centre, New Haven, CT, USA. ²¹Faculdade de Medicina da Universidade de São Paulo (FMUSP), São Paulo, Brazil. ²²Instituto Nacional de Psiquiatria Ramón de la Fuente Muñiz, Mexico City, Mexico. ²³Icahn School of Medicine at Mount Sinai, New York, NY, USA. ²⁴Utrecht University, Utrecht, The Netherlands. ²⁵Ospedale San Raffaele Milano, Milano, Italy. ²⁶IRCSS Santa Lucia Foundation, Rome, Italy. ²⁷University of Southern California Keck School of Medicine, Los Angeles, CA, USA. ²⁸University of Illinois at Chicago, Chicago, IL, USA. ²⁹Yale University, New Haven, CT, USA. ³⁰Bio4Dreams, Milano, Italy. ³¹Vanderbilt University, Nashville, TN, USA. ³²AP-HP, Robert Debré Hospital, Paris, France. ³³Amsterdam UMC, location AMC, Amsterdam, The Netherlands. ³⁴QIMR Berghofer Medical Research Institute, Brisbane, Australia. ³⁵University of New South Wales, Sydney, Australia. ³⁶Ludwig-Maximilians-University Munich, Munich, Germany. ³⁷Columbia University Medical Centre, New York, NY, USA. ³⁸Harvard Medical School, Boston, MA, USA. ³⁹University Medicine Greifswald, Greifswald, Germany. ⁴⁰Alpert Medical School of Brown University, Providence, RI, USA. ⁴¹University Hospital of Psychiatry Zürich, University of Zürich, Zürich Centre for Integrative Human Physiology, Neuroscience Center Zürich, Zürich, Switzerland. ⁴²National Institute of Mental Health, Bethesda, MD, USA. ⁴³Massachusetts General Hospital, Boston, MA, USA. ⁴⁴McLean Hospital, Belmont, MA, USA. ⁴⁵University of Chicago, Chicago, IL, USA. ⁴⁶Centre for Addiction and Mental Health, Toronto, ON, Canada. ⁴⁷Janssen Pharmaceuticals, Spring House, PA, USA. ⁴⁸State University of New York, Downstate Health Sciences University, Brooklyn, NY, USA. ⁴⁹Harvard T.H. Chan School of Public Health, Boston, MA, USA. ⁵⁰Grupo Médico Carracci, Mexico City, Mexico. ⁵¹Hôpitaux universitaires Henri-Mondor, Créteil, France. ⁵²Stellenbosch University, Stellenbosch, South Africa. ⁵³University of California, Irvine, Irvine, CA, USA. ⁵⁴Johns Hopkins University, Baltimore, MD, USA. ⁵⁵University of Florida College of Medicine, Gainesville, FL, USA. ⁵⁶Aarhus University, Aarhus, Denmark. ⁵⁷Karolinska, Stockholm, Sweden. ⁵⁸Dalhousie University,

Nova Scotia, NS, Canada. ⁵⁹UCLA Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, USA. ⁶⁰Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, Mexico. ⁶¹VU University Amsterdam, Amsterdam, The Netherlands. ⁶²Sunnybrook Health Sciences Centre, Toronto, ON, Canada. ⁶³Charité - Universitätsmedizin, Berlin, Germany. ⁶⁴University of Cologne, Cologne, Germany. ⁶⁵Federal University of Bahia, Salvador, Brazil. ⁶⁶Amsterdam University Medical Center, Amsterdam, The Netherlands. ⁶⁷SA MRC Unit on Risk & Resilience in Mental Disorders, University of Cape Town, Cape Town, South Africa. ⁶⁸University of British Columbia, BC Children's Hospital Research; BC Mental Health and Substance Use Research, Vancouver, BC, Canada. ⁶⁹New York State Psychiatric Institute, New York, NY, USA. ⁷⁰University Hospital Bonn, Bonn, Germany. ⁷¹Takeda Pharmaceuticals, Cambridge, MA, USA