Serotonin system genes and obsessive-compulsive trait dimensions in a population-based, pediatric sample: A genetic association study

Authors

Vanessa M. Sinopoli[a,b], Lauren Erdman[b,c], Christie L. Burton[b,d], Laura S. Park[a,b], Annie Dupuis[e,f], Janet Shan[d], Tara Goodale[d], S-M Shaheen[g], Jennifer Crosbie[d,h], Russell J. Schachar[a,d,h], Paul D. Arnold[b,g,i,*]

Affiliations

[a]*Institute of Medical Science, University of Toronto, Toronto ON, Canada*
[b]*Program in Genetics & Genome Biology, The Hospital for Sick Children, Toronto ON, Canada*
[c]*Department of Computer Science, University of Toronto, Toronto ON, Canada*
[d]*Program in Neurosciences and Mental Health, The Hospital for Sick Children, Toronto ON, Canada*
[e]*Clinical Research Services, The Hospital for Sick Children, Toronto ON, Canada*
[f]*Dalla Lana School of Public Health, University of Toronto, Toronto ON, Canada*
[g]*Mathison Centre for Mental Health Research & Education, Hotchkiss Brain Institute, University of Calgary, Calgary AB, Canada*
[h]*Department of Psychiatry, University of Toronto, Toronto ON, Canada*
[i]*Departments of Psychiatry and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary AB, Canada*

* Corresponding Author: Paul Daniel Arnold, MD, PhD, FRCPC, Mathison Centre for Mental Health Research and Education, 4th floor, Teaching, Research and Wellness (TRW) Building, 3280 Hospital Dr NW, Calgary AB, T2N 4Z6, Canada; Email: paul.arnold@ucalgary.ca; Phone: (403) 210-6464

Conflicts of Interest

Russell J. Schachar has acted as a consultant for Highland Therapeutics, ehave, and Eli Lilly Canada Inc. and is the Toronto Dominion Bank Financial Group Chair in Child and Adolescent Psychiatry. Vanessa M. Sinopoli, Lauren Erdman, Christie L. Burton, Laura S. Park, Annie Dupuis, Janet Shan, Tara Goodale, S-M Shaheen, Jennifer Crosbie, and Paul D. Arnold have no conflicts of interest to declare.

CITATION

Sinopoli, V., Erdman, L., Burton, C.L., Park, L.S., Dupuis, A., Shan, J., Goodale, T., Shaheen, S-M, Crosbie, J., Schachar, R.J., & Arnold, P.D. (2019). Serotonin system genes and obsessive-compulsive trait dimensions in a population-based, pediatric sample: a genetic association study. *Journal of Child Psychology and Psychiatry,* 60(12): 1289-1299. https://doi.org/10.1111/jcpp.13079

ABSTRACT

Background: Serotonin system genes are commonly studied in obsessive-compulsive disorder (OCD), but genetic studies to date have produced inconsistent results, possibly because phenotypic heterogeneity has not been adequately accounted for. In this paper, we studied candidate serotonergic genes and homogenous phenotypic subgroups as presented through obsessive-compulsive (OC) trait dimensions in a general population of children and adolescents. We hypothesized that different serotonergic gene variants are associated with different OC trait dimensions and, furthermore, that they vary by sex. **Methods**: Obsessive-compulsive trait dimensions (Cleaning/ Contamination, Counting/Checking, Symmetry/Ordering, Superstition, Rumination, and Hoarding) were examined in a total of 5,213 pediatric participants in the community using the Toronto Obsessive-Compulsive Scale (TOCS). We genotyped candidate serotonin genes (directly genotyping the 5-HTTLPR polymorphism in SLC6A4 for 2018 individuals and using single nucleotide polymorphism (SNP) array data for genes SLC6A4, HTR2A, and HTR1B for 4711 individuals). We assessed the association between variants across these genes and each of the OC trait dimensions, within males and females separately. We analyzed OC traits as both (a) dichotomized based on a threshold value and (b) quantitative scores. **Results**: The [LG + S] variant in 5-HTTLPR was significantly associated with hoarding in males (p-value of 0.003 and 0.004 for categorical and continuous analyses, respectively). There were no significant findings for 5-HTTLPR in females. Using SNP array data, there were significant findings for rumination in males for HTR2A SNPs (p-value of 1.04e-6 to 5.20e-6). **Conclusions**: This represents the first genetic association study of OC trait dimensions in a community-based pediatric sample. Our strongest results indicate that hoarding and rumination may be distinct in their association with serotonin gene variants and that serotonin gene variation may be specific to sex. Future genetic association studies in OCD should properly account for heterogeneity, using homogenous subgroups stratified by symptom dimension, sex, and age group.

Keywords: Obsessive-compulsive disorder; symptom dimensions; phenotypic heterogeneity; population-based; serotonin system; serotonin genes; SLC6A4; 5-HTTLPR; HTR2A; HTR1B; genetic association.

Introduction

Obsessive-compulsive disorder (OCD) is a debilitating psychiatric disorder characterized by recurring, disturbing thoughts and/or repetitive behaviors carried out in response to mounting anxiety (American Psychiatric Association [APA], 2013). The estimated worldwide prevalence of OCD is 2%–3% (Angst et al., 2004; Kessler et al., 2005; Murphy et al., 2013) and 0.5%–2% in children and adolescents, although OCD is often underdiagnosed in the pediatric population (Alvarenga et al., 2015; Geller & March, 2012). The most efficacious, pharmacological treatments for OCD are medications that act on the serotonin system, collectively referred to as serotonin reuptake inhibitors (SRIs), which include selective serotonin reuptake inhibitors (SSRIs), combined serotonin-norepinephrine reuptake inhibitors (SNRIs), and clomipramine (a tricyclic compound) (Murphy, Lerner, Rudnick, & Lesch, 2004).

Obsessive-compulsive disorder is familial and heritable with estimates ranging from 27% to 47% in adults and from 45% to 65% in children (Pauls, Abramovitch, Rauch, & Geller, 2014; Van Grootheest, Cath, Beekman, & Boomsma, 2005). Genetic studies have focused on serotonin system gene variants because of the efficacy of SRIs for treating OCD (Millan, Goodwin, Meyer-Lindenberg, & Ove Ogren, 2015). SRIs block serotonin reuptake in the brain via binding and allosteric modulation of the serotonin transporter (SERT) (Billett et al., 1997; Blier, de Montigny, & Chaput, 1990). The most commonly studied variants are in the genes coding 5-HT2A (*HTR2A*) and 5-HT1B (*HTR1B*) (Sinopoli, Burton, Kronenberg, & Arnold, 2017; Taylor, 2013).

One of the most researched serotonin gene polymorphisms is *5-HTTLPR* (*5-HTT*-linked polymorphic region), located within the promoter region of SLC6A4 (Sinopoli et al., 2017). *5-HTTLPR* exists in two major forms: the long (L) or short (S) variant. The L variant is made up of 16 sets of 20–23 base pair (bp) tandem repeats, while the S variant is made up of 14 sets of 20–23 bp tandem repeats. The L variant results in greater SERT expression than the S variant (Heils et al., 1996; Lesch et al., 1996). Recently, an intrinsic single nucleotide polymorphism (SNP), rs25531, was discovered within the additional repeats in the L variant. When the less frequent G allele of rs25531 is present in the L variant, SERT expression is similar to the S variant. This results in two functional groups: LA variant with high SERT expression and LG/S variant with low SERT expression (Hu et al., 2006). The rs25531 SNP was not accounted for in early association studies, which may contribute to some of the discrepancies in the literature to date (Sinopoli et al., 2017). In a recent meta-analysis, the LA variant of *5-HTTLPR* and the A allele of SNP rs6311 or the linked T allele of SNP rs6313 in HTR2A were associated with OCD (Taylor, 2013, 2016).

Obsessive-compulsive disorder is phenotypically complex and can be present with various types of symptoms. A meta-analysis of previous factor analyses of the gold standard measure of OCD, the Yale Brown Obsessive-Compulsive Scale (Y-BOCS), in adults and children with OCD, reported four symptom dimensions: (a) Symmetry (b) Forbidden thoughts, (c) Cleaning, and (d) Hoarding. Although similar factors were identified in children and adults, the specific items

contributing to each dimension differed based on age group (Bloch, Landeros-Weisenberger, Rosario, Pittenger, & Leckman, 2008). Factor analysis also demonstrates that symptom dimensions exist for obsessive-compulsive (OC) traits in the general population (Burton et al., 2018). The prevalence and distribution of symptom dimensions in clinical and in general population samples also differ between males and females (Alvarenga et al., 2015; Flament, 1990; Ruscio, Stein, Chiu, & Kessler, 2010). OCD symptom dimensions are heritable, with evidence of both shared and/or unique genetic influences between them in clinical and population-based samples (Burton et al., 2018; Katerberg et al., 2010; Van Grootheest, Boomsma, Hettema, & Kendler, 2008), and they differ in terms of their neurobiology, comorbidities, and response to treatment (Pauls et al., 2014). For example, response to SRIs varies depending on prevailing OCD symptoms or symptom dimensions (Landeros-Weisenberger et al., 2010). Thus, OCD symptom dimensions may be mediated by different underlying biological mechanisms.

The genetic basis of OCD is unclear, likely because it is often treated as a unidimensional phenomenon. The purpose of this study was to examine whether different candidate serotonin system gene variants are associated with different obsessive-compulsive trait dimensions. Given that pediatric OCD is often underdiagnosed, and that children and adolescents with mild to moderate symptom severity may not present in the clinic (Geller & March, 2012), we used a population-based sample to capture symptoms that may otherwise remain undetected. To account for previously reported genetic differences in OCD between sexes (Mattina & Steiner, 2016), we ran our analyses in males and females separately. Since candidate serotonin genes have been implicated in OCD and given that OCD symptom dimensions differ in terms of their heritability, neurobiology, and clinical presentation, we hypothesized that we would find different serotonin gene variants associated with different OC trait dimensions and that these findings would vary based on sex.

Methods

Participants

Seventeen thousand two hundred and sixty-three children and adolescents, aged 6–18 years, were recruited and assessed at the Ontario Science Centre in Toronto, Canada, for the Thoughts, Actions, and Genes (TAG) project, described in detail elsewhere (Burton et al., 2016; Crosbie et al., 2013; Park et al., 2016). Informed consent, and assent where applicable, was obtained for all participants. The study was approved by the Hospital for Sick Children Research Ethics Board. 16,718 participants had complete demographic information and questionnaires from either parent-respondent (N = 13,680) or self-respondent (N = 3,038). From this sample, we obtained two smaller, partially overlapping groups of participants who had four grandparents of European descent. The first group had 2,100 participants with DNA samples readily available to directly genotype the *5-HTTLPR* region of interest in *SLC6A4*. The second group had 4,810 participants with genome-wide association study (GWAS) data available, from which candidate gene SNPs were derived.

Obsessive-compulsive features

Participants or their parents completed the 21-item Toronto Obsessive Compulsive Scale (TOCS) that measures OC traits (Park et al., 2016). Each of the 21 items was scored on a scale of -3 to +3, where -3 signifies that the child performs the action of interest far less often than his or her peers, and + 3 signifies that the child performs the action of interest far more often than his or her peers. A previous factor analysis of the TOCS identified six factors: (a) Cleaning/Contamination, (b) Counting/Checking, (c) Symmetry/Ordering, (d) Superstition, (e) Rumination, and (f) Hoarding (Burton et al., 2018). Heritability estimates for each of these dimensions range from 30% to 77% (Burton et al., 2018). For our primary set of analyses, we dichotomized each item, whereby a score of ≥2 on an item indicates presence of the obsessive-compulsive trait and any other score indicates absence of that trait. Individuals were then assigned 'affected' status for the OC group if they had at least one reported obsessive-compulsive trait. They were also assigned 'affected' status for groups, based on the previous factor analysis, if they had at least one reported trait within the dimension. Note that individuals could be deemed 'affected' for more than one dimension. Individuals who did not have any OC traits were deemed controls. For our secondary set of analyses, we treated the items as continuous.

DNA collection and extraction

Saliva was collected from all participants using OrageneDNA (OG-500) kits and DNA extracted using the protocol recommended by the manufacturer (DNA Genotek). We also centrifuged the samples at 10,000 rpm for another 10 min to remove any precipitated carbohydrates. DNA was quantified using the Quant-iTTM PicoGreen dsDNA Assay Kit (Invitrogen). DNA samples were run on agarose gels to ensure good quality before using microarrays. Samples with DNA concentration below 60 ng/ll [if genotyped on the HumanCoreExome-12 v1.0 microarray (Illumina)] and samples with poor DNA quality were excluded.

Selection of candidate genes

We selected serotonin candidate genes *SLC6A4*, *HTR2A*, and *HTR1B* based on their inclusion in a recent OCD meta-analysis (Taylor, 2013). Given our interest in the serotonin system, we included three serotonin genes in which the polymorphisms with the best evidence of association with OCD were located.

5-HTTLPR

Direct genotyping. Variation in *5-HTTLPR* is not captured via microarrays and, thus, was directly genotyped in a two-step process (Wendland, Martin, Kruse, Lesch, & Murphy, 2006), in conjunction with The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children in Toronto:

(a) Identifying L or S variant by polymerase chain reaction (PCR) amplification of *5-HTTLPR* polymorphism: *5-HTTLPR* was PCR amplified using forward primer 50 -ATGCCAGCACCTA ACCCCTAATGT-30 , which was 50 -labeled with HEX fluorescent dye for visualization, and reverse primer 50 -GGACCGCAAGG TGGGCGGGA-30 (Gelernter, Kranzler, & Cubells, 1997). PCR involved 50 ng of our genomic template DNA, 10 mM dNTP mix (Fermentas, Life Technologies), the PCRx Enhancer System (Invitrogen, Life Technologies) inclusive of 10X

PCRx Enhancer Solution, 50 mM MgSO4, 10X PCRx Amplification Buffer, and 5,000 U/ml Taq DNA polymerase (New England BioLabs). PCR cycling conditions involved 30 cycles of 95°C for 30 s, 55°C for 45 s, and 68°C for 60 s. One microliter volumes of the PCR products were suspended in a 10 ll mixture of 7 ll GeneScan 500 LIZ dye Size Standard in 993 ll Hi-Di formamide (Applied Biosystems). Samples were subsequently run on an ABI3730XL genetic analyzer using the POP-7 polymer and Dye Set G5 (Applied Biosystems). The Peak Scanner Software v1.0 (Thermo Fisher Scientific) was used to analyze results. The amplified *5- HTTLPR* products were predicted to be 419 and 375 bp in length for the L and S variants, respectively. We found the mobility of the amplified fragments on the ABI3730XL genetic analyzer to read as 412 and 370 bp for L and S, respectively.

(b) Determining rs25531 SNP by digestion of identified L or S variant: Once the *5-HTTLPR* was identified as L or S, the DNA was digested to determine an A or G SNP at rs25531. A restriction enzyme digested the amplified PCR product to cut the sequence at rs25531 if a G was present. As per the manufacturer's protocol, we used 0.1–0.5 lg of PCR DNA, 10X Buffer Tango, and 0.5–2 ll of restriction enzyme MspI (Fermentas, Life Technologies). One microliter volumes of the resulting DNA products were suspended in a 10 ll mixture of 7 ll GeneScan 500 LIZ dye Size Standard in 993 ll Hi-Di formamide (Applied Biosystems). Samples were run on an ABI3730XL genetic analyzer using the POP-7 polymer and Dye Set G5 (Applied Biosystems), and the Peak Scanner Software v1.0 (Thermo Fisher Scientific) was used to analyze results. We identified that if no G was present at rs25531, the digested L variant yielded a 321-bp fragment and the digested S variant yielded a 278-bp fragment. If a G, but not an A, was present at rs25531, the enzyme cut at rs25531 to yield a 149-bp fragment. It is documented that the G SNP is located in the additional stretch of DNA within the L variant (Hu et al., 2006), but there have been reports of a rare SG variant that carries a G SNP at a site resulting in the same length fragment as LG when digested (Voyiaziakis et al., 2011; Wendland et al., 2006). Thus, the resulting digested fragments would be the same length for both LG and SG variants. It is for this reason that the L versus S variant data were obtained from the initial PCR amplification, prior to digestion. L versus S data were then combined with data from the rs25531 SNP restriction enzyme method to complete *5-HTTLPR* genotyping and accurately identify LA, LG, and S variants (inclusive of an SA variant and the rare SG variant) (Figure 1). We then grouped our variants into two allelic groups, based on functionality, as follows: LA and [LG + S]. We also took into account that the alleles are additive in their effect (Hu et al., 2006).

For the *5-HTTLPR* sample, quality control (QC) was implemented when visualizing and analyzing genotypes, using the standard Peak Scanner Software v1.0 (Thermo Fisher Scientific) program settings. Individuals were genotyped with a 97% completion rate. Reported siblings and individuals of indeterminate ethnicity were removed. When siblings were identified, we kept the individual in the OC group if possible. We identified 12 individuals carrying the rare SG variant in our final, total *5-HTTLPR*-genotyped sample. Given that we found no evidence of differential expression of the SG in the literature, we chose to include individuals with this variant in our analysis and considered both SA and the rare SG variants as 'S'. After QC and removal of ensuing individuals, we had a final sample size of 2,018.

Statistical analyses.

a. Primary analyses using dichotomized trait scores: Analyses were conducted using the program, R, version 3.0.1 (https://www.R-project.org). The expected number of tested alleles (0, 1, or 2 [LG + S] alleles) was used in a logistic regression as a predictor of affected versus

control status for each of the seven trait groups. An additive model was used. Analyses were conducted in male and in female subjects separately, using age and questionnaire respondent type (parent or self) as covariates. To account for multiple comparisons, our threshold of significance was a p-value

b. Secondary analyses using quantitative trait scores: We reanalyzed our *5-HTTLPR* data using a quantitative trait approach (without dichotomizing the trait scores, but instead treating the traits as continuous). For this set of analyses, we used a linear model and the traits were made normal using ordered quantile normalization applied to the sum of each trait.

Candidate gene SNPs

Genotyping. Samples were genotyped on the HumanCoreExome-12 v1.0 microarray (Illumina) for collection of SNP data at TCAG. The microarray data were available from an ongoing GWAS conducted by our group (Burton et al., 2015). We analyzed SNPs across *SLC6A4*, *HTR2A*, and *HTR1B* (UCSC Genome Browser, hg19). Standard QC removed SNPs with call rate <0.97 and samples with sex misspecification and ambiguity. Only one member per family was used in the analyses using the PI_HAT cutoff = 0.12. European ethnicity was verified via principal component analysis (PCA) using EIGENSTRAT version 3.2.10 (Price et al., 2006) with HapMap populations CEU and TSI to ensure that our individuals' genetic data clustered with Europeans, relative to other ethnicities, and then via another PCA to identify and remove outliers within our own population. An additional PCA was conducted without outliers, followed by formation of a scree plot, to determine that there were seven principal components (PCs) we needed to use to control for population structure in our analyses. After QC and removal of ensuing individuals, we had a final sample size of 4,711.

Genetic data were imputed over the genetic regions for *SLC6A4*, *HTR2A*, and *HTR1B* along with ± 50 kb flanking regions for each gene using SHAPEIT2 version 2.790 (Delaneau & Marchini, 2014) and IMPUTE2 version 2.3.1 (Howie, Donnelly, & Marchini, 2009). 1000 Genomes Phase 3 reference data were used for both pre-phasing and imputation. After imputation, SNPs were removed for low-quality imputation (information score <0.8). SNPs with a minor allele frequency (MAF) <0.05 were excluded from the analysis. Data were kept as dosage calls for all analyses. Table 1 shows final SNP counts across each of our three genes of interest.

Statistical analyses

a. Primary analyses using dichotomized trait scores: Using R, version 3.0.1 (https://www.R-project.org), logistic regressions for each SNP were conducted using the expected number of tested alleles (0, 1, or 2) to predict affected versus control status for each of the seven trait groups. An additive model was used. Regressions were conducted for males and females separately, using age, respondent, and the identified PCs as covariates. To address multiple comparisons, we used the Genetic Type 1 error calculator (GEC), developed to account for dependent SNPs (Li, Yeung, Cherny, & Sham, 2012). Our threshold of significance was a p-value <2.94e-5 (2.06e-4/7 analyses) for males and a p-value <2.90e-5 (2.03e-4/7 analyses) for females.

b. Secondary analyses using quantitative trait scores: We reanalyzed our SNP data using a quantitative trait approach (without dichotomizing the trait scores, but instead treating the traits as continuous). For this set of analyses, we used a linear model and the traits were made normal using ordered quantile normalization applied to the sum of each trait.

Table 1: SNP counts across 3 candidate genes, including all genotyped and imputed SNPs: Common SNPs (MAF ≥ 0.05) are shown for each gene and its flanking regions included in the analyses for males and in the analyses for females.

Gene	Number of SNPs in Males	Number of SNPs in Females
SLC6A4	121	142
HTR2A	411	361
HTR1B	242	242

Abbreviations: SNP, single nucleotide polymorphism; *SLC6A4*, solute carrier family 6 member 4 (serotonin transporter gene); *HTR2A*, 5-HT_{2A} receptor gene; *HTR1B*, 5-HT_{1B} receptor gene.

Sample size considerations

The final sample size for *5-HTTLPR* (N = 2,018) was smaller than the final sample size for our candidate gene SNPs (N = 4,711), with a 29% overlap between these two groups (502 participants genotyped in the *5-HTTLPR* group only; 3,195 participants genotyped in the candidate gene SNP group only; 1,516 participants genotyped for and used in both the *5-HTTLPR* and candidate gene SNP groups). The smaller sample size for *5-HTTLPR* was deemed more than sufficient given the less stringent statistical correction required for the single variant versus the set of candidate gene SNPs, and given the increased labor and cost required to directly genotype *5-HTTLPR*.

Results

***5-HTTLPR* analyses**

5-HTTLPR: dichotomized trait analyses. Demographics for the *5-HTTLPR*-genotyped individuals for the dichotomized trait analyses are reported in Table 2. Using t-tests, we identified no significant differences for age between affected males and male controls, or between affected females and female controls. We identified a significant difference for respondent between affected males and male controls (absolute t-statistic = 3.09, p-value = 0.002), but not between affected females and female controls.

As shown in Figure 2, after correction for multiple comparisons, we found a significant association between the [LG + S] variant and hoarding in males (odds ratio, OR, of 1.35; p-value of 0.003). *5-HTTLPR* was not significantly associated with OC traits or any specific OC trait dimension aside from hoarding in males, after correction for multiple comparisons. *5-HTTLPR* was not significantly associated with OC traits or any specific OC trait dimension in females.

5-HTTLPR: quantitative trait analyses. After correction for multiple comparisons, the [LG + S] variant in 5-HTTLPR was significantly associated with hoarding in males (effect of 0.19; p-value of 0.004). We report no other significant findings in males or in females.

Table 2: 5-HTTLPR-genotyped individuals for the dichotomized trait analyses: Demographics and group characteristics are shown for total individuals combined, all affected individuals, and control individuals

Total Individuals		
	Male	**Female**
N	1050	968
Mean Age (SD)	10.7 (2.5)	11.3 (2.9)
Age Range	6.1 - 17.9	6.3 - 17.9
Respondent: % Parent-report, % Self-report	90.1, 9.9	81.5, 18.5
All Affected Individuals		
	Male	**Female**
N	590	571
Mean Age (SD)	10.6 (2.5)	11.2 (3.0)
Age Range	6.3 - 17.9	6.3 - 17.9
OCD Diagnosis, N (%)	32 (5.4)	30 (5.3)
Mood Disorder Diagnosis, N (%)	12 (2.0)	24 (4.2)
ADHD Diagnosis, N (%)	106 (18.0)	39 (6.8)
Anxiety Disorder Diagnosis, N (%)	74 (12.5)	66 (11.6)
ASD Diagnosis, N (%)	59 (10.0)	12 (2.1)
Tic Disorder Diagnosis, N (%)	28 (4.7)	16 (2.8)
Taking SRI Medication, N (%)	20 (3.4)	19 (3.3)
Respondent: % Parent-report, % Self-report	89.8, 10.2	81.3, 18.7
Allele Frequency: L_A, [L_G + S]	0.48, 0.52	0.52, 0.48
Control Individuals		
	Male	**Female**
N	460	397
Mean Age (SD)	10.8 (2.5)	11.3 (2.9)
Age Range	6.1 - 17.7	6.3 - 17.9
OCD Diagnosis, N (%)	0 (0.0)	0 (0.0)
Mood Disorder Diagnosis, N (%)	3 (0.7)	1 (0.3)
ADHD Diagnosis, N (%)	25 (5.4)	10 (2.5)
Anxiety Disorder Diagnosis, N (%)	3 (0.7)	5 (1.3)
ASD Diagnosis, N (%)	1 (0.2)	0 (0.0)
Tic Disorder Diagnosis, N (%)	2 (0.4)	1 (0.3)
Taking SRI Medication, N (%)	0 (0.0)	0 (0.0)
Respondent: % Parent-report, % Self-report	90.4, 9.6	81.9, 18.1
Allele Frequency: L_A, [L_G + S]	0.52, 0.48	0.53, 0.47

Abbreviations: *5-HTTLPR*, 5-HTT-linked polymorphic region; N, sample size; SD, standard deviation; OCD, obsessive-compulsive disorder; ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; SRI, serotonin reuptake inhibitor; L_A, long variant of *5-HTTLPR* with A allele at SNP rs25531; L_G, long variant of *5-HTTLPR* with G allele at SNP rs25531; S, short variant of *5-HTTLPR*.

Candidate gene SNP analyses

Candidate gene SNPs: dichotomized trait analyses. Demographics for the individuals included in the dichotomized trait analyses of candidate gene SNPs are reported in Table 3. Using t-tests, we identified no significant differences for age or for respondent between affected males and male controls, or between affected females and female controls. Most of the top SNPs reported in males corresponded with *HTR2A* and were associated with rumination (OR of 0.62–0.65; p-value of 1.34e-3 to 3.63e-3; Table 4). For females, all of the top reported SNPs corresponded with *HTR1B* and were associated with hoarding (OR of 1.33 to 1.42; p-value of 1.05e-3 to 4.60e-3; Table 5). None of these findings remained significant after correction for multiple comparisons.

Candidate gene SNPs: quantitative trait analyses. For this set of analyses, we report SNP findings that are significant after correction for multiple comparisons in males (effect of 0.19 to 0.20; pvalue of 1.04e-6 to 5.20e-6; Table 6). The significant SNP findings all corresponded with *HTR2A* and rumination in males. We report no other significant findings, after correction for multiple comparisons, in males or in females.

Analysis after removal of individuals with OCD diagnosis

We reanalyzed our data after the removal of individuals who reported an OCD diagnosis. There were no significant *5-HTTLPR* findings, after correction for multiple comparisons, for OC traits overall or for any of the OC trait dimensions in males or in females. Similarly, there were no significant SNP findings, after correction for multiple comparisons, for OC traits overall or for any of the OC trait dimensions in males or in females.

Table 3: Individuals genotyped for candidate gene SNPs for the dichotomized trait analyses: Demographics and group characteristics are shown for total individuals combined, all affected individuals, and control individuals

Total Individuals		
	Male	**Female**
N	2439	2272
Mean Age (SD)	10.7 (2.6)	11.1 (2.9)
Age Range	6.1 - 17.9	6.1 - 17.9
Respondent: % Parent-report, % Self-report	89.7, 10.3	81.0, 19.0
All Affected Individuals		
	Male	**Female**
N	1025	1043
Mean Age (SD)	11.1 (2.8)	11.8 (3.1)
Age Range	6.1 - 17.9	6.4 - 17.9
OCD Diagnosis, N (%)	28 (2.7)	27 (2.6)
Mood Disorder Diagnosis, N (%)	18 (1.8)	30 (2.9)
ADHD Diagnosis, N (%)	157 (15.3)	44 (4.2)
Anxiety Disorder Diagnosis, N (%)	95 (9.3)	82 (7.9)
ASD Diagnosis, N (%)	65 (6.3)	11 (1.1)
Tic Disorder Diagnosis, N (%)	30 (2.9)	15 (1.4)
Taking SRI Medication, N (%)	30 (2.9)	20 (1.9)
Respondent: % Parent-report, % Self-report	82.8, 17.2	71.9, 28.1
Control Individuals		
	Male	**Female**
N	1414	1229
Mean Age (SD)	10.3 (2.3)	10.6 (2.6)
Age Range	6.1 - 17.7	6.1 - 17.9
OCD Diagnosis, N (%)	0 (0)	0 (0)
Mood Disorder Diagnosis, N (%)	8 (0.6)	2 (0.2)
ADHD Diagnosis, N (%)	95 (6.7)	39 (3.2)
Anxiety Disorder Diagnosis, N (%)	18 (1.3)	22 (1.8)
ASD Diagnosis, N (%)	10 (0.7)	1 (0.1)
Tic Disorder Diagnosis, N (%)	11 (0.8)	5 (0.4)
Taking SRI Medication, N (%)	0 (0)	0 (0)
Respondent: % Parent-report, % Self-report	94.8, 5.2	89.7, 10.3

Abbreviations: SNP, single nucleotide polymorphism; N, sample size; SD, standard deviation; OCD, obsessive-compulsive disorder; ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; SRI, serotonin reuptake inhibitor.

Table 4: Association between serotonin gene SNPs and OC traits overall/OC trait dimensions in males for the dichotomized trait analyses: Top SNP findings with p-value <1.00e-2

SNP	Gene	Tested Allele	MAF	OR	P-value	Trait Group
rs3742278	*HTR2A*	G	0.132	0.64	1.34E-03	Rumination
rs9567735	*HTR2A*	G	0.132	0.64	1.47E-03	Rumination
rs9562684	*HTR2A*	C	0.131	0.64	1.60E-03	Rumination
rs1923884	*HTR2A*	T	0.131	0.64	1.65E-03	Rumination
rs9562685	*HTR2A*	A	0.131	0.64	1.65E-03	Rumination
rs9567736	*HTR2A*	A	0.132	0.64	1.81E-03	Rumination
rs62416430	*HTR1B*	G	0.066	0.51	2.38E-03	Rumination
rs17068986	*HTR2A*	T	0.133	0.65	2.90E-03	Rumination
rs62416428	*HTR1B*	G	0.067	0.52	3.06E-03	Rumination
rs62416429	*HTR1B*	T	0.067	0.52	3.06E-03	Rumination
rs6505166	*SLC6A4*	A	0.052	1.93	3.50E-03	Counting/Checking
rs9534505	*HTR2A*	G	0.085	0.62	3.63E-03	Rumination
rs2020932	*SLC6A4*	A	0.051	2.04	3.96E-03	Counting/Checking
rs62416430	*HTR1B*	G	0.068	0.37	5.83E-03	Superstition
rs2770301	*HTR2A*	C	0.231	0.63	8.68E-03	Superstition

Abbreviations: SNP, single nucleotide polymorphism; MAF, minor allele frequency; OR, odds ratio; *SLC6A4*, solute carrier family 6 member 4 (serotonin transporter gene); *HTR2A*, 5-HT$_{2A}$ receptor gene; *HTR1B*, 5-HT$_{1B}$ receptor gene.

Table 5: Association between serotonin gene SNPs and OC traits overall/OC trait dimensions in females for the dichotomized trait analyses: Top SNP findings with p-value <1.00e-2

HTR1B, 5-HT1B receptor gene; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphism

SNP	Gene	Tested Allele	MAF	OR	P-value	Trait Group
rs1777762	*HTR1B*	T	0.154	1.39	1.05E-03	Hoarding
rs1777763	*HTR1B*	T	0.154	1.39	1.07E-03	Hoarding
rs2798528	*HTR1B*	T	0.154	1.39	1.09E-03	Hoarding
rs1738506	*HTR1B*	A	0.154	1.39	1.09E-03	Hoarding
rs2223832	*HTR1B*	C	0.154	1.39	1.11E-03	Hoarding
rs1343334	*HTR1B*	A	0.154	1.39	1.17E-03	Hoarding
rs2207053	*HTR1B*	C	0.153	1.38	1.32E-03	Hoarding
rs1228806	*HTR1B*	C	0.153	1.38	1.34E-03	Hoarding
rs1228805	*HTR1B*	G	0.154	1.38	1.48E-03	Hoarding
rs2207056	*HTR1B*	T	0.154	1.38	1.49E-03	Hoarding
rs1228797	*HTR1B*	T	0.154	1.38	1.49E-03	Hoarding
rs1228798	*HTR1B*	G	0.154	1.38	1.49E-03	Hoarding
rs1228800	*HTR1B*	G	0.154	1.38	1.49E-03	Hoarding
rs1343336	*HTR1B*	T	0.154	1.38	1.51E-03	Hoarding
rs1228802	*HTR1B*	A	0.153	1.37	1.65E-03	Hoarding
rs1145827	*HTR1B*	A	0.153	1.37	1.73E-03	Hoarding
rs2798529	*HTR1B*	G	0.151	1.37	1.81E-03	Hoarding
rs1228804	*HTR1B*	G	0.155	1.37	1.82E-03	Hoarding
rs6453980	*HTR1B*	G	0.153	1.37	1.90E-03	Hoarding
rs58608035	*HTR1B*	T	0.153	1.37	1.90E-03	Hoarding
rs60174069	*HTR1B*	C	0.153	1.37	1.90E-03	Hoarding
rs59414600	*HTR1B*	C	0.153	1.37	2.03E-03	Hoarding
rs1228803	*HTR1B*	T	0.154	1.36	2.03E-03	Hoarding
rs61295513	*HTR1B*	T	0.153	1.37	2.03E-03	Hoarding
rs6938832	*HTR1B*	T	0.152	1.36	2.24E-03	Hoarding
rs6939163	*HTR1B*	C	0.152	1.36	2.24E-03	Hoarding
rs4708338	*HTR1B*	C	0.152	1.36	2.26E-03	Hoarding
rs6938765	*HTR1B*	A	0.152	1.36	2.27E-03	Hoarding
rs4708340	*HTR1B*	C	0.152	1.36	2.28E-03	Hoarding
rs1777767	*HTR1B*	C	0.148	1.37	2.33E-03	Hoarding
rs10806097	*HTR1B*	T	0.121	1.42	2.33E-03	Hoarding
rs17272333	*HTR1B*	C	0.121	1.41	2.78E-03	Hoarding
rs112390126	*HTR1B*	C	0.121	1.41	2.79E-03	Hoarding
rs11755194	*HTR1B*	C	0.122	1.40	3.17E-03	Hoarding
rs55636038	*HTR1B*	T	0.122	1.40	3.17E-03	Hoarding
rs11753559	*HTR1B*	A	0.120	1.40	3.20E-03	Hoarding
rs12110491	*HTR1B*	T	0.123	1.40	3.21E-03	Hoarding
rs17272829	*HTR1B*	A	0.121	1.40	3.50E-03	Hoarding
rs62416427	*HTR1B*	A	0.121	1.40	3.59E-03	Hoarding
rs11757592	*HTR1B*	C	0.150	1.33	4.60E-03	Hoarding

Discussion

Previous studies have suggested that variants in serotonin system genes may be involved in OCD, but findings have been mixed. One reason for the mixed findings may be that previous studies have not examined whether or not different serotonin gene variants are associated with different symptom dimensions (Sinopoli et al., 2017; Taylor, 2013, 2016). Our study was the first to examine whether serotonin genes were associated with OC trait dimensions in males and in females in a pediatric, population-based sample. Overall, our findings support our hypothesis that different OC trait dimensions may have different genetic underpinnings and that genetic influences may be sex-specific, particularly showing that *5-HTTLPR* is significantly associated with hoarding traits in males and that a subset of SNPs corresponding with *HTR2A* are significantly associated with rumination in males.

For our primary, dichotomized/categorical analyses, our findings suggest that hoarding traits are biologically distinct from other OC traits. When comparing the two functionally distinct variants of *5-HTTLPR* (LA vs. [LG + S]), the [LG + S] variant, associated with lower SERT expression, was significantly associated with hoarding traits in males, but not in females. In our post hoc analysis combining males and females, the association between [LG + S] and hoarding was no longer statistically significant, suggesting that males are driving the association. Many studies have hypothesized an increase in SERT expression in OCD (Murphy et al., 2013; Sinopoli et al., 2017), but findings have been inconsistent overall. Our finding suggests that a decrease in SERT expression may be correlated with hoarding in males.

Our hoarding-driven findings are consistent with recent literature suggesting that hoarding is clinically and biologically distinct from OCD (Mataix-Cols & Pertusa, 2012; Snowdon, Pertusa, & Mataix-Cols, 2012). The Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) now classifies hoarding as a distinct disorder in the 'Obsessive-Compulsive and Related Disorders' category (APA, 2013). Recent research has also identified phenotypic, biological/genetic, and sex differences in patients with primary hoarding symptoms versus OCD patients without hoarding symptoms (Murphy et al., 2013) and shown that hoarding symptoms may have distinct neurological and genetic underpinnings (Van Ameringen, Patterson, & Simpson, 2014). As of yet, there is no clear evidence linking serotonergic genes to hoarding specifically. Our data suggest that particular serotonin gene variants may be associated with hoarding traits and, furthermore, that this association varies by sex. *5-HTTLPR* appears to be involved in hoarding symptomatology in males and there was a trend (in the dichotomized trait analyses) for variation downstream of *HTR1B* to be associated with hoarding symptomatology in females. Our data suggest that distinct biological underpinnings may also be driving other symptom dimensions within OCD and that these underpinnings vary by sex, as we observed a trend for variation across *HTR2A* to be association between the rumination trait dimension and in males.

5-HTTLPR and hoarding in males were significantly associated whether hoarding was treated categorically or quantitatively. Quantitative analyses also revealed a significant association between *HTR2A* SNPs and rumination in males not identified in the categorical analyses, where the same SNPs were approaching significance. Therefore, the quantitative analyses improved our

power to detect an association in males for rumination. Quantitative trait approaches have been shown to boost power in psychiatric genetic studies with phenotypes that have considerable variation as compared to dichotomous/categorical approaches (van der Sluis, Posthuma, Nivard, Verhage, & Dolan, 2013). This boost in power is increased in traits that are normally distributed rather than skewed (van der Sluis et al., 2013), and both hoarding and rumination were among our most normally distributed trait dimensions (Park et al., 2016). The fact that our significant association between *5-HTTLPR* and hoarding in males emerged using either approach suggests that this genetic finding is robust enough to be identified no matter how the data are analyzed, further supporting the uniqueness of hoarding disorder versus other OCD symptoms.

Our findings show that when studying the genetic basis of OCD, and of psychiatric disorders in general, it is important to reduce heterogeneity wherever possible to increase the power of genetic analyses and to study whether different trait dimensions have shared or distinct etiologies. The National Institute of Mental Health (NIMH) launched the Research Domain Criteria (RDoC) project to help researchers account for heterogeneity when studying mental health. The framework suggests researching basic dimensional constructs (identified through molecular, genetic, neurocircuitry-related, and behavioral analyses) that underlie higher-level behaviors and thus complex psychiatric disorders (NIMH, 2008). The major strength of our study includes our effort to address heterogeneity and use a dimensional approach consistent with RDoC.

Some inconsistencies in previous genetic studies could be attributed to failure to account for phenotypic heterogeneity in OCD, particularly stratifying data by both symptom dimensions and sex. For example, three studies used Y-BOCS-derived factors and analyzed association with *5-HTTLPR* in adults with OCD. Results varied across all three with one identifying an association between L-carriers and a religious/somatic factor (Kim, Lee, & Kim, 2005), another identifying an association between the S allele and the S/S genotype and symmetry obsessions/repeating counting and ordering/arranging compulsions (Hasler, Kazuba, & Murphy, 2006), and another reporting an associating between the L/L genotype (vs. L/S or S/S) of *5-HTTLPR* and counting and repeating rituals in individuals with comorbid tics (Cavallini, Di Bella, Siliprandi, Malchiodi, & Bellodi, 2002). Mixed results from these early studies may have been due to a failure to account for sex differences in addition to symptom dimensions. For example, another study examined SNPs in a glutamate receptor gene *GRIN2B* and found no association between *GRIN2B* SNPs and OCD. After stratifying by symptom dimension and sex, rs1805476 was significantly associated with contamination obsessions and cleaning compulsions in male patients (Alonso et al., 2012). Our results, although requiring replication, highlight the need to account for key sources of heterogeneity (including symptom dimensions and sex) to ensure that important associations are identified and to gain a fuller understanding of the complex etiology of OCD and traits that contribute to the disorder.

Limitations

One of the limitations of our study was sample size. Although adequate for a candidate gene approach, the size of our sample precluded an alternative, more comprehensive approach such as performing a GWAS in which we could comprehensively analyze all common variation across

the genome. A GWAS of multiple symptom dimensions would require very stringent correction for multiple comparisons (standard GWAS plus correction for multiple phenotypes), and our sample would be underpowered to detect associations. Therefore, we instead elected to perform a more focused study to examine whether candidate serotonergic gene variants were associated with OC trait dimensions in our pediatric, population-based sample. Another limitation was a difference in the initial sample collection approach used for our *5-HTTLPR*-genotyped group versus our candidate gene SNP group. We initially began collecting individuals for the *5-HTTLPR*-genotyped group using an extreme trait approach, which would result in higher TOCS scores in affected individuals and which likely explains why there are more individuals who reported an OCD diagnosis in this group (relative to our larger group of individuals genotyped for candidate gene SNPs). Though we cannot rule out whether or not our significant hoarding findings were being driven by a higher number of OCD-diagnosed individuals in the upper extreme, future work by our group is exploring the effect of non-hoarding OC traits on our main hoarding finding.

Conclusion

Our paper highlights the importance of addressing phenotypic heterogeneity based on clinical symptomatology and based on sex. More specifically, our findings suggest that hoarding and rumination are distinct in their underlying serotonin gene variation, relative to other OC dimensions. Future studies are necessary to replicate our findings of association. Additional studies are also required to verify that these trends go beyond the general population and occur in clinical settings. Although our findings were strongest for hoarding and rumination, we predict that different gene variants (serotonin genes and genes in other gene systems) are implicated in different OCD symptom dimensions and we, therefore, encourage prospective studies to use our approach to heterogeneity to build on our findings in independent, large datasets. Future studies should also examine additional sources of heterogeneity including gene–gene interaction and gene–environment interaction. Improved understanding of the relationship between genetic variants and specific OC traits will ultimately lead to more refined, targeted treatments for patients with OCD and related disorders.

Acknowledgements

This work was supported by the Canadian Institutes of Health Research (P.D.A., MOP-106573 and R.J.S., MOP-93696). P.D.A. receives funding from the Alberta Innovates Health Solutions (AIHS) Translational Health Chair in Child and Youth Mental Health. V.M.S. received funding from the Canadian Institutes of Health Research (CIHR) Master's Award: Frederick Banting and Charles Best Canada Graduate Scholarships, Ontario Graduate Scholarship (OGS), and the Hospital for Sick Children Restracomp Studentship. The authors also wish to acknowledge the contributions of Ms. Clara Tam in study coordination, Mr. Bingbin Li for laboratory technical assistance, and the Ontario Science Centre for collaborating in the recruitment of participants. R.J.S. consults for Highland Therapeutics, Purdue Pharma, and Ehave, and is the Toronto Dominion Bank Financial Group Chair in Child and Adolescent Psychiatry. The remaining authors have declared that they have no competing or potential conflicts of interest.

Correspondence

Paul D. Arnold, Mathison Centre for Mental Health Research and Education, 4th floor, Teaching, Research and Wellness (TRW) Building, University of Calgary, 3280 Hospital Dr. NW, Calgary AB T2N 4Z6, Canada; Email: paul.arnold@ucalgary.ca

Key points

- OCD is a heterogeneous disorder with several overlapping symptom dimensions and sex differences reflected in the general population.
- *5-HTTLPR* variant [LG + S] is significantly associated with hoarding in pediatric males in the community.
- *HTR2A* SNPs are significantly associated with rumination in males.
- Obsessive-compulsive trait dimensions may have distinct underlying serotonin system genetics and the exact associated gene variation may be sex-specific.
- Our findings emphasize that heterogeneity in OCD should be properly accounted for using homogenous subgroups reflecting symptom dimension, sex, and age group.

References

Alonso, P., Gratacos, M., Segalas, C., Escaramis, G., Real, E., Bayes, M., ... & Menchon, J.M. (2012). Association between the NMDA glutamate receptor GRIN2B gene and obsessivecompulsive disorder. *Journal of Psychiatry and Neuroscience: JPN*, 37, 273–281.

Alvarenga, P.G., Cesar, R.C., Leckman, J.F., Moriyama, T.S., Torres, A.R., Bloch, M.H., ... & do Rosario, M.C. (2015). Obsessive-compulsive symptom dimensions in a population-based, cross-sectional sample of schoolaged children. *Journal of Psychiatric Research*, 62, 108– 114.

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders (5th edn). Arlington, VA: Author.

Angst, J., Gamma, A., Endrass, J., Goodwin, R., Ajdacic, V., Eich, D., & Rossler, W. (2004). Obsessive-compulsive severity spectrum in the community: Prevalence, comorbidity, and course. *European Archives of Psychiatry and Clinical Neuroscience*, 254, 156–164.

Billett, E.A., Richter, M.A., King, N., Heils, A., Lesch, K.P., & Kennedy, J.L. (1997). Obsessive compulsive disorder, response to serotonin reuptake inhibitors and the serotonin transporter gene. *Molecular Psychiatry*, 2, 403–406.

Blier, P., de Montigny, C., & Chaput, Y. (1990). A role for the serotonin system in the mechanism of action of antidepressant treatments: Preclinical evidence. *The Journal of Clinical Psychiatry*, 51(Suppl), 14–20; discussion 21.

Bloch, M.H., Landeros-Weisenberger, A., Rosario, M.C., Pittenger, C., & Leckman, J.F. (2008). Meta-analysis of the symptom structure of obsessive-compulsive disorder. *The American Journal of Psychiatry*, 165, 1532–1542.

Burton, C.L., Crosbie, J., Dupuis, A., Mathews, C.A., Soreni, N., Schachar, R., & Arnold, P.D. (2016). Clinical correlates of hoarding with and without comorbid obsessive-compulsive symptoms in a community pediatric sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55, 114–121. e2.

Burton, C.L., Crosbie, J., Erdman, L., Dupuis, A., Park, L.S., Sinopoli, V., ... & Arnold, P.D.. (2015) A hypothesis-driven genome-wide association study of an obsessive-compulsive quantitative trait in a community-based sample of children and adolescents. Oral talk at the 23nd World Congress of Psychiatric Genetics Meeting October 2015, Toronto, ON.

Burton, C.L., Park, L.S., Corfield, E.C., Forget-Dubois, N., Dupuis, A., Sinopoli, V.M., ... & Arnold, P.D. (2018). Heritability of obsessive-compulsive trait dimensions in youth from the general population. *Translational Psychiatry*, 8, 191.

Cavallini, M.C., Di Bella, D., Siliprandi, F., Malchiodi, F., & Bellodi, L. (2002). Exploratory factor analysis of obsessivecompulsive patients and association with 5-HTTLPR polymorphism. *American Journal of Medical Genetics*, 114, 347– 353.

Crosbie, J., Arnold, P., Paterson, A., Swanson, J., Dupuis, A., Li, X., ... & Schachar, R.J. (2013). Response inhibition and ADHD traits: Correlates and heritability in a community sample. *Journal of Abnormal Child Psychology*, 41, 497–507.

Delaneau, O., & Marchini, J., & 1000 Genomes Project Consortium. (2014). Integrating sequence and array data to create an improved 1000 genomes project haplotype reference panel. *Nature Communications*, 5, 3934.

Flament, M. (1990). Epidemiology of obsessive-compulsive disorder in children and adolescents. [Epidemiologie du trouble obsessionnel-compulsif chez l'enfant et l'adolescent] *L'Encephale*, 16 Spec No, 311–316.

Gelernter, J., Kranzler, H., & Cubells, J.F. (1997). Serotonin transporter protein (SLC6A4) allele and haplotype frequencies and linkage disequilibria in african- and europeanamerican and japanese populations and in alcohol-dependent subjects. *Human Genetics*, 101, 243–246.

Geller, D.A., & March, J., The AACAP Committee on Quality Issues (CQI). (2012). Practice parameter for the assessment and treatment of children and adolescents with obsessivecompulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51, 98–113.

Hasler, G., Kazuba, D., & Murphy, D.L. (2006). Factor analysis of obsessive-compulsive disorder YBOCS-SC symptoms and association with 5-HTTLPR SERT polymorphism. *American Journal of Medical Genetics Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 141B, 403–408.

Heils, A., Teufel, A., Petri, S., Stober, G., Riederer, P., Bengel, D., & Lesch, K.P. (1996). Allelic variation of human serotonin transporter gene expression. *Journal of Neurochemistry*, 66, 2621–2624.

Howie, B.N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, 5, e1000529.

Hu, X.Z., Lipsky, R.H., Zhu, G., Akhtar, L.A., Taubman, J., Greenberg, B.D., ... & Goldman, D. (2006). Serotonin transporter promoter gain-of-function genotypes are linked to obsessive-compulsive disorder. *American Journal of Human Genetics*, 78, 815–826.

Katerberg, H., Delucchi, K.L., Stewart, S.E., Lochner, C., Denys, D.A., Stack, D.E., ... & Cath, D.C. (2010). Symptom dimensions in OCD: Item-level factor analysis and heritability estimates. *Behavior Genetics*, 40, 505–517.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., & Walters, E.E. (2005). Lifetime prevalence and age-ofonset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62, 593–602.

Kim, S.J., Lee, H.S., & Kim, C.H. (2005). Obsessive-compulsive disorder, factor-analyzed symptom dimensions and serotonin transporter polymorphism. *Neuropsychobiology*, 52, 176–182.

Landeros-Weisenberger, A., Bloch, M.H., Kelmendi, B., Wegner, R., Nudel, J., Dombrowski, P., ... & Coric, V. (2010). Dimensional predictors of response to SRI pharmacotherapy in obsessive-compulsive disorder. *Journal of Affective Disorders*, 121, 175–179.

Lesch, K.P., Bengel, D., Heils, A., Sabol, S.Z., Greenberg, B.D., Petri, S., ... & Murphy, D.L. (1996). Association of anxietyrelated traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* (New York, NY), 274, 1527–1531.

Li, M.X., Yeung, J.M., Cherny, S.S., & Sham, P.C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics*, 131, 747–756.

Mataix-Cols, D., & Pertusa, A. (2012). Annual research review: Hoarding disorder: Potential benefits and pitfalls of a new mental disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 53, 608–618.

Mattina, G.F., & Steiner, M. (2016). The need for inclusion of sex and age of onset variables in genetic association studies of obsessive-compulsive disorder: Overview. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 67, 107–116.

Millan, M.J., Goodwin, G.M., Meyer-Lindenberg, A., & Ove Ogren, S. (2015). Learning from the past and looking to the future: Emerging perspectives for improving the treatment of psychiatric disorders. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, 25, 599–656.

Murphy, D.L., Lerner, A., Rudnick, G., & Lesch, K.P. (2004). Serotonin transporter: Gene, genetic disorders, and pharmacogenetics. *Molecular Interventions*, 4, 109–123. Murphy, D.L.,

Moya, P.R., Fox, M.A., Rubenstein, L.M., Wendland, J.R., & Timpano, K.R. (2013). Anxiety and affective disorder comorbidity related to serotonin and other neurotransmitter systems:

Obsessive-compulsive disorder as an example of overlapping clinical and genetic heterogeneity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368, 20120435.

National Institute of Mental Health (2008). National Institute of Mental Health strategic plan. Available from: http:// www.nimh.nih.gov/about/strategic-planningreports/nimh_ strategicplanforresearch_508compliant_corrected_final_ 149979.pdf [last accessed 21 May 2019].

Park, L.S., Burton, C.L., Dupuis, A., Shan, J., Storch, E.A., Crosbie, J., Schachar, R.J., & Arnold, P.D. (2016). The Toronto Obsessive-Compulsive Scale: Psychometrics of a dimensional measure of obsessive-compulsive traits. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55, 310–318. e4.

Pauls, D.L., Abramovitch, A., Rauch, S.L., & Geller, D.A. (2014). Obsessive-compulsive disorder: An integrative genetic and neurobiological perspective. *Nature Reviews, Neuroscience*, 15, 410–424.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38, 904–909.

Ruscio, A.M., Stein, D.J., Chiu, W.T., & Kessler, R.C. (2010). The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry*, 15, 53e63.

Sinopoli, V.M., Burton, C.L., Kronenberg, S., & Arnold, P.D. (2017). A review of the role of serotonin system genes in obsessive-compulsive disorder. *Neuroscience and Biobehavioral Reviews*, 80, 372–381.

Snowdon, J., Pertusa, A., & Mataix-Cols, D. (2012). On hoarding and squalor: A few considerations for DSM-5. *Depression and Anxiety*, 29, 417–424.

Taylor, S. (2013). Molecular genetics of obsessive-compulsive disorder: A comprehensive meta-analysis of genetic association studies. *Molecular Psychiatry*, 18, 799–805.

Taylor, S. (2016). Disorder-specific genetic factors in obsessive-compulsive disorder: A comprehensive meta-analysis. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 171B, 325–332.

Van Ameringen, M., Patterson, B., & Simpson, W. (2014). DSM-5 obsessive-compulsive and related disorders: Clinical implications of new criteria. *Depression and Anxiety*, 31, 487–493.

van der Sluis, S., Posthuma, D., Nivard, M.G., Verhage, M., & Dolan, C.V. (2013). Power in GWAS: Lifting the curse of the clinical cut-off. *Molecular Psychiatry*, 18, 2–3.

Van Grootheest, D.S., Boomsma, D.I., Hettema, J.M., & Kendler, K.S. (2008). Heritability of obsessive-compulsive symptom dimensions. *American Journal of Medical Genetics. Part B,*

Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics, 147B, 473– 478.

Van Grootheest, D.S., Cath, D.C., Beekman, A.T., & Boomsma, D.I. (2005). Twin studies on obsessive-compulsive disorder: A review. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, 8, 450– 458.

Voyiaziakis, E., Evgrafov, O., Li, D., Yoon, H.J., Tabares, P., Samuels, J., , ... & Knowles, J.A. (2011). Association of SLC6A4 variants with obsessive-compulsive disorder in a large multicenter US family study. *Molecular Psychiatry*, 16, 108–120.

Wendland, J.R., Martin, B.J., Kruse, M.R., Lesch, K.P., & Murphy, D.L. (2006). Simultaneous genotyping of four functional loci of human SLC6A4, with a reappraisal of 5- HTTLPR and rs25531. *Molecular Psychiatry*, 11, 224–226

FIGURES

Figure 1

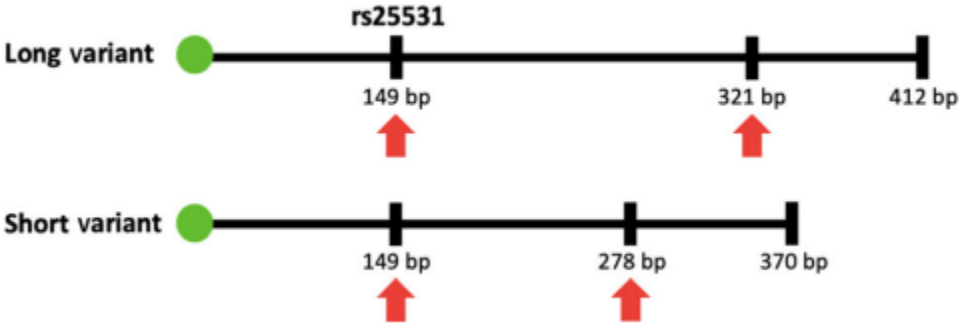

Figure 1 5-HTTLPR; L and S variants and rs25531: Green circles represent fluorescent probes. Red arrows represent where restriction enzyme, MspI, cuts the DNA at its appropriate recognition sequence if present. Upon digestion, the following sizes of fluorescently probed fragments are identified: LA = 321 bp, SA = 278 bp, LG and rare SG = 149 bp. This information is combined with the L versus S information from the original PCR to distinguish between LG and SG. 5-HTTLPR, 5-HTTlinked polymorphic region; L, long variant; S, short variant; LA, long variant with single nucleotide polymorphism A at rs25531; SA, short variant with single nucleotide polymorphism A at rs25531; LG, long variant with single nucleotide polymorphism G at rs25531; SG, short variant with single nucleotide polymorphism G at rs25531; PCR, polymerase chain reaction

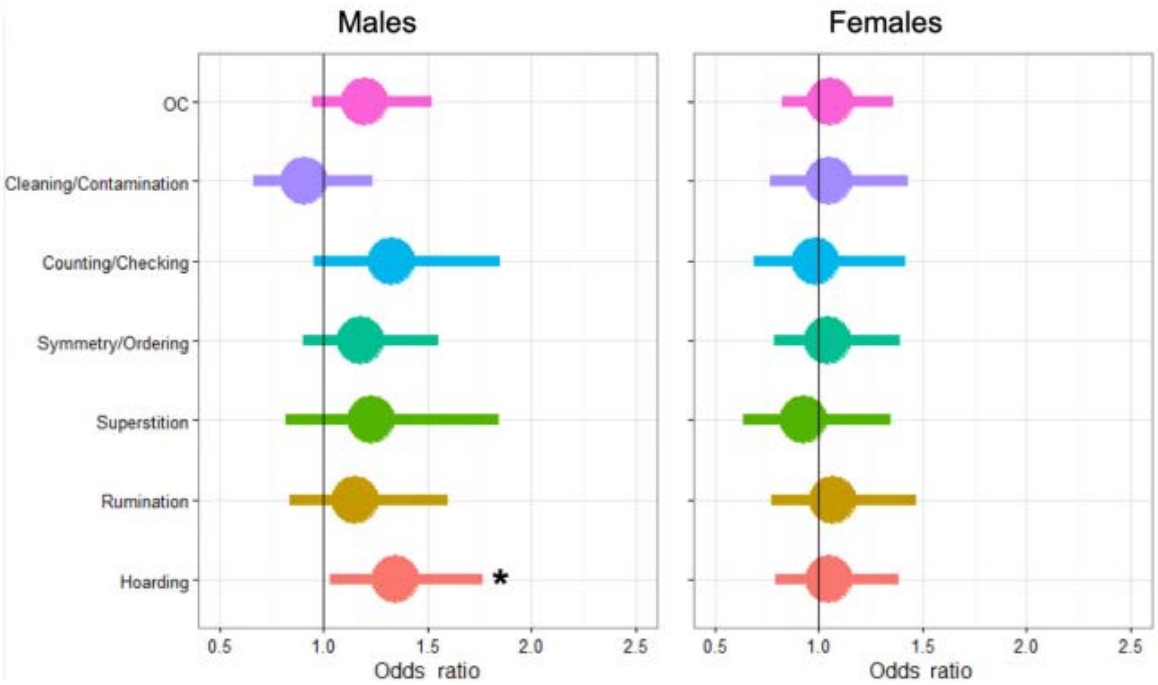

Figure 2. Association between 5-HTTLPR and OC traits overall/OC trait dimensions for the dichotomized trait analyses: Forest plots for males (left) and females (right). Colored points show the odds ratios for the [LG + S] allele, representing the effect size estimate for the allele in each trait group relative to controls. The length of each horizontal line is the confidence interval for each effect size estimate. Significant finding is indicated by *. OC, obsessive-compulsive; 5-HTTLPR, 5-HTT-linked polymorphic region