Genome-wide Association Study of Pediatric Obsessive-Compulsive Traits: Shared Genetic Risk between Traits and Disorder

Christie L. Burton, PhD[1], Mathieu Lemire, PhD[1], Bowei Xiao, MSc[2], Elizabeth C. Corfield, PhD[1], Lauren Erdman, MSc[2], Janita Bralten PhD[3], Geert Poelmans MD, PhD [3], Dongmei Yu, MSc[4,5], S-M Shaheen, MSc, MBA[6,7], Tara Goodale, MSc[1], OCD Working Group of the Psychiatric Genomics Consortium, Noam Soreni, MD[8], Gregory L. Hanna, MD[9], Kate D. Fitzgerald, MD[9], David Rosenberg, MD[10,11], Gerry Nestadt, MBBCh, MPH[12], Andrew D. Paterson, MD[2,13], Lisa Strug, PhD[2,14], Russell J. Schachar, MD[1,15], Jennifer Crosbie, PhD[1,15], Paul D. Arnold, PhD, MD[2,6,7]

[1]Neurosciences and Mental Health, [2]Genetics and Genome Biology Hospital for Sick Children, Toronto, Canada

[3]Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands

[4]The Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

[5]The Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[6]Mathison Centre for Mental Health Research and Education, [7]Departments of Psychiatry & Medical Genetics, Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, Canada

[8]Department of Psychiatry and Behavioural Neurosciences, McMaster University

[9]Department of Psychiatry, University of Michigan, Ann Arbor, USA

[10]Department of Psychiatry and Behavioural Neurosciences, Wayne State University, Detroit, USA

[11]The Children's Hospital of Michigan, Detroit, USA

[12]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University

[13]Divisions of Epidemiology and Biostatistics, Dalla Lana School of Public Health, [14]Department of Statistical Sciences, Faculty of Arts and Science, [15]Department of Psychiatry, Faculty of Medicine, University of Toronto, Canada

Corresponding Author*:
Christie L. Burton, PhD
Research Associate
The Hospital for Sick Children (Neurosciences and Mental Health/Psychiatry)
555 University Ave, 4289B, Toronto, ON M5G 1X8
Phone: 416.813.7654 ex 202346/Fax: 416.813.6565
Email: christie.burton@sickkids.ca

Abstract

This study examined the genetic correlates of obsessive-compulsive (OC) traits and their shared genetic risks with obsessive-compulsive disorder (OCD). We conducted genome-wide association analyses on OC traits in 5018 unrelated Caucasian children and adolescents. Overall OC traits and trait dimensions (e.g., cleaning/contamination) were measured with the Toronto Obsessive-Compulsive scale (TOCS). One locus tagged by rs7856850 in an intron of *PTPRD* (protein tyrosine phosphatase δ) was associated with OC traits at the genome-wide significance level (p=2.48x10^{-8}). A variant in *GRID2* was significantly associated with only the symmetry/ordering dimension (p=3.2x10^{-8}). We tested the role of central nervous system (CNS) and glutamate gene-sets using hypothesis-driven methods. A stratified False Discovery Rate found OC traits were associated with SNPs in three CNS genes: *NPAS2* (p=7.8x10^{-7}), *GRID2* (p=1.6x10^{-6}) and *SH3GL2* (p=1.9x10^{-7}). The combined effect of neither the CNS development nor the glutamate gene-set were associated with OC traits using the competitive gene-set test implemented with MAGMA. We replicated the SNP in *PTPRD* in a meta-analysis of three independent OCD case/control genome-wide datasets (p=0.0069, cases=3384, controls=8363). Polygenic risk from OC traits was significantly associated with OCD in a sample of childhood-onset OCD and vice versa (p's<0.01). OC traits were highly but not significantly correlated with OCD (r_g=0.83, p=0.07). We report the first replicated genome-wide significant variant for OCD traits. Our results indicate that OC traits in the general population share genetic risk with OCD in independent samples. This study demonstrates the feasibility and power of using trait-based approaches in community samples in psychiatric genomics.

Obsessive-compulsive disorder (OCD) is a common (1-2% prevalence [1]) psychiatric disorder characterized by intrusive, recurrent thoughts and repeated, ritualized behaviors. OCD symptoms cluster into several distinct dimensions (e.g., cleaning/contamination [2]). Childhood-onset OCD (before the age of 18) occurs in 30-50% of cases [3] and is more heritable compared to adult-onset OCD [4]. Two genome-wide association studies (GWAS) in clinical populations with mixed ages of OCD-onset and a meta-analysis of these studies did not identify genome-wide significant findings [5–7]. Top hits from previous GWAS include SNPs within *DLGAP1*, *BTBD3, GRID2* and close to *PTPRD*. These genes have been functionally linked to glutamate neurotransmission and neurodevelopment, which are recurrent themes in the OCD genetic and imaging literature [8]. A previous study of obsessive-compulsive (OC) symptoms in a community-based sample of adult twins identified a genome-wide significant SNP in *MEF2B* (rs8100480) [9]. However, this SNP was not replicated in an independent sample [6]. OCD symptom dimensions have some shared but some distinct genetic risks [10, 11], however, there have been no GWAS on OCD dimensions to date.

We conducted a GWAS of quantitative OC traits, and secondarily OC trait dimensions, using the Toronto Obsessive-Compulsive Scale (TOCS [12]) in a large pediatric, community-based sample: Spit for Science [13]. TOCS scores are heritable, factor into six commonly-observed OCD symptom dimensions [11] and include 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). The strength-to-weakness format generates scores with a more normal distribution than observed with existing OCD scales. The latter generate J-shaped distributions, especially in community-based samples where the prevalence of OC symptoms is low [14, 15]. We checked if the distribution of TOCS scores would affect the power of genetic discovery [16] by collapsing the strengths/negative scores into scores of zero, thereby replicating a J-shaped distribution. We used the collapsed TOCS measure in a secondary GWAS with another OCD symptom measure: Child Behavior Checklist –

Obsessive-Compulsive Scale (CBCL-OCS [17]). To further understand the biology of OC traits, we used hypothesis-driven genome-wide approaches to test the role of genetic variants annotated to genes implicated in two leading biological hypotheses for OCD: brain development and glutamate function [18]. Genome-wide significant variants were tested for association with brain expression quantitative trait loci (eQTLs). We calculated SNP-based heritability of total OC trait scores and, secondarily, trait dimension scores to estimate the contribution of common genetic factors. We also examined the genetic correlation of total OC trait scores with other medical/mental health disorders and traits. 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 and examined if the top hits from the previous GWAS of OC symptoms [9] replicated in our study.

Materials and Methods

Discovery

Participants

The Spit for Science sample is described in detail elsewhere [13]. 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 Hospital for Sick Children Research Ethics Board. 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 [11, 12]. 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 and OC dimensions scores are described in the supplement. We tested the impact of the strength/weakness structure of the TOCS by comparing it to an OCD symptom measure with a j-shaped distribution - CBCL-OCS [17]. We also re-scored the TOCS by collapsing all strengths (negative scores) into zero. This 'collapsed' TOCS total score created a distribution similar to the CBCL-OCS (see supplement and Supplemental Figure S1 for details).

Genetic Data

DNA was extracted manually from saliva using standard methods (see the supplement for additional details). We excluded any samples with concentrations <60ng/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 1 140 449 markers on the Omni array.

Quality control (QC) was conducted separately for each array using standard methods with PLINK v1.90 [19]. Sample exclusion and selection criteria are described in the supplemental methods and Supplemental Figure S2. Imputation was performed separately for all platforms and sample sets, using Beagle v4.1 [20, 21] 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).

Analyses

GWAS were 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>0.3) and used the standard genome-wide threshold of $p \leq 5x10^{-8}$. In the supplement, we describe secondary analyses including GWAS with non-standardized TOCS total scores, gene-based GWAS, analyses for CBCL-OCS, collapsed TOCS total score and the six TOCS OC dimensions. Using hypothesis-driven methods, we conducted GWAS that prioritized SNPs within two gene-sets involved in brain/central nervous system (CNS) development and glutamate receptors and transporters selected based on previous literature [8](see supplement/Supplemental Table S1 for gene-set details). We used a stratified False Discovery Rate (sFDR - [22]) to test the significance of individual SNPs. In separate analyses for each gene-set, all the SNPs in the gene-set were assigned to high priority groups and the remaining SNPs were assigned to low priority groups and FDR was controlled separately in the high and low priority groups. We then tested the association of each gene-set collectively using the MAGMA competitive gene-set test [23] with a Bonferroni correction to account for testing two hypotheses (α=0.025).

We tested each genome-wide significant variant for co-localization with brain eQTLs using LocusFocus (https://locusfocus.research.sickkids.ca/ [24]; see supplement).

We estimated SNP heritability using both GCTA [25] v1.91.2-beta (http://cnsgenomics.com/software/gcta/) with further exclusion of cousins and SNPs with AR2>0.9 and LDSC (v1.0.0, https://github.com/bulik/ldsc [26]) calculated from SNPs in HapMap3. We secondarily

examined the heritability of the six TOCS dimensions. We used LDSC [27] to examine the genetic correlation of TOCS total scores with the 850 phenotypes available on LD Hub (http://ldsc.broadinstitute.org/ldhub/).

Replication

Participants

For replication analyses, we investigated three independent OCD case/control cohorts: 1) the International OCD Foundation Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS) meta-analysis [7], 2) the Philadelphia Neurodevelopmental Cohort (PNC) from the Children's Hospital of Philadelphia (CHOP [28]) and 3) the Michigan/Toronto OCD Imaging Genomics Study [29]. See the supplement and Table 1 for sample sizes.

Analyses

GWAS summary stats from each replication sample were meta-analyzed using fixed-effect inverse variance methods. We tested if the results from the gene-based GWAS and hypothesis-driven methods for the TOCS total score replicated in the OCD samples by conducting the same genome-wide analyses as described above or in the supplement. Polygenic risk score (PRS) analyses were performed using LDpred v1.06 ([30]; 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 replication cohorts (target sample: CHOP, Michigan/Toronto and a subset of the IOCDF-GC/OCGAS - see supplement). Second, we derived PRS from the combined OCD replication 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 replication samples using genetic correlations estimated with LDSC [27]. Finally, we examined the top

variants in the only previous GWAS of OC symptoms (only 20 loci reported [9]) in the results from the TOCS total score GWAS.

Results

Discovery

We used 5018 participants for GWAS analyses after sample exclusion and selection (see supplement and Supplemental Figures S2/S3). In the primary analysis, rs7856850 in *PTPRD* was significantly associated with TOCS total scores at the genome-wide level (p=2.48x10^{-8}, β=0.14, s.e.=0.025, R^2=0.618%: Figure 1a, top hits listed in Supplemental Table S2). 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 (Figure 1B). 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; Figure 1C). eQTL results for this and all genome-wide significant SNPs are presented in the supplement. In the secondary analyses, the results for rs7856850 were similar for the non-standardized TOCS total score (see supplement. However, 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 still had the same direction of effect (p=0.00045 and p=0.025 respectively; see supplement for details). Only one TOCS OC dimension, symmetry/ordering, yielded a genome-wide significant association (rs5860287 in *GRID2*, p=3.2x10^{-8}; β=0.118; s.e.=0.0213; R^2=0.610%; Supplemental Figure S4).

A gene-based GWAS by MAGMA using FUMA identified four genome-wide significant genes: *SH3GL2* (p=1.53x10^{-7}); *PDXDC1* (p=3.95x10^{-7}); *RIMBP2* (p=1.513x10^{-6}) and *RRN3* (p=1.56x10^{-6}; see supplement for details and Supplemental Figure S5).

Results from the sFDR are shown in Supplemental Table S3 and genes included in the analysis are reported in the supplement and Supplemental Table S1. Among the CNS development gene-set, sFDR revealed three loci associated with TOCS total score: rs3754673 (*NPAS2;*q_{SFDR}=0.0123, R^2=0.486%), rs5860287 (*GRID2:*q_{SFDR}=0.0165, R^2=0.458%) and rs2182089 (*SH3GL2:*q_{SFDR}=0.0096, R^2=0.540%; Figure 2, Supplemental Figure S6). rs2182089/*SH3GL2* is close to (9.2 Mbp), but not in LD with, rs7856850/*PTPRD*. rs2182089 is still associated with the TOCS total score when conditioning on rs7856850 (p=1.7x10^{-7}), showing that these loci are independent.

Among the glutamate gene-set, only SNPs in *GRID2* (a gene also annotated to the CNS gene-set) approached significance (Supplemental Figure S7, Table S3). When we compared CNS development and glutamate gene-sets to all other gene-sets in the genome with the MAGMA competitive gene-set test, neither set were significantly associated with TOCS total scores (*p's*>=0.33).

The heritability of the TOCS total score was h^2=0.068 (s.e.=0.052, p=0.19) using GCTA and h^2=0.073 (s.e.=0.064; p=0.25) using LDSC when the intercept was constrained to 1. None of the OC dimensions were significantly heritable (Supplemental Table S4). TOCS total score was not significantly associated with any phenotypes on LD Hub (see supplement).

Replication

Following standard QC and sample exclusion where applicable (see supplement), we had a total of 3369 cases and 8611 controls in our replication sample (Table 1). We tested if the genome-wide SNP associated with TOCS total scores in Spit for Science replicated in the meta-analyzed OCD 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], Figure 3, Supplemental Figure S8). A gene-based GWAS by MAGMA using FUMA did not identify any genome-wide significant genes. Therefore, none of the genes identified in the gene-based GWAS of OC traits were replicated (see supplement). We then tested if the 56 SNPs in

the CNS development list identified in the sFDR analysis were associated with OCD case/control status in the replication meta-analysis. None of these SNPs replicated even without a Bonferroni correction for multiple testing (see supplement).

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 replication samples (Nagelkerke's pseudo r^2=0.277%, p=0.0045 at ρ=0.003). Figure 4b shows that PRS constructed from the OCD replication sample were significantly associated with TOCS total scores in Spit for Science (r^2=0.24%; p=0.00057 at ρ=0.1).

The genetic correlation between standardized TOCS total scores and OCD meta-analysis was r_g=0.825 (s.e.=0.428; p=0.073) when intercepts are constrained to 1.

One of the top-ranked SNPs from a previous GWAS of OC symptoms [9] 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 top hit (9p24.1), but not in LD (r^2=0.004, D'=0.517). Another 16 of the reported top hits in den Braber [9], 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 S5).

Discussion

Using a trait-based approach in a community sample, we identified the first replicated genome-wide significant variant related to OCD (rs7856850). The hypothesis-driven approach showed that genetic variants related to CNS development, particularly in *NPAS2*, *SH3GL2* and *GRID2*, were associated with OC traits. A variant in *GRID2* was also significantly associated with the OCD dimension symmetry/ordering and four genes were significantly associated with OC traits; none of these findings were replicated. 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]). This variant replicated in a meta-analysis of three independent OCD cohorts making it the first variant associated with OC traits and OCD. The small size of the replication sample likely precluded finding genome-wide hits in the meta-analysis. 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 [5, 9]. 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 [31–33]. Rare CNVs in *PTPRD* have been identified in cases with OCD [29], ADHD [34] and with brain malformations at birth [35]. SNPs in *PTPRD* were genome-wise significantly associated with ASD [36], restless legs syndrome [37], and self-reported mood instability [38]. *Ptprd*-deficient mice show learning deficits and altered long-term potentiation magnitudes in hippocampal synapses [39]. *PTPRD* is expressed highly in the brain compared to non-brain tissues, especially in myelinating axons and growth cones [40–42] in the prenatal cerebellum [43]. The presynaptically located PTPRD is involved in axon outgrowth and guidance [44, 45] and interacts with postsynaptic proteins such as Slitrk-2, interleukin-1 receptor and TrK to mediate synapse adhesion and organization in mice [46–48] and the development of excitatory and inhibitory synapses [49]. Members of the Slitrk and interleukin protein families have been associated with OC behaviors in humans and mice [50–52].

SNPs in *GRID2* were significant in our sFDR analysis of CNS development genes and were associated with the symmetry/ordering dimension. *GRID2* codes for glutamate ionotropic receptor, δ2 (GluD2). *GRID2* is highly expressed in the testis and in the brain, particularly in brain regions commonly associated with OCD including the striatum, anterior cingulate cortex and cerebellum (GTEx V8 [53], https://gtexportal.org/home/). In the largest OCD meta-analysis, one of the top associated variants was

in *GRID2* (rs1030757, p=1.1x10^{-6}, OR=1.18 [54]). Additionally, this variant was associated with TOCS in the Spit for Science sample (β=0.09, s.e.=0.021, p=8.87x10^{-6}) and was in the same haplotype block as the genome-wide significant *GRID2* SNPs identified in our study (r^2=0.66-0.92) suggesting that these variants may be tagging the same locus. Neither rs1030757, nor the SNP in *GRID2* with the lowest p-value in the sFDR analysis (rs5860287), were associated with any brain eQTLS. Rare inherited CNVs in *GRID2* have been identified in ASD cases [55] and SNPs in *GRID2* have been associated with cognitive deficits in schizophrenia [56]. Glutamate has been strongly implicated in OCD [32, 57]. Although we did not identify any genome-wide significant variants from the glutamate gene list and the glutamate gene-set was not associated with OC traits, many of our significant variants from the CNS development list were linked to glutamate, including a glutamate receptor gene *GRID2* and *NPAS2*.

Within the CNS development gene-set, we identified two additional significant loci. One locus was in *NPAS2*, which codes for neuronal per-arnt-sim (PAS) domain protein 2. NPAS2 is a core transcriptional factor in the molecular clock that maintains circadian rhythms and is expressed exclusively in the brain in the first week of life [58]. Although not annotated to the glutamate gene-set, *NPAS2* regulates the uptake of glutamate into astrocytes [59], which maintains appropriate levels of extracellular glutamate [60]. The other locus was in *SH3GL2*, which codes for Endophilin A1. This gene was also significant in the gene-based GWAS of OC traits. Endophillin-A1 is involved in synaptic vesicle endocytosis in presynaptic terminals [61] and is required for dendrite development driven by brain-derived neurotrophic factor (BDNF) in mice [62]. Although none of the *NPAS2* or *SH3GL2* SNPs replicated in the OCD cohorts, their association with brain eQTLs in LocusFocus [24] provides support for these variants.

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, similar to a recent study (r_g=0.42, p=0.095 [63]). Lack of power is the most likely explanation for the absence of significant results. Previous

studies on other psychiatric disorders report 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 [38, 64, 65]. 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. Previous studies similarly report lower SNP-based heritability for self-reported OC symptoms (0.058; [63]) than for clinical OCD (0.28-.37; [54, 66]. A similar trend for lower trait vs. diagnosis SNP heritability has been observed for schizophrenia [67, 68] and ADHD [69, 70]. The reason for the disparity in SNP heritability between traits and diagnosis is unclear. One possible explanation could be differences in the informant as shown previously for ADHD [70]. Regardless of a non-significant SNP heritability for OC traits from our sample, we still identified and replicated a genome-wide significant variant.

OCD is a heterogeneous disorder with several accepted symptom dimensions. OCD dimensions have shared, but distinct, genetic variance in twin studies [10, 71, 72]. When we conducted a GWAS for each dimension separately, only the symmetry/ordering dimension had a genome-wide significant hit. This variant in *GRID2* was also significant in the sFDR for CNS development genes in the TOCS total score. Our results highlight the importance of including phenotypes beyond just diagnosis or overall OC traits/symptoms.

The type of OC trait measure we used may have increased our power to identify 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 integer [12]. The TOCS total score was associated with a genome-wide significant variant, which was not significant when we used two measures with j-shaped distributions (collapsed TOCS score and CBCL-OCS). Therefore, using trait-based scales that capture strengths and weaknesses and have a less skewed distribution could improve power, especially in population samples where the prevalence of clinically significant OC symptoms is relatively low.

Conclusions

We identified the first replicated genome-wide significant variant for OC traits and demonstrated the sharing of genetic risk between OC traits and OCD. This supports 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

This study was funded by the Canadian Institutes of Health Research (PDA: MOP-106573; RJS: MOP–93696). PDA is supported by the Alberta Innovates Translational Health Chair in Child and Youth Mental Health. The OCD Collaborative Genetics Association Study (OCGAS) was funded by the following NIMH Grant Numbers: MH071507 (GN), MH079489 (DAG), MH079487 (JM), MH079488 (AF) and MH079494 (JK). Yao Shugart and Wei Guo 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 (DLP), MH073250 (DLP), S40024 (JMS), MH 085057 (JMS), and MH087748 (CAM). 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 RJS, JC and PDA) and the National Institutes of Mental Health (MH101493 awarded to GH and PDA; MH085321 awarded to GH, DR and PDA). 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), GEFOS (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 (haemotological 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) and TAG (Tobacco and Genetics

Consortium), TRICL (Transdisciplinary Research in Cancer of the Lung consortium), 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).

Conflicts of Interest

Russell Schachar 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 have no conflicts.

References

1. Ruscio AM, Stein DJ, Chiu WT, Kessler RC. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. Mol Psychiatry. 2010;15:53–63.
2. Bloch MH, Landeros-Weisenberger A, Rosario MC, Pittenger C, Leckman JF. Meta-analysis of the symptom structure of obsessive-compulsive disorder. Am J Psychiatry. 2008;165:1532–1542.
3. Zohar A. The epidemiology of obsessive-compulsive disorder in children and adolescents. Child Adolesc Psychiatr Clin N Am. 1999;8:445–460.
4. Van Grootheest DS, Cath DC, Beekman AT, Boomsma DI. Twin studies on obsessive-compulsive disorder: A review. Twin Res Hum Genet. 2005;8:450–458.
5. Mattheisen M, Samuels JF, Wang Y, Greenberg BD, Fyer AJ, McCracken JT, et al. Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. Mol Psychiatry. 2014. 2014. https://doi.org/10.1038/mp.2014.43.
6. Stewart SE, Yu D, Scharf JM, Neale BM, Fagerness JA, Mathews CA, et al. Genome-wide association study of obsessive-compulsive disorder. Mol Psychiatry. 2013;18:788–798.
7. 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. 2018;23.
8. Pauls DL, Abramovitch A, Rauch SL, Geller DA. Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective. Nat Rev Neurosci. 2014;15:410–424.
9. den Braber A, Zilhão NR, Fedko IO, Hottenga JJ, Pool R, Smit DJA, 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. 2016;6:1–7.
10. Iervolino AC, Rijsdijk F V, Cherkas L, Fullana M a, Mataix-Cols D. A multivariate twin study of obsessive-compulsive symptom dimensions. Arch Gen Psychiatry. 2011;68:637–644.
11. Burton CL, Park LS, Corfield EC, Forget-Dubois N, Dupuis A, Sinopoli VM, et al. Heritability of obsessive–compulsive trait dimensions in youth from the general population. Transl Psychiatry.

2018;8.

12. Park LS, Burton CL, Dupuis A, Shan J, Storch EA, Crosbie J, et al. The Toronto Obsessive-Compulsive Scale: Psychometrics of a Dimensional Measure of Obsessive-Compulsive Traits. J Am Acad Child Adolesc Psychiatry. 2016;55.

13. Crosbie J, Arnold P, Paterson A, Swanson J, Dupuis A, Li X, et al. Response inhibition and ADHD traits: correlates and heritability in a community sample. J Abnorm Child Psychol. 2013;41:497–507.

14. Burns GL, Formea GM, Keortge S, Sternberger LG. The utilization of nonpatient samples in the study of obsessive compulsive disorder. Behav Res Ther. 1995;33:133–144.

15. Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, et al. Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. J Am Acad Child Adolesc Psychiatry. 1997;36:844–852.

16. van der Sluis S, Posthuma D, Nivard MG, Verhage M, Dolan C V. Power in GWAS: lifting the curse of the clinical cut-off. Mol Psychiatry. 2012;18:2–3.

17. Hudziak JJ, Althoff RR, Stanger C, van Beijsterveldt CEM, Nelson EC, Hanna GL, 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. 2006;47:160–166.

18. Pauls DL, Abramovitch A, Rauch SL, Geller DA. Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective. Nat Rev Neurosci. 2014;15:410–424.

19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–575.

20. Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. Am J Hum Genet. 2011;88:173–182.

21. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007;81:1084–1097.

22. Sun L, Rommens JM, Corvol H, Li W, Li X, Chiang TA, et al. Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. Nat Genet. 2012;44:562–569.

23. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol. 2015;11.

24. Gong J, Wang F, Xiao B, Panjwani N, Lin F, Keenan K, et al. Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. PLoS Genet. 2019;15.

25. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.

26. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD score regression

distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–295.

27. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P, et al. An Atlas of Genetic Correlations across Human Diseases and Traits. Nat Genet. 2016;47:1236–1241.

28. Calkins ME, Merikangas KR, Moore TM, Burstein M, Behr MA, Satterthwaite TD, et al. The Philadelphia Neurodevelopmental Cohort : constructing a deep phenotyping collaborative. J Child Psychol Psychiatry. 2015;56:1356–1369.

29. Gazzellone MJ, Zarrei M, Burton CL, Walker S, Uddin M, Shaheen SM, et al. Uncovering obsessive-compulsive disorder risk genes in a pediatric cohort by high-resolution analysis of copy number variation. J Neurodev Disord. 2016;8.

30. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. MoVilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., … Price, A. L. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. The American Journal of Human Genetics, 97(4), 576–592. https://doi. Am J Hum Genet. 2015;97:576–592.

31. Hanna GL, Veenstra-VanderWeele J, Cox NJ, Boehnke M, Himle JA, Curtis GC, et al. Genome-wide linkage analysis of families with obsessive-compulsive disorder ascertained through pediatric probands. Am J Med Genet - Neuropsychiatr Genet. 2002;114:541–552.

32. Stewart SE, Mayerfeld C, Arnold PD, Crane JR, O'Dushlaine C, Fagerness JA, 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. 2013;162B:367–379.

33. Willour VL, Yao Shugart Y, Samuels J, Grados M, Cullen B, Bienvenu OJ, et al. Replication study supports evidence for linkage to 9p24 in obsessive-compulsive disorder. Am J Hum Genet. 2004;75:508–513.

34. Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, et al. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Mol Psychiatry. 2010;15:637–646.

35. Schumann M, Hofmann A, Krutzke SK, Hilger AC, Marsch F, Stienen D, et al. Array-based molecular karyotyping in fetuses with isolated brain malformations identifies disease-causing CNVs. J Neurodev Disord. 2016;8.

36. Liu X, Shimada T, Otowa T, Wu YY, Kawamura Y, Tochigi M, et al. Genome-wide Association Study of Autism Spectrum Disorder in the East Asian Populations. Autism Res. 2016;9:340–349.

37. Schormair B, Zhao C, Bell S, Tilch E, Salminen A V, Pütz 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. 2017;16:898–907.

38. Ward J, Tunbridge EM, Sandor C, Lyall LM, Ferguson A, Strawbridge RJ, 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. 2019. 2019. https://doi.org/10.1038/s41380-019-0439-8.

39. Uetani N, Kato K, Ogura H, Mizuno K, Kawano K, Mikoshiba K, et al. 2000_EMBO J_Uetani et al_Impaired learning with enhanced LTP in PTPd KO.pdf. 2000;19.

40. Zhu Q, Tan Z, Zhao S, Huang H, Zhao X, Hu X, et al. Developmental expression and function analysis of protein tyrosine phosphatase receptor type D in oligodendrocyte myelination. Neuroscience. 2015;308:106–114.

41. Stoker AW. Isoforms of a novel cell adhesion molecule-like protein tyrosine phosphatase are implicated in neural development. Mech Dev. 1994;46:201–217.

42. Gershon TR, Baker MW, Nitabach M, Macagno ER. The leech receptor protein tyrosine phosphatase HmLAR2 is concentrated in growth cones and is involved in process outgrowth. Development. 1998;125:1183–1190.

43. Ohi K, Shimada T, Yasuyama T, Kimura K, Uehara T, Kawasaki Y. Spatial and temporal expression patterns of genes around nine neuroticism-associated loci. Prog Neuro-Psychopharmacology Biol Psychiatry. 2017;77:164–171.

44. Van Vactor D. Protein tyrosine phosphatases in the developing nervous system. Curr Opin Cell Biol. 1998;10:174–181.

45. Johnson KG, Van Vactor D. Receptor protein tyrosine phosphatases in nervous system development. Physiol Rev. 2003;83:1–24.

46. Yoshida T, Shiroshima T, Lee S-J, Yasumura M, Uemura T, Chen X, et al. Interleukin-1 receptor accessory protein organizes neuronal synaptogenesis as a cell adhesion molecule. J Neurosci. 2012;32:2588–2600.

47. Yoshida T, Yasumura M, Uemura T, Lee SJ, Ra M, Taguchi R, 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. 2011;31:13485–13499.

48. Yamagata A, Sato Y, Goto-Ito S, Uemura T, Maeda A, Shiroshima T, et al. Structure of Slitrk2-PTPδ complex reveals mechanisms for splicing-dependent trans-synaptic adhesion. Sci Rep. 2015;5:9686.

49. Takahashi H, Craig AM. Protein tyrosine phosphatases PTPδ, PTPς, and LAR: Presynaptic hubs for synapse organization. Trends Neurosci. 2013;36:522–534.

50. Gray SM, Bloch MH. Systematic review of proinflammatory cytokines in obsessive-compulsive disorder. Curr Psychiatry Rep. 2012;14:220–228.

51. Shmelkov S V, Hormigo A, Jing D, Proenca CC, Bath KG, Milde T, et al. Slitrk5 deficiency impairs corticostriatal circuitry and leads to obsessive-compulsive-like behaviors in mice. Nat Med. 2010;16:598–602, 1p following 602.

52. Ozomaro U, Cai G, Kajiwara Y, Yoon S, Makarov V, Delorme R, et al. Characterization of SLITRK1 Variation in Obsessive-Compulsive Disorder. PLoS One. 2013;8.

53. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. 2013.

54. Arnold PD, Askland KD, Barlassina C, Bellodi L, Bienvenu OJ, Black D, et al. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. Mol Psychiatry. 2018;23:1181–1188.

55. Gazzellone MJ, Zhou X, Lionel AC, Uddin M, Thiruvahindrapuram B, Liang S, et al. Copy number

variation in Han Chinese individuals with autism spectrum disorder. J Neurodev Disord. 2014;6:1–7.

56. Greenwood TA, Lazzeroni LC, Calkins ME, Freedman R, Green MF, Gur RE, et al. Genetic assessment of additional endophenotypes from the Consortium on the Genetics of Schizophrenia Family Study. Schizophr Res. 2016;170:30–40.

57. Rajendram R, Kronenberg S, Burton CL, Arnold PD. Glutamate Genetics in Obsessive-Compulsive Disorder: A review. J Can Acad Child Adolesc Psychiatry. 2017;26.

58. Zhou YD, Barnard M, Tian H, Li X, Ring HZ, Francke U, et al. Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system. Proc Natl Acad Sci U S A. 1997;94:713–718.

59. Beaulé C, Swanstrom A, Leone MJ, Herzog ED. Circadian modulation of gene expression, but not glutamate uptake, in mouse and rat cortical astrocytes. PLoS One. 2009;4.

60. Nedergaard M, Takano T, Hansen AJ. Beyond the role of glutamate as a neurotransmitter. Nat Rev Neurosci. 2002;3:748–755.

61. Reutens AT, Begley CG. Endophilin-1: A multifunctional protein. Int J Biochem Cell Biol. 2002;34:1173–1177.

62. Fu X, Yang Y, Xu C, Niu Y, Chen T, Zhou Q, et al. Retrolinkin cooperates with endophilin A1 to mediate BDNF-TrkB early endocytic trafficking and signaling from early endosomes. Mol Biol Cell. 2011;22:3684–3698.

63. Smit DJA, Cath D, Zilhão NR, Ip HF, Denys D, Den A, et al. Genetic meta-analysis of obsessive-compulsive disorder and self-report compulsive symptoms . 2019. 2019.

64. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.

65. Robinson EB, Anttila V, Kosmicki JA, Bulik-sullivan B, Grove J, Maller J, et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nat Genet. 2016;48:552–555.

66. Davis LK, Yu D, Keenan CL, Gamazon ER, Konkashbaev AI, Derks EM, et al. Partitioning the Heritability of Tourette Syndrome and Obsessive Compulsive Disorder Reveals Differences in Genetic Architecture. PLoS Genet. 2013;9.

67. Shi H, Kichaev G, Pasaniuc B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. Am J Hum Genet. 2016;99:139–153.

68. Ward J, Strawbridge RJ, Bailey MES, Graham N, Ferguson A, Lyall DM, et al. Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia. Transl Psychiatry. 2017;7.

69. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for ADHD. 2017. 2017.

70. Middeldorp CM, Hammerschlag AR, Ouwens KG, Groen-Blokhuis MM, St. Pourcain B, Greven CU, 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. 2016;55:896-905.e6.

71. Burton CL, Park LS, Corfield EC, Forget-Dubois N, Dupuis A, Sinopoli VM, et al. Heritability of obsessive-compulsive trait dimensions in youth from the general population. Transl Psychiatry. 2018;8:191.

72. Van Grootheest DS, Boomsma DI, Hettema JM, Kendler KS. Heritability of obsessive-compulsive symptom dimensions. Am J Med Genet Part B Neuropsychiatr Genet. 2008;147:473–478.

Figure Legends

Figure 1: Genome-Wide Significant Locus in *PTPRD* Associated with OC Traits in Spit for Science

a) Manhattan plot for GWAS of the TOCS standardized total score. rs7856850 in one the introns of *PTPRD* surpassed the genome-wide threshold (p=5x10^{-8}; grey line). b) Locus zoom plot for the genome-wide significant locus from the GWAS of the TOCS standardized total score. c). QQ Plot for the GWAS of the TOCS standardized total score. n=5018

Figure 2: Genome-wide significant loci in CNS Development Gene-Set Identified with sFDR

Manhattan plot from stratified False Discovery Rate (sFDR) test prioritizing central nervous system (CNS) development genes. Three loci reached genome-wide significance: *NPAS2* (q_{SFDR}=0.0123 for rs3754673 [A/C]), *GRID2* (q_{SFDR}=0.0362 for rs5860287 [AT/A indel]) and *SH3GL2* (q_{SFDR}=0.0096 for rs2182089 [A/T]). Light grey circles represent SNPs in the low priority group (i.e., SNPs in genes not in the CNS development gene-set). n=5018

Figure 3: Replication of Locus in PTPRD in OCD Replication Samples

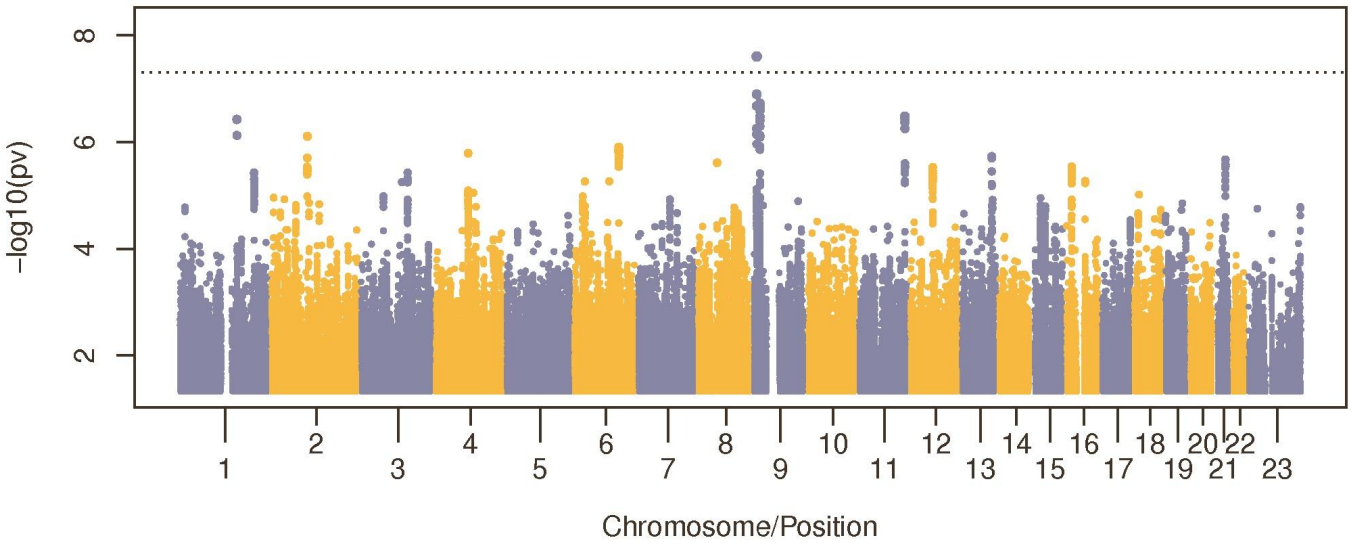
Forrest plot of genome-wide significant variant (rs7856850) across all the replication samples and sub-samples: 1) IOCDF/OCGAS (International Obsessive-Compulsive Disorder Foundation Collaborative and OCD Collaborative Genetics Association Studies samples), 2) CHOP (Philadelphia Neurodevelopmental Cohort (PNC) from the Children's Hospital of Philadelphia and 3) Michigan/Toronto OCD Imaging Genomics Study. OR = Odds Ratio

Figure 4: OC Traits in the Community and OCD Share Polygenic Risk

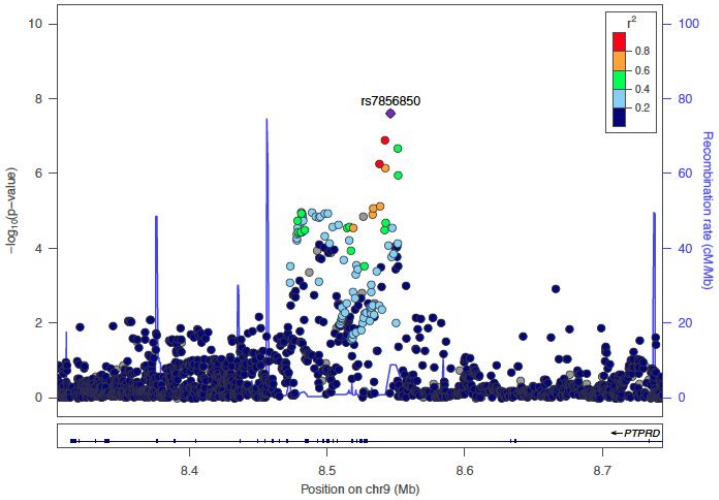
a) Variance explained (R^2) in OCD case/control status in replication samples by polygenic risk for OC traits from Spit for Science b) variance explained in OC traits in Spit for Science sample by polygenic risk for OCD from replication samples across a range of prior proportion of causal variants (ρ). Analyses conducted using LDpred. * = $p < 0.01$

Figure 1

A


−log10(pv)
8
6
4
2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Chromosome/Position

B


rs7856850
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
PTPRD
8.4
8.5
8.6
8.7
Position on chr9 (Mb)

C

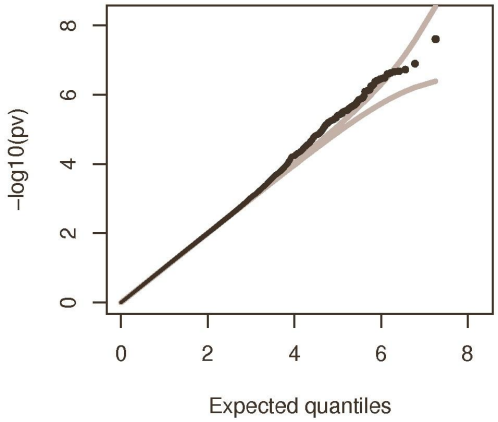

−log10(pv)
0
2
4
6
8
Expected quantiles

Figure 2

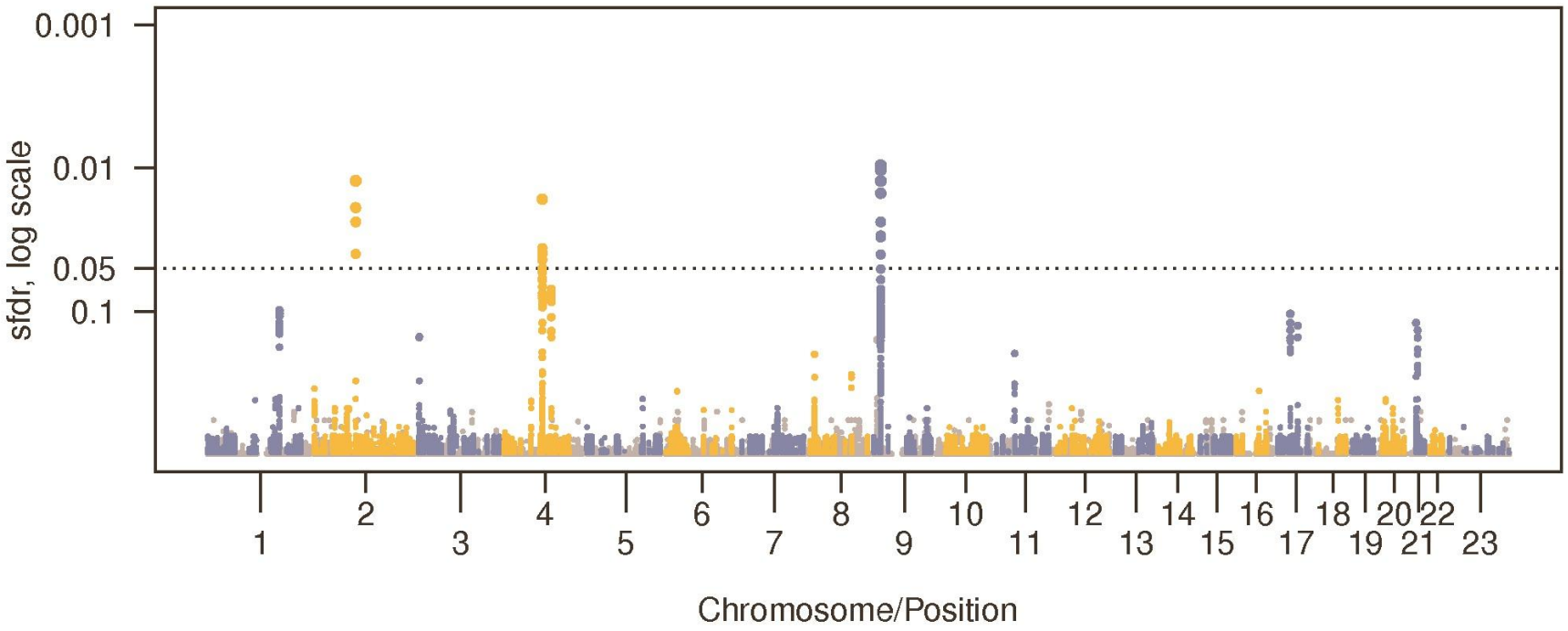

0.001
0.01
0.05
0.1
sfdr, log scale
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Chromosome/Position

Figure 3

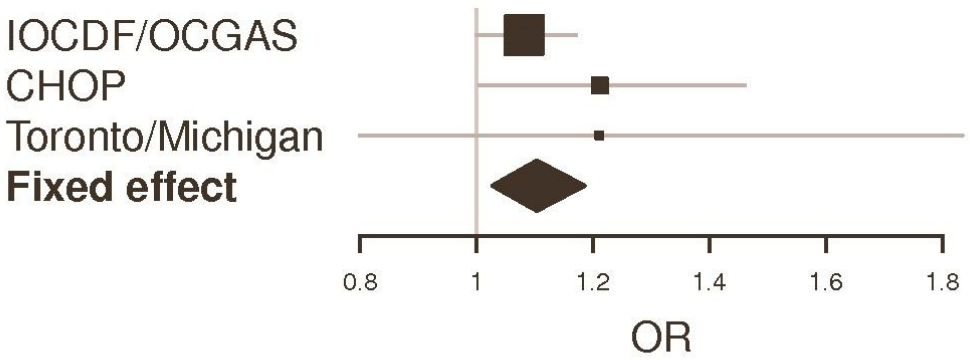

IOCDF/OCGAS
CHOP
Toronto/Michigan
Fixed effect
0.8
1
1.2
1.4
1.6
1.8
OR

Figure 4

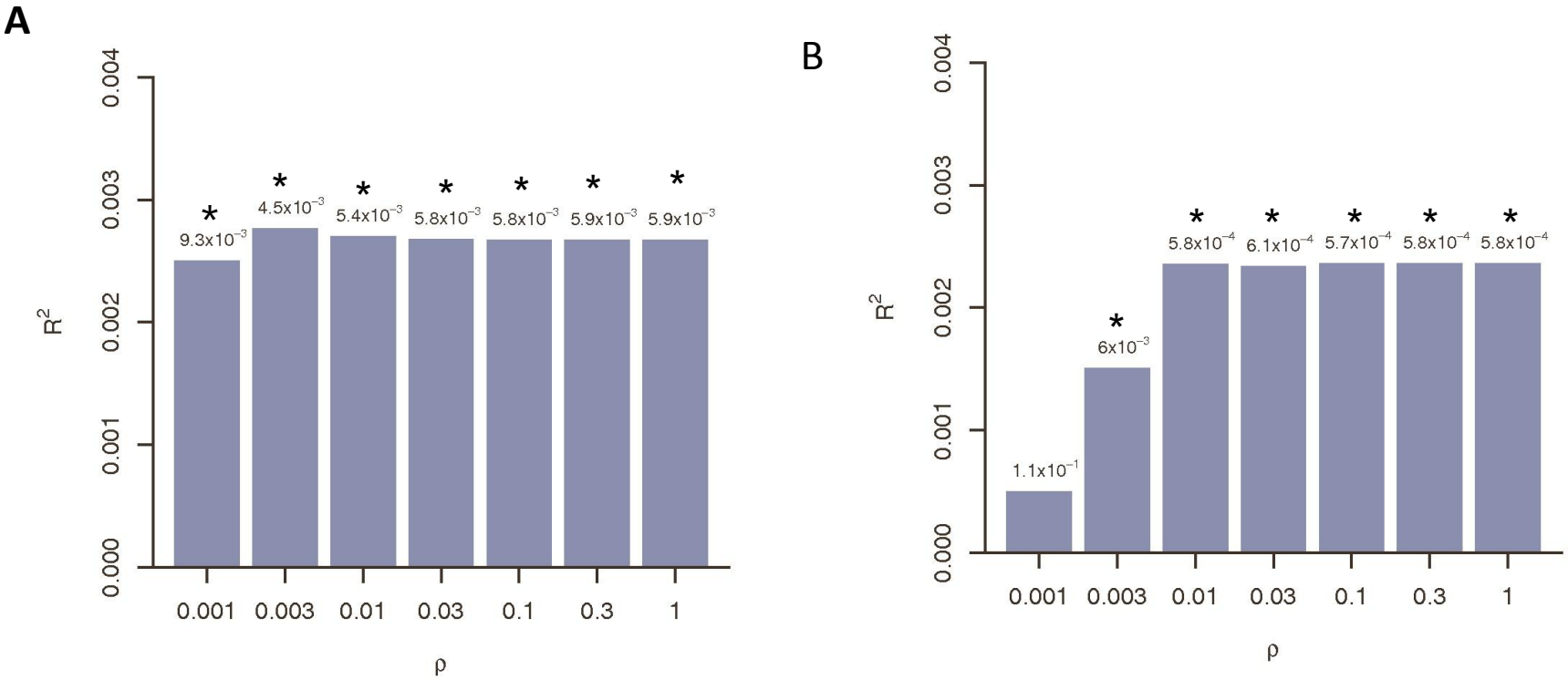

A
B
R^2
0.000
0.001
0.002
0.003
0.004
*
9.3x10^{-3}
4.5x10^{-3}
5.4x10^{-3}
5.8x10^{-3}
5.8x10^{-3}
5.9x10^{-3}
5.9x10^{-3}
1.1x10^{-1}
6x10^{-3}
5.8x10^{-4}
6.1x10^{-4}
5.7x10^{-4}
5.8x10^{-4}
5.8x10^{-4}
0.001
0.003
0.01
0.03
0.1
0.3
1
ρ

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†	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			**11980**	**3369**	**8611**

† There were 62 individuals that had a parent- or self-reported diagnosis of OCD. Number of samples reflect 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