DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Genome-Wide Association Study in Obsessive-Compulsive Disorder: Results from the OCGAS

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

Citation	Mattheisen, M., J. F. Samuels, Y. Wang, B. D. Greenberg, A. J. Fyer, J. T. McCracken, D. A. Geller, et al. 2014. "Genome-Wide Association Study in Obsessive-Compulsive Disorder: Results from the OCGAS." Molecular psychiatry 20 (3): 337-344. doi:10.1038/mp.2014.43. http://dx.doi.org/10.1038/mp.2014.43.
Published Version	doi:10.1038/mp.2014.43
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:22856935
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Genome-Wide Association Study in Obsessive-Compulsive Disorder: Results from the OCGAS

Manuel Mattheisen, M.D[1,2,3,*], **Jack F. Samuels, Ph.D.**[4], **Ying Wang, M.S.**[4], **Benjamin D. Greenberg, M.D., Ph.D.**[5], **Abby J. Fyer, M.D.**[6], **James T. McCracken, M.D.**[7], **Daniel A. Geller, M.D.**[8], **Dennis L. Murphy, M.D.**[9], **James A. Knowles, M.D., Ph.D.**[10], **Marco A. Grados, M.D.**[4], **Mark A. Riddle, M.D.**[4], **Steven A. Rasmussen, M.D.**[5], **Nicole C. McLaughlin, Ph.D.**[5], **Erica Nurmi, M.D., Ph.D.**[7], **Kathleen D. Askland, M.D.**[5], **Hai-De Qin, Ph.D.**[11], **Bernadette A. Cullen, M.D.**[4], **John Piacentini, Ph.D.**[7], **David L. Pauls, Ph.D.**[8], **O. Joseph Bienvenu, M.D., Ph.D.**[4], **S. Evelyn Stewart, M.D.**[8,12], **Kung-Yee Liang, Ph.D.**[13], **Fernando S. Goes, M.D.**[4], **Brion Maher, Ph.D.**[13], **Ann E. Pulver, Sc.D.**[4], **Yin-Yao Shugart, Ph.D.**[11], **David Valle, M.D.**[14], **Cristoph Lange, Ph.D.**[2,3], and **Gerald Nestadt, M.D.**[4,*]

[1]Department of Biomedicine and Center for Integrated Sequencing (iSEQ), Aarhus University, Aarhus, Denmark

[2]Harvard School of Public Health, Department of Biostatistics, Boston, MA, USA

[3]Department of Genomic Mathematics, University of Bonn, Bonn, Germany

[4]Johns Hopkins University School of Medicine, Department of Psychiatry and Behavioral Sciences, Baltimore, MD, USA

[5]Brown Medical School, Department of Psychiatry and Human Behavior, Providence, RI, USA

[6]College of Physicians and Surgeons at Columbia University, New York State Psychiatric Institute, New York, NY, USA

[7]University of California, Los Angeles School of Medicine, Department of Psychiatry and Biobehavioral Sciences, Los Angeles, CA, USA

[8]Massachusetts General Hospital and Harvard Medical School, Department of Psychiatry, Boston, MA, USA

[9]National Institute of Mental Health, Laboratory of Clinical Science, Bethesda, MD, USA

[10]Keck School of Medicine at the University of Southern California, Department of Psychiatry and Behavioral Sciences, Los Angeles, CA, USA

[11]National Institute of Mental Health, Unit of Statistical Genomics, Intramural Research Program, Division of Intramural Research Program, Bethesda, MD, USA

[12]University of British Columbia, Department of Psychiatry, Vancouver, BC, Canada

*Corresponding authors: Dr. Gerald Nestadt, Johns Hopkins University, 600 N. Wolfe St. - Meyer 113, Baltimore, Maryland 21287 (gnestadt@jhmi.edu; phone +1 410 955 4838) and Dr. Manuel Mattheisen, Department of Biomedicine, Aarhus University, Wilhjelm Meyers Alle 4, 8000 Aarhus C, Denmark (mm@hum-gen.au.dk; phone +45 87 16 84 97).

[13]Johns Hopkins University Bloomberg School of Public Health, Department of Mental Health, Baltimore, MD, USA

[14]Johns Hopkins University School of Medicine, Institute of Human Genetics, Baltimore, MD, USA

Abstract

Obsessive-compulsive disorder (OCD) is a psychiatric condition characterized by intrusive thoughts and urges and repetitive, intentional behaviors that cause significant distress and impair functioning. The OCD Collaborative Genetics Association Study (OCGAS) is comprised of comprehensively assessed OCD patients, with an early age of OCD onset. After application of a stringent quality control protocol, a total of 1 065 families (containing 1 406 patients with OCD), combined with population-based samples (resulting in a total sample of 5 061 individuals), were studied. An integrative analyses pipeline was utilized, involving association testing at SNP- and gene-levels (via a hybrid approach that allowed for combined analyses of the family- and population-based data). The smallest P-value was observed for a marker on chromosome 9 (near *PTPRD*, $P=4.13\times10^{-7}$). Pre-synaptic PTPRD promotes the differentiation of glutamatergic synapses and interacts with SLITRK3. Together, both proteins selectively regulate the development of inhibitory GABAergic synapses. Although no SNPs were identified as associated with OCD at genome-wide significance level, follow-up analyses of GWAS signals from a previously published OCD study identified significant enrichment ($P=0.0176$). Secondary analyses of high confidence interaction partners of DLGAP1 and GRIK2 (both showing evidence for association in our follow-up and the original GWAS study) revealed a trend of association ($P=0.075$) for a set of genes such as *NEUROD6*, *SV2A*, *GRIA4*, *SLC1A2, and PTPRD*. Analyses at the gene-level revealed association of *IQCK* and *C16orf88* (both $P<1\times10^{-6}$, experiment-wide significant), as well as *OFCC1* ($P=6.29\times10^{-5}$). The suggestive findings in this study await replication in larger samples.

Keywords

Obsessive-Compulsive; protein tyrosine phosphatase delta; CDH10; CDH9; schizophrenia

Introduction

Obsessive-compulsive disorder (OCD) is a psychiatric condition characterized by persistent, intrusive, senseless thoughts and urges (obsessions) and repetitive, intentional behaviors (compulsions). Affected individuals tend to recognize that their thoughts and behaviors are excessive and unreasonable, and often struggle to resist them. The lifetime prevalence of OCD is estimated to be between 1–3%, based on national and international population-based surveys [1, 2]. Patients experience a chronic or episodic course with exacerbations that can substantially impair social, occupational, and academic functioning; according to the World Health Organization, OCD is among the most disabling medical conditions worldwide [3].

Since the early twentieth century, clinicians have suspected that heredity plays an important role in OCD susceptibility. Consistent with this, increased OCD prevalence was identified among first-degree relatives of case probands (11.7%), compared to relatives of controls

(2.7%; the Hopkins OCD family study) [4]. In the same study, the prevalence of OCD in siblings of early onset probands was 17.9% (λsib =7.8) and thus within the range found in other psychiatric disorders for which a heritable component was reported, such as bipolar disorder and panic disorder. Family studies of child and adolescent OCD probands report prevalence rates of 7% to 15% in first-degree relatives, and 20–30% of these probands had one or more first-degree relative with OCD [5]. These findings are consistent with previous reports of an increased familial loading in probands with early age at onset. In addition, a recent population-based study suggested that (while there is no significant effect of a shared environment on OCD risk) the contribution of genetic factors to OCD risk is as high as 50% [6].

Two genetic linkage studies of OCD have found a suggestive linkage peak on chromosomal region 9p24 [7, 8]. The OCD Collaborative Genetics Study (OCGS) also found suggestive genetic linkage peaks on chromosomal regions 3q27-28, 6q, 7p, 1q, and 15q, but none were genome-wide significant [9].

There have been numerous candidate gene association studies in OCD [10]. A positional candidate of particular interest is the *neuronal and epithelial glutamate transporter* gene (*SLC1A1*). It is located in the region of the replicated linkage peak on chromosome 9p24, and there is considerable evidence from imaging, animal, and treatment studies that abnormal glutamatergic transmission may be involved in the pathophysiology of OCD [11, 12]. Several studies have reported positive associations of OCD with this gene, but with different SNPs in each [13–19]. A recent meta-analysis of these studies did not resolve the issue and was not able to report experiment-wise significance for SNPs annotated to *SLC1A1* [20].

A genome-wide association study was recently completed by the International OCD Foundation Genetic Collaborative (IOCDF-GC) [21] with a combined study sample from 22 sites, comprising 1 465 cases and 5 557 ancestry-matched controls, as well as 400 complete trios. In case-control analyses, the lowest *P*-values were found for two *SNPs* in perfect LD (rs11081062 and rs11663827; $r^2=1$, D′=1 in reference samples from the 1000 genomes project [22]; $P=2.49\times10^{-6}$ and $P=3.44\times10^{-6}$) Both were located on chromosome 18 within the *discs, large (Drosophila) homolog-associated protein 1 gene (DLGAP1)*, a member of the neuronal postsynaptic density complex. In the trio analysis, a SNP near the *BTB (POZ) domain containing 3 gene (BTBD3;* rs6131295, chromosome 20), exceeded the genome-wide significance threshold with a *P*-value=3.84×10^{-8}. However, when trios were meta-analyzed together with the case-control samples, the *P*-value for this variant was 3.62×10^{-5}, losing genome-wide significance. Although no SNPs were found to be associated with OCD at a genome-wide significant level in the combined trio–case–control sample, a significant enrichment of methylation QTLs ($P<0.001$) and frontal lobe expression quantitative trait loci (eQTLs) ($P=0.001$) was observed within the top-ranked SNPs ($P<0.01$) from the trio–case–control analysis, potentially pointing to a broad role in gene expression in the brain, and possibly in the etiology of OCD.

In this paper, we report the results of the most recent genome-wide association study of OCD, the OCD Collaborative Genetics Association Study (OCGAS). Investigators at 8 research centers in the United States, including Brown University, Columbia University,

Johns Hopkins University, Massachusetts General Hospital, Harvard School of Public Health, the National Institute of Mental Health, University of California Los Angeles, and the University of Southern California, conducted this collaborative study. The aim of the study was to identify common variants associated with OCD using an integrative analyses pipeline that involved association testing at both SNP and gene levels. The approach included OCD-affected patients (in a family-based analytic setting) with an early onset of the disorder (eighteen years and younger) who had been comprehensively assessed via an identical diagnostic approach at all recruitment centers.

Materials and methods

Sample

A total of 1 065 families were included in this study (comprising 1 406 patients with OCD and 2 895 individuals in total); 621 families were recruited and assessed specifically for this study at one of the five participating recruitment sites or the National Institute of Mental Health; 444 families had previously been evaluated in one of the earlier studies at Hopkins University or by one of the collaborating sites. The sample comprised of 460 complete trios (including an affected proband and both parents); 155 pedigrees with a proband and an unaffected sibling and 450 families with another structure (complex family structure). An additional 192 probands without an additional family member present in the study (singletons) were included. A breakdown of the families by site is found in Supplementary Table S1.

For study inclusion, probands were required to meet DSM-IV criteria for OCD [23] with onset of obsessions and/or compulsions before the age of 18 years (mean = 9.4 years; SD=6.35). Subjects disease, schizophrenia, severe mental retardation that does not permit an evaluation to characterize the psychiatric disorder, Tourette disorder (TS), or OCD occurring exclusively in the context of depression (secondary OCD) were excluded. In addition, individuals were removed from the sample if they were previously diagnosed with brain pathology including brain tumors, Huntington's Disease, Parkinson's Disease, or Alzheimer's Disease. Each case was evaluated by a PhD-level clinical psychologist using the Structured Clinical Interview for DSM-IV (SCID) modified and extended to include additional symptom and diagnostic information as indicated in the supplementary material. Final diagnostic status was assigned based on the consensus of two psychiatrists or psychologists reviewing the case independently. Both parents of the proband were also recruited whenever possible. When parents were unavailable for participation, unaffected siblings were recruited. Genotyping was performed at the Johns Hopkins SNP Center using Illumina's HumanOmniExpress bead chips (Illumina Inc., San Diego, CA, USA). More details on the diagnostic assessment and the genotyping process are provided in the supplementary materials and methods.

To increase the power of the study to detect significant association, we also included 1 984 unrelated controls (genotyped with Illumina's HumanOMNI1-QUAD bead chip) from a previously published study on Parkinson's disease (dbGaP accession number phs000196.v2.p1) [24, 25]. Individuals with a self-reported or diagnosed neuropsychiatric disorder at the time of enrollment were excluded from the present study.

Quality control of genotyping data

We followed a stringent quality control (QC) protocol that was designed to minimize occurrence of false positive signals, which included checking the relatedness of samples (i.e. verifying the relationship reported from the participating clinical centers) and reported sex (based on deviations from expected heterozygosity rates based on x-chromosomal markers in the analysis; PLINK standard parameters were used). In addition, evidence for genotyping errors at the sample / marker level was evaluated by searching for an excess of "Mendelian inconsistencies" in the data (as indicated by substantial deviation from the empirical distributions in both measurements). Multidimensional scaling (MDS) analyses were performed on singleton OCD cases and unselected controls, as implemented into Plink [26]. Samples were removed when they significantly deviated in the first two MDS dimensions (> 4 standard deviation from the mean). Although inherently robust against population stratification, substantial heterogeneity in the population-based study cohorts would have lead to decreased power for our analytical approach (see below). More details on the QC process are provided in the supplementary materials and methods section. Use of the described filters resulted in a final analysis dataset comprising 5 061 individuals from the OCGAS samples and additional controls, with available genotypes for 549 123 autosomal markers.

Statistical analyses

For the single-marker association analyses, we used a recently published method for combining family-based and population-based data. More precisely, family-based association testing of within-family information is combined with population-based analyses of between-family information and the association analyses of unrelated study subjects [27]. This hybrid approach is inherently robust to population stratification and potentially increases statistical power compared to a classic meta-analysis design [27]. As part of our analytical pipeline we therefore computed *P*-values for all autosomal markers and both the within and the between family information using PBAT [28]. The two *P*-values were subsequently combined using a weighted Z-score statistic as implemented via METAL [29].

Gene-based statistics were derived as fixed Z scores, as implemented in FORGE [30]. More information on the approach is provided elsewhere [31]. Here we used the single-marker results of our combined OCGAS GWAS. Information about the correlation pattern in the data was provided through usage of HapMap phase 2 samples [32]. A maximum of 1 000 000 permutations were used per gene (adaptive approach) and analyses included an additional +/− 20kb sequence information based on positions obtained from ENSEMBL v70 [33]. We used the gene-based results in two ways: First we used them to agnostically search for genes that are associated with OCD. Second, we used them to follow-up on gene-set based results from the IOCDF-GC study, which reported an enrichment of association signals for two gene sets that comprised high confidence targets of two miRNA families [21]. In addition, we used information from a global interactome for Homo sapiens in order to identify high confidence interactors of DLGAP1 and GRIK2 (confidence threshold > 0.95) [34]. No sub-network reduction was applied and only genes representing first neighbors of *DLGAP1* and/or *GRIK2* in the global interactome were considered for this analysis. We used gene-based results in order to identify an enrichment of association with OCD in interaction partners of

DLGAP1 and the ionotropic kainate 2 glutamate receptor (GRIK2). This analysis was motivated by the assumption that biologically closely related genes of these previously described OCD risk genes would make reasonable candidates for hypothesis driven downstream analyses. Based on an assumed similarity between these genes with regards to an involvement in common biological processes and provided that, on a broader level, these biological processes themselves are associated with the phenotype under study it seems reasonable to hypothesize that focusing on biologically closely related genes (see above) helps to identify new disease genes (through reduction of some of the multiple testing burden in standard GWAS). More information on the usage of the interactome data and the visualization of the resulting networks are found in the supplementary material and methods. In brief, we would like to emphasize that our analyses represent a simple, hypothesis driven approach for inclusion of the human interactome data, rather than an exhaustive network science approach (making use among others of topological features of the interactome other then the status of direct interaction). While the latter has been successfully used with neuropsychiatric traits in the past, we felt that this evolving field still needs some further improvements and therefore decided to refrain from these kinds of analyses.

Results

Single SNP association

Analysis of the 549 123 autosomal markers revealed a nominally significant result at the level of α=0.05 for 27 283 markers. Among these, 456 markers reached a *P*-value of less than 0.001 and 54 markers were identified with a *P*<0.0001. As expected due to the stringent study design, no evidence for population stratification was observed (see also Supplementary Figure S1). Figure 1 shows the 'Manhattan' plot of the association results across the autosomal genome and Table 1 reports the association between single SNPs and OCD with $P<1\times10^{-5}$. No marker tested for association met the standard of genome-wide significance. The smallest *P*-value for our study was observed with a marker on chromosome 9 (rs4401971) at a *P*-value of 4.13×10^{-7}. This SNP is 1.28 Mb from the 5′ end of the *protein tyrosine phosphatase, receptor type, D* gene (*PTPRD,* isoform 1; according to RefSeq). The second most significantly associated SNP (rs6876547, $P=1.76\times10^{-6}$) is located in a region of cadherin clusters and it is of note, that a second, independent SNP (rs6452234, $P=1.13\times10^{-5}$, $r^2<0.2$, distance ~650kb) is located in the same region. The nearest flanking protein coding genes are the *cadherin 9, type 2* (*CDH9*, 1 308 kb) and the *cadherin 10, type 2* (*CDH10*, 927 kb) genes. A 'regional association plot' of these two top regions is provided in Figure 2.

We used the single-maker association signals to follow up the results from the recent IOCDF-GC GWAS [21]. For the markers listed in Table 1 of that study, we found evidence for association with a marker (rs2205748, same effect allele and direction of effect) on chromosome 6 near the genes *GRIK2* (distance to gene 1.94 Mb) and the *HECT domain and ankyrin repeat containing E3 ubiquitin protein ligase 1* gene (*HACE1,* 713 kb) with an uncorrected *P*-value of 0.0493 in our study ($P=8.52\times10^{-6}$ in the original IOCDF-GC trio-case-control analysis). One additional region (harboring the gene for the *cAMP-specific phosphodiesterase 4D*, *PDE4D*) showed a nominal significant signal; however, the effect

(based on the same effect allele) was in the opposite direction (compared with the IOCDF-GC finding). Though no other marker among the tested SNPs from the IOCDF-GC GWAS reached a significant result, it is of note that, for the region around *DLGAP1* an independent SNP in our analyses (rs3866988) yielded a *P*-value of 2.67×10^{-4} and thus indicating this region as a susceptibility factor for OCD (r^2=0.023, D′=0.316 between rs3866988 and rs11081062/ rs11663827, based on 1 000 genomes data). It is also of note, that 12 out of 15 markers (counting the *DLGAP1* markers in perfect LD only once) show the same direction of effects for the effect alleles in both the IOCDF-GC and OCGAS analyses (binominal test with probability 0.5 = sign test: *P*=0.0176). Detailed results for the follow-up of the IOCDF-GC results are found in Table 2.

Gene-based findings

Gene-based analyses for 21 567 genes (protein-coding genes and miRNAs) resulted in an experiment-wise significant result for two genes, the *IQ motif containing K* gene (*IQCK*) and the *chromosome 16 open reading frame 88* gene (*C16orf88;* P_{corr}<0.0215; P<1×10^{-6}). Both genes are located in the same chromosomal region (16p12.3) and share their leading SNP, i.e. the SNP with the lowest *P*-value that is annotated to the gene (rs1544352; *P*=1.94×10^{-5}). The only other gene that shows a *P*-value < 1×10^{-4} is the *orofacial cleft candidate1* gene (*OFCC1*; P=6.29×10^{-5}). A more detailed list of association results for the gene-based analyzes are given in Table 3 and a 'regional association plot' for the *OFCC1* locus is provided in Figure 2.

Query of the interactome [34] for high confidence interactors of *DLGAP1* and *GRIK2* identified 169 interactors of *DLGAP1* and 161 interactors for *GRIK2* (246 interactors for both of them together). Gene-based analyses revealed that 21 (out of 246) genes showed a nominal significant *P*-value (one-sided *P*=0.075). This included 16 genes for *DLGAP1* (one-sided *P*=0.069) and 14 genes for *GRIK2* (one-sided *P*=0.135). It is of note that *GRIK2* itself was identified as interactor of *DLGAP1* and showed a nominal significant gene-based *P*-value of 0.03. Among the list of other nominal significant genes are the *neuronal differentiation 6* gene (*NEUROD6, P*=0.010), the *synaptic vesicle glycoprotein 2A* gene (*SV2A, P*=0.026), the *ionotropic, AMPA 4 glutamate receptor* gene (*GRIA4, P*=0.039), and the solute carrier family 1 (glial high affinity glutamate transporter), member 2 gene (*SLC1A2, P*=0.035). A full list of these results is given in Supplementary Table S4.

We also attempted to replicate a reported enrichment of association signals for two gene sets that comprised high confidence targets of two miRNA families (see the Stewart et al OCD GWAS). Analyses of the miRNA families of predicted miR-130ac/301ab/301b/301b-3p/454/721/4295/3666 (miRNA set 1) and miRNA-219-5p/508/508-3p/4782-3p (miRNA set 2) targets showed nominal significance for 3 out of 45 high confidence targets in miRNA set 1 and 14 out of 145 high confidence targets in miRNA set 2. The number of nominal significant genes for both sets are higher then expected by chance; however, for both sets, no significant enrichment for OCD associated genes was found in either set (one-sided *P*=0.323 and one-sided *P*=0.087, respectively). It is of note that one of the high confident targets in miRNA set 2 (the *epiregulin* gene, *EREG*) did demonstrate an experiment-wise significance

level (correcting for the number of tests performed in miRNA set2; P=3.06×10^{-4}; P_{corr}=0.044).

Discussion

Here we report the results of a GWAS including a sample of 1 406 patients with OCD, that was predominantly family-based, but which also included a case-control subsample to increase power (resulting in a total sample of 5 061 individuals). The study identified interesting candidate genes for OCD, but failed to detect any genome-wide significant findings. This is similar to what has been observed for other psychiatric phenotypes such as schizophrenia, for which genome-wide significant findings were only achieved after either starting with more samples in the discovery step [35, 36] or making use of large follow-up samples, that were orders of magnitude larger then the original discovery samples [37, 38]. We suspect that, with additional samples, the findings of genetic studies for OCD will be more robust, and our currently suggestive findings may reach genome-wide significance.

Of particular interest is the signal near the *PTPRD* gene (rs4401971), the most significant found in this study (P=4.13×10^{-7}). PTPRD is a member of the receptor protein tyrosine phosphatase family, which comprises transmembrane signaling molecules that regulate a variety of cellular processes including cell growth and differentiation [39]. Pre-synaptic *PTPRD* promotes the differentiation of glutamatergic synapses [40–43] and interacts with Slit and NTRK-like family member 3 (*SLITRK3*) which acts as a postsynaptic adhesion molecule. Together both proteins selectively regulate inhibitory GABAergic synapse development [44]. This is interesting because molecules in the same family, *SLITRK5* and *SLITRK1*, have been shown to be associated with TS and OCD, the former in a mouse model and the latter in a TS genetic study [45, 46].

Mice deficient in PTPRD show impairment in learning and memory tasks, especially spatial learning, and exhibit enhanced long-term potentiation (LTP), a form of activity dependent plasticity [47]. This is relevant to OCD, in that memory deficits have been reported for this condition [48]. PTPRD also has been shown to be associated with restless leg syndrome (RLS) [49] and hemizygous deletions were detected in four unrelated attention deficit hyperactivity disorder (ADHD) probands [50].

Our second strongest association finding is located in a region of cadherin clusters. It is known that CDH9 has been reported associated with OCD [51]. Furthermore, six SNPs between cadherin 10 (*CDH10*) and cadherin 9 (*CDH9*) were significantly (as well as in a replication sample) associated with autism spectrum disorders (ASD) [52] and was a top (although not significant) association signal in anorexia nervosa [53].

Analysis of previously identified GWAS hit regions for the IOCDF-GC study [21] revealed that 12 out of the 15 strongest signals in that study showed associations with the same direction of effects (based on the same effect alleles) in both the IOCDF-GC and OCGAS analyses (sign test P=0.0176). This observation supports the hypothesis that future collaborative studies (meta- and mega-analyses) will be able to identify OCD genes at the level of genome-wide significance. Support for association with OCD was particularly

found in two previously identified genetic regions (see above). The most interesting region harbors the *DLGAP1* gene. In the IOCDF-GC study signals in this gene were the top signals from the case-control analysis. Although the specific SNP from the former study was not found significant in this study, there was a prominent signal for an independent nearby marker (P=2.67×10^{-4}). This observation potentially serves as independent evidence for association of markers in this region. *DLGAP1* is a member of the neuronal postsynaptic density complex and is in the same family as the *DLGAP3* gene, which has been shown in a convincing way to be responsible for OCD-like behaviors in *DLGAP3* knockout mice models [54]. A marker in the second genetic region (harboring *GRIK2*), also a top signal in the IOCDF-GC GWAS showed nominal significance in our study (P=0.045). Although failing to reach experiment-wise significance it is of note that *DLGAP1* and *GRIK2* interact (FUNCOUP [34]). Following up these signals in a gene-set analysis for high confidence interaction partners of *DLGAP1* and *GRIK2* (identified using FUNCOUP [34]) showed a trend for association and pointed to a potential role of a set of *DLGAP1* and *GRIK2* interactors in the etiology of OCD, involving genes such as *NEUROD6*, *SV2A*, *GRIA4*, and *SLC1A2*. Interestingly *PTPRD* was part of this gene set. It is of note that for both *DLGAP1* and *GRIK2* we identified more then 160 interactors each, thus indicating they are highly connected nodes (hubs) in the interactome. Earlier findings have suggested that only genes that are essential in (early) development ("essential genes") tend to encode hub proteins, while the vast majority of disease genes are nonessential and show no tendency to encode hub proteins [55]. This might point to *DLGAP1* and GRIK2 as being essential genes in early (neuro-)development. However, more studies are warranted to confirm this hypothesis.

Gene-based association analyses identify significant signals in two genes, *IQCK* and *C16orf88*; however, this finding is not amenable to interpretation at this time. The next strongest gene-based signal in *OFCC1* is intriguing because a variant within the exon of this gene has been reported in seven affected individuals of family segregating OCD, TS, and ADHD [56].

This study employed careful and comprehensive phenotyping; MD and PhD-trained psychologists conducted all assessments and recruitment sites used the same assessment approach and all cases were reviewed at a single site (JHU) to ensure diagnostic replicability. Given the likely etiologic heterogeneity of OCD, the comprehensive phenotyping provides opportunities for future studies (through subtyping based e.g. on factor-analytical approaches).

A limitation of this study is that the power to detect significant association signals for small effect sizes is limited by the sample size. This has been clearly demonstrated in many GWAS efforts for other neuropsychiatric conditions (e.g., schizophrenia) in which studies of comparable size have failed to identify significant association signals. Yet, when sample sizes have increased (in some case 10 or more fold), then multiple significant signals were identified. Power was improved in this study by using a hybrid analytic approach that included both family and case-control samples in the analysis. In conclusion, while this OCD GWAS study did not identify a study-wide significant association finding, several of the strongest findings are particularly interesting. There are both plausible biologic hypotheses and prior genetic evidence, for either OCD or related conditions, for the two

most significant association findings, *PTPRD* and *CDH9/CDH10*. The finding that the signals in this GWAS concur with respect to allele and direction significantly with the top signals in a second reported OCD GWAS, suggests that similar genetic underpinnings of OCD are identified in both studies. Moreover, while not a replication in the full sense, findings for the genes *DLGAP1* and *GRIK2* suggest that they are both relevant candidates genes for OCD. The suggestive findings in this study await replication in larger samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Weissman MM, Bland RC, Canino GJ, Greenwald S, Hwu HG, Lee CK, et al. The cross national epidemiology of obsessive compulsive disorder. The Cross National Collaborative Group. The Journal of clinical psychiatry. 1994; 55 (Suppl):5–10. [PubMed: 8077177]

2. Karno, M.; Golding, J. Obsessive compulsive disorder. In: Robins, L.; Regier, D., editors. Psychiatric Disorders in America: The epidemiologic catchment area study. Free Press; New York: 1991. p. 204-219.

3. Murray, C.; Lopez, A. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Vol. 1. Harvard University Press; Cambridge: 1996.

4. Nestadt G, Samuels J, Riddle M, Bienvenu OJ 3rd, Liang KY, LaBuda M, et al. A family study of obsessive-compulsive disorder. Archives of general psychiatry. 2000; 57(4):358–363. [PubMed: 10768697]

5. Riddle MA, Scahill L, King R, Hardin MT, Towbin KE, Ort SI, et al. Obsessive compulsive disorder in children and adolescents: phenomenology and family history. Journal of the American Academy of Child and Adolescent Psychiatry. 1990; 29(5):766–772. [PubMed: 2228931]

6. Mataix-Cols D, Boman M, Monzani B, Ruck C, Serlachius E, Langstrom N, et al. Population-based, multigenerational family clustering study of obsessive-compulsive disorder. JAMA psychiatry. 2013; 70(7):709–717. [PubMed: 23699935]

7. 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. American journal of medical genetics. 2002; 114(5):541–552. [PubMed: 12116192]

8. Willour VL, Yao Shugart Y, Samuels J, Grados M, Cullen B, Bienvenu OJ 3rd, et al. Replication study supports evidence for linkage to 9p24 in obsessive-compulsive disorder. American journal of human genetics. 2004; 75(3):508–513. [PubMed: 15272418]

9. Shugart YY, Samuels J, Willour VL, Grados MA, Greenberg BD, Knowles JA, et al. Genomewide linkage scan for obsessive-compulsive disorder: evidence for susceptibility loci on chromosomes 3q, 7p, 1q, 15q, and 6q. Molecular psychiatry. 2006; 11(8):763–770. [PubMed: 16755275]

10. Nestadt G, Grados M, Samuels JF. Genetics of obsessive-compulsive disorder. The Psychiatric clinics of North America. 2010; 33(1):141–158. [PubMed: 20159344]

11. Pittenger C, Bloch MH, Williams K. Glutamate abnormalities in obsessive compulsive disorder: neurobiology, pathophysiology, and treatment. Pharmacology & therapeutics. 2011; 132(3):314–332. [PubMed: 21963369]

12. Wu K, Hanna GL, Rosenberg DR, Arnold PD. The role of glutamate signaling in the pathogenesis and treatment of obsessive-compulsive disorder. Pharmacology, biochemistry, and behavior. 2012; 100(4):726–735.

13. Arnold PD, Sicard T, Burroughs E, Richter MA, Kennedy JL. Glutamate transporter gene SLC1A1 associated with obsessive-compulsive disorder. Archives of general psychiatry. 2006; 63(7):769–776. [PubMed: 16818866]

14. Dickel DE, Veenstra-VanderWeele J, Cox NJ, Wu X, Fischer DJ, Van Etten-Lee M, et al. Association testing of the positional and functional candidate gene SLC1A1/EAAC1 in early-onset obsessive-compulsive disorder. Archives of general psychiatry. 2006; 63(7):778–785. [PubMed: 16818867]

15. Wang Y, Adamczyk A, Shugart YY, Samuels JF, Grados MA, Greenberg BD, et al. A screen of SLC1A1 for OCD-related alleles. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2010; 153B(2):675–679.

16. Samuels J, Wang Y, Riddle MA, Greenberg BD, Fyer AJ, McCracken JT, et al. Comprehensive family-based association study of the glutamate transporter gene SLC1A1 in obsessive-compulsive disorder. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2011; 156B(4):472–477.

17. Shugart YY, Wang Y, Samuels JF, Grados MA, Greenberg BD, Knowles JA, et al. A family-based association study of the glutamate transporter gene SLC1A1 in obsessive-compulsive disorder in 378 families. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2009; 150B(6):886–892.

18. Stewart SE, Platko J, Fagerness J, Birns J, Jenike E, Smoller JW, et al. A genetic family-based association study of OLIG2 in obsessive-compulsive disorder. Archives of general psychiatry. 2007; 64(2):209–214. [PubMed: 17283288]

19. Wendland JR, Moya PR, Timpano KR, Anavitarte AP, Kruse MR, Wheaton MG, et al. A haplotype containing quantitative trait loci for SLC1A1 gene expression and its association with obsessive-compulsive disorder. Archives of general psychiatry. 2009; 66(4):408–416. [PubMed: 19349310]

20. 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. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2013; 162B(4):367–379.

21. Stewart SE, Yu D, Scharf JM, Neale BM, Fagerness JA, Mathews CA, et al. Genome-wide association study of obsessive-compulsive disorder. Molecular psychiatry. 2013; 18(7):788–798. [PubMed: 22889921]

22. Consortium GP, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467(7319):1061–1073. [PubMed: 20981092]

23. American Psychiatric Association. Diagnostic and statistical manual of mental disorders : DSM-IV. 4. Vol. xxvii. American Psychiatric Association; Washington, DC: 1994. American Psychiatric Association. Task Force on DSM-IV; p. 886

24. Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nature genetics. 2010; 42(9):781–785. [PubMed: 20711177]

25. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. Nature genetics. 2007; 39(10):1181–1186. [PubMed: 17898773]

26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81(3):559–575. [PubMed: 17701901]

27. Lasky-Su J, Won S, Mick E, Anney RJ, Franke B, Neale B, et al. On genome-wide association studies for family-based designs: an integrative analysis approach combining ascertained family samples with unselected controls. American journal of human genetics. 2010; 86(4):573–580. [PubMed: 20346434]

28. Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM. PBAT: tools for family-based association studies. American journal of human genetics. 2004; 74(2):367–369. [PubMed: 14740322]

29. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26(17):2190–2191. [PubMed: 20616382]

30. Pedroso I, Breen G. Gene set analysis and network analysis for genome-wide association studies. Cold Spring Harbor protocols. 2011; 2011(9)

31. Pedroso I, Lourdusamy A, Rietschel M, Nothen MM, Cichon S, McGuffin P, et al. Common genetic variants and gene-expression changes associated with bipolar disorder are over-represented in brain signaling pathway genes. Biological psychiatry. 2012; 72(4):311–317. [PubMed: 22502986]

32. International HapMap C. The International HapMap Project. Nature. 2003; 426(6968):789–796. [PubMed: 14685227]

33. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics. 2010; 26(16): 2069–2070. [PubMed: 20562413]

34. Alexeyenko A, Schmitt T, Tjarnberg A, Guala D, Frings O, Sonnhammer EL. Comparative interactomics with Funcoup 2.0. Nucleic acids research. 2012; 40(Database issue):D821–828. [PubMed: 22110034]

35. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. Nature genetics. 2011; 43(10):969–976. [PubMed: 21926974]

36. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, et al. Common variants conferring risk of schizophrenia. Nature. 2009; 460(7256):744–747. [PubMed: 19571808]

37. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nature genetics. 2008; 40(9):1053–1055. [PubMed: 18677311]

38. Rietschel M, Mattheisen M, Degenhardt F, Genetic R, Outcome in P, Muhleisen TW, et al. Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. Molecular psychiatry. 2012; 17(9):906–917. [PubMed: 21747397]

39. Chagnon MJ, Uetani N, Tremblay ML. Functional significance of the LAR receptor protein tyrosine phosphatase family in development and diseases. Biochemistry and cell biology = Biochimie et biologie cellulaire. 2004; 82(6):664–675. [PubMed: 15674434]

40. Dunah AW, Hueske E, Wyszynski M, Hoogenraad CC, Jaworski J, Pak DT, et al. LAR receptor protein tyrosine phosphatases in the development and maintenance of excitatory synapses. Nature neuroscience. 2005; 8(4):458–467.

41. Woo J, Kwon SK, Choi S, Kim S, Lee JR, Dunah AW, et al. Trans-synaptic adhesion between NGL-3 and LAR regulates the formation of excitatory synapses. Nature neuroscience. 2009; 12(4): 428–437.

42. Kwon SK, Woo J, Kim SY, Kim H, Kim E. Trans-synaptic adhesions between netrin-G ligand-3 (NGL-3) and receptor tyrosine phosphatases LAR, protein-tyrosine phosphatase delta (PTPdelta), and PTPsigma via specific domains regulate excitatory synapse formation. The Journal of biological chemistry. 2010; 285(18):13966–13978. [PubMed: 20139422]

43. Takahashi H, Arstikaitis P, Prasad T, Bartlett TE, Wang YT, Murphy TH, et al. Postsynaptic TrkC and presynaptic PTPsigma function as a bidirectional excitatory synaptic organizing complex. Neuron. 2011; 69(2):287–303. [PubMed: 21262467]

44. Takahashi H, Katayama K, Sohya K, Miyamoto H, Prasad T, Matsumoto Y, et al. Selective control of inhibitory synapse development by Slitrk3-PTPdelta trans-synaptic interaction. Nature neuroscience. 2012; 15(3):389–398. S381–382.

45. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, et al. Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science. 2005; 310(5746):317–320. [PubMed: 16224024]

46. Shmelkov SV, 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. Nature medicine. 2010; 16(5):598–602. 591p following 602.

47. Uetani N, Kato K, Ogura H, Mizuno K, Kawano K, Mikoshiba K, et al. Impaired learning with enhanced hippocampal long-term potentiation in PTPdelta-deficient mice. The EMBO journal. 2000; 19(12):2775–2785. [PubMed: 10856223]

48. Jaafari N, Frasca M, Rigalleau F, Rachid F, Gil R, Olie JP, et al. Forgetting what you have checked: a link between working memory impairment and checking behaviors in obsessive-compulsive disorder. European psychiatry : the journal of the Association of European Psychiatrists. 2013; 28 (2):87–93. [PubMed: 21924871]

49. Schormair B, Kemlink D, Roeske D, Eckstein G, Xiong L, Lichtner P, et al. PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nature genetics. 2008; 40(8):946–948. [PubMed: 18660810]

50. 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. Molecular psychiatry. 2010; 15(6):637–646. [PubMed: 19546859]

51. Mas S, Plana MT, Castro-Fornieles J, Gasso P, Lafuente A, Moreno E, et al. Common genetic background in anorexia nervosa and obsessive compulsive disorder: preliminary results from an association study. Journal of psychiatric research. 2013; 47(6):747–754. [PubMed: 23337130]

52. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature. 2009; 459(7246):528–533. [PubMed: 19404256]

53. Wang K, Zhang H, Bloss CS, Duvvuri V, Kaye W, Schork NJ, et al. A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. Molecular psychiatry. 2011; 16(9):949–959. [PubMed: 21079607]

54. Welch JM, Lu J, Rodriguiz RM, Trotta NC, Peca J, Ding JD, et al. Cortico-striatal synaptic defects and OCD-like behaviours in Sapap3-mutant mice. Nature. 2007; 448(7156):894–900. [PubMed: 17713528]

55. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(21):8685–8690. [PubMed: 17502601]

56. Sundaram SK, Huq AM, Sun Z, Yu W, Bennett L, Wilson BJ, et al. Exome sequencing of a pedigree with Tourette syndrome or chronic tic disorder. Annals of neurology. 2011; 69(5):901–904. [PubMed: 21520241]
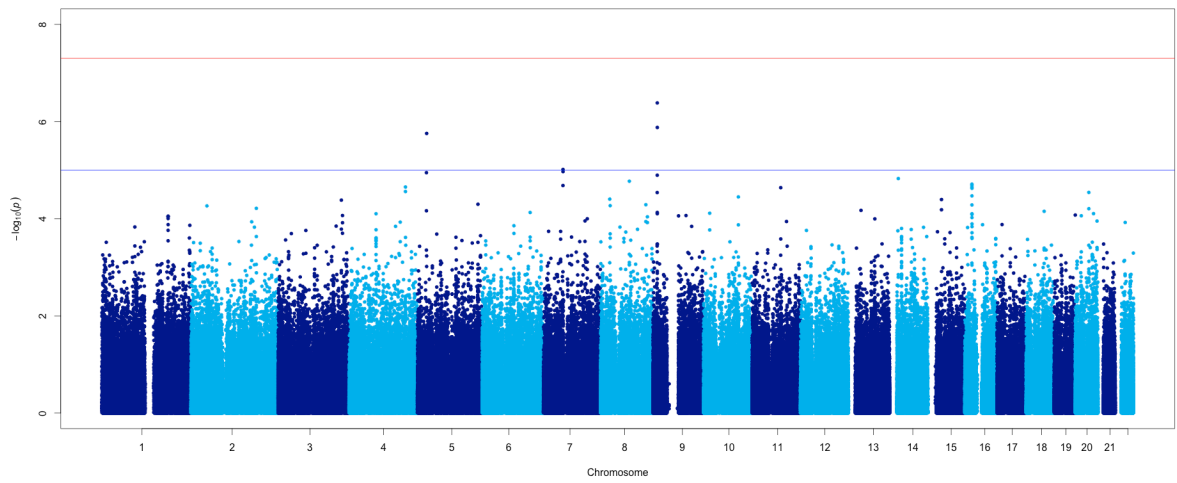
Figure 1. Manhattan plot for OCGAS GWAS

Shown are the result for the hybrid analysis of the within and between family component of the OCGAS GWAS. A thin blue line indicates level of suggestive evidence for association (1×10^{-5}) and a thin red line indicates genome-wide significance (5×10^{-8}).
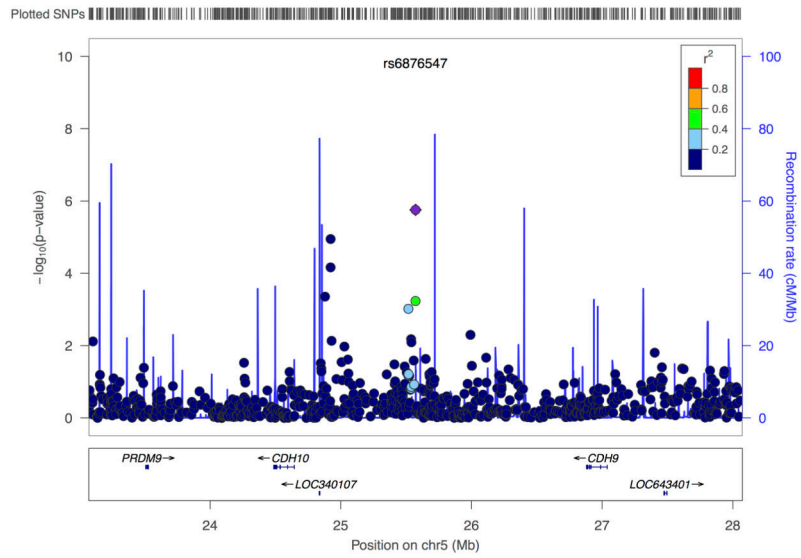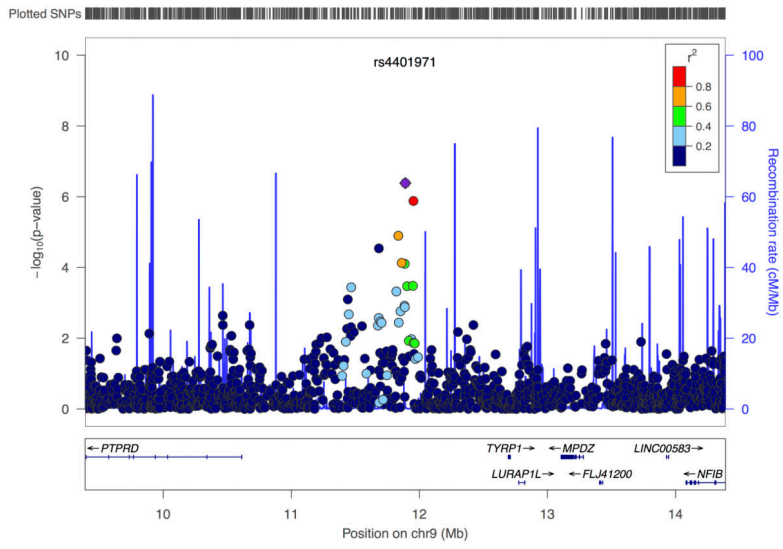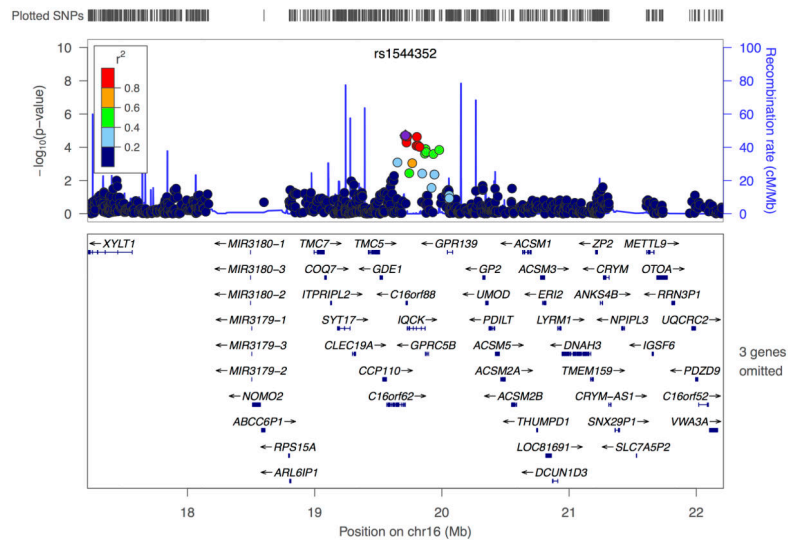
Figure 2. Regional association plots for top regions in OCGAS GWAS

The most associated marker in the region (see table 2 and S2, purple dot) is centered in a genomic window of 5 Mb (hg19). P-values for the OCGAS GWAS are given. The linkage disequilibrium (LD) strength (r^2; data from the 1000 genomes project European samples) between the sentinel single nucleotide polymorphism and its flanking markers is demonstrated by the coloring of the dots for the neighboring markers (ranging from red = high to blue = low). The recombination rate (cM/Mb; second y axis) is plotted in blue. Plots are given for the a) chromosome 9 region harboring *PTPRD* (clump 1 in table 2), b) the chromosome 5 region (clump 2 in table 2), and c) the chromosome 16 region harboring the two genes that were identified in the gene-based analysis, *IQCK* and *C16orf88*.

Table 1

Results for OCGAS GWAS

SNPs listed are strongest associated GWAS variants ($P < 0.0001$) in the hybrid analysis of the within and between family component. Only one SNP ("lead SNP" = SNP with lowest p-value) per clump is given. Clumps are defined using PLINK (--clump-p1 0.0001, --clump-p2 0.05, --clump-r2 0.20, --clump-kb 5000, LD based on additional controls that have also been used for the hybrid analysis). Headings are as follows: MARKERINFO (# - Rank place for clump for which this SNP is the lead SNP; CHR – Chromosome; SNP – marker name for lead SNP in clump; BP – position; MA – minor allele in additional controls; MAF – frequency of this SNP in additional controls;), OCGAS (INFO – number of informative families in the within family component analysis; DIR – direction of effect with respect to the MA; P – p-value for the lead SNP in the clump; N – number of additional SNPs in the clump meeting the aforementioned criteria), and ANNOTATION (GENE – gene in which lead SNP is physically located; LEFT GENE and RIGHT GENE – nearest flanking genes).

#	CHR	MARKERINFO					OCGAS				ANNOTATION	
		SNP	BP	MA	MAF	INFO	DIR	P	N	GENE	LEFT GENE	RIGHT GENE
1	9	rs4401971	11890045	A	0.41	619	– –	4.13×10^{-7}	31	NA	LOC646114	LOC100049717
2	5	rs6876547	25572301	G	0.19	471	++	1.76×10^{-6}	6	NA	CDH10	MSNL1
3	7	rs1343795	54313171	A	0.08	229	– –	9.69×10^{-6}	6	NA	FLJ45974	LOC222005
4	5	rs6452234	24922789	A	0.21	429	– –	1.13×10^{-5}	10	NA	CDH10	MSNL1
5	14	rs1014951	24039559	T	0.1	311	++	1.49×10^{-5}	6	JPH4	LOC100131731	DHRS2
6	8	rs7462051	80266553	A	0.18	462	++	1.69×10^{-5}	21	NA	IL7	LOC100128963
7	16	rs1544352	19713882	C	0.16	445	++	1.94×10^{-5}	19	NA	C16orf62	C16orf88
8	4	rs1532154	157279663	C	0.21	427	– –	2.22×10^{-5}	3	NA	FTHP2	hCG_1814936
9	11	rs509876	79811268	G	0.13	406	++	2.29×10^{-5}	4	NA	LOC646112	LOC729790
10	20	rs10392	37550935	A	0.19	467	– –	2.87×10^{-5}	18	PPP1R16	ACTR5	FAM83D
11	9	rs2821204	11683901	G	0.2	492	++	2.89×10^{-5}	29	NA	LOC646114	LOC100049717
12	10	rs1088258	97109019	A	0.33	577	– –	3.56×10^{-5}	22	SORBS1	PDLIM1	LOC643981
13	8	rs2278144	25634077	T	0.13	382	– –	3.94×10^{-5}	2	NA	CDCA2	EBF2
14	15	rs8026755	37247722	A	0.33	594	– –	4.03×10^{-5}	8	MEIS2	LOC145845	LOC390576
15	3	rs833209	176352763	C	0.34	616	– –	4.14×10^{-5}	5	NA	LOC730168	TBL1XR1
16	5	rs1773562	168653502	C	0.1	312	++	5.03×10^{-5}	4	SLIT3	LOC728095	CCDC99
17	8	rs7003102	126303493	G	0.14	416	++	5.14×10^{-5}	6	NSMCE2	KIAA0196	TRIB1
18	8	rs1254734	26833198	A	0.2	486	– –	5.38×10^{-5}	11	NA	LOC100127897	LOC100132229

#	CHR	MARKERINFO		MA	MAF	OCGAS				ANNOTATION		
		SNP	BP			INFO	DIR	P	N	GENE	LEFT GENE	RIGHT GENE
19	2	rs7593878	44358504	C	0.15	378	– –	5.44×10^{-5}	6	NA	LRPPRC	PPM1B
20	2	rs1686740	182189405	T	0.11	345	++	6.10×10^{-5}	5	NA	LOC729026	LOC100127923
21	13	rs9541148	35095884	C	0.37	624	– –	6.73×10^{-5}	5	NA	LOC100130499	LOC100129452
22	18	rs1671253	50040561	T	0.29	555	– –	7.04×10^{-5}	19	DCC	LOC100132995	LOC100133176
23	6	rs973714	133578733	A	0.34	615	++	7.44×10^{-5}	12	EYA4	LOC285735	TCF21
24	10	rs3902042	17065067	T	0.39	641	++	7.71×10^{-5}	12	CUBN	RSU1	TRDMT1
25	20	rs8120171	50712059	T	0.19	414	– –	7.85×10^{-5}	1	ZFP64	SALL4	ERP28P
26	4	rs1685101	75216365	A	0.24	467	– –	7.89×10^{-5}	15	NA	EPGN	EREG
28	3	rs9845643	179531062	G	0.2	457	– –	8.58×10^{-5}	7	PEX5L	USP13	LOC647249
29	9	rs1126590	92077132	C	0.39	634	– –	8.60×10^{-5}	12	SEMA4D	PA2G4P6	LOC100128670
30	20	rs1699787	16533162	C	0.23	518	++	8.68×10^{-5}	12	KIF16B	LOC100131642	RPLP0P1
31	9	rs2183738	71211200	G	0.14	366	++	8.73×10^{-5}	8	NA	C9orf71	LOC347097
32	1	rs1209698	185201061	C	0.34	612	++	8.87×10^{-5}	32	C1orf26	LOC100129295	IVNS1ABP
33	8	rs7005206	130620813	A	0.15	377	– –	9.13×10^{-5}	7	NA	LOC100129525	LOC100130376

Table 2

Follow up of IOCDF-GC GWAS results

SNPs listed are strongest associated GWAS variants in trio, case-control and combined trio-case-control samples reported in table 1 of Stewart et al. 2013. Only the results from the trio-case-control sample are listed in the table (IOCDF –GC: P) along the results from the current GWAS (OCGAS: P). Directions of effects (IOCDF-GC: DIR and OCGAS: DIR) are given based on the A1 alleles in the IOCF analyses (IOCDF-GC: A1). For the IOCDF-GC GWAS directions are given for all study samples (EU, AJ, SA, trios). In case of the OCGAS GWAS the directions refer to the combined results (within and between family component). For SNPs not present in the OCGAS results a proxy SNP was used that was identified using the SNAP tool (OCGAS: PROXY; max distance to query SNP 500kb, rsq min 0.8, reference data: 1000 genomes project pilot 1 release).

| MARKERINFO | | | IOCDF-GC | | | OCGAS | | | |
SNP	CHR	BP (HG19)	P	A1/A2	DIR	P	DIR	PROXY (rsq)	GENES (kb)
rs11898020	2	144282078	2.65E-04	A/G	+++–	0.6357	–	N/A	ARHGAP15 (intronic)
rs10165908	2	158315629	0.0169	C/T	+++–	0.5816	+	rs2198581 (0.83)	CYTIP (15), ACVR1C (68)
rs1838733	5	58533392	3.82E-05	T/C		0.0482	+	rs6858946 (0.95)	PDE4D (intronic)
rs26728	5	106946056	1.01E-04	T/C	–++–	0.3289	–	rs154191 (0.85)	EFNA5 (intronic)
rs4868342	5	173504522	3.20E-05	C/T	++++	0.8425	–	N/A	HMP19 (intronic)
rs9499708	6	104445367	2.96E-06	C/T	––+–	0.1287	–	N/A	GRIK2 (1927), HACE1 (731)
rs2205748	**6**	**104462555**	**8.52E-06**	**G/A**	**++++**	**0.0493**	**+**	**N/A**	**GRIK2 (1944), HACE1 (713)**
rs182320	6	130073291	2.25E-05	C/T	++++	0.5801	+	N/A	ARHGAP18 (42), C6orf191 (79)
rs6531002	8	12722703	0.0067	T/C	–++–	0.4897	–	rs7823534 (0.96)	LONRF1 (110), KIAA1456 (80)
rs11611761	12	33025612	0.115	A/C	––+–	0.2676	–	rs17543624 (0.87)	PKP2 (intronic)
rs297941	12	50319086	4.99E-07	G/A		0.8846	–	N/A	FAIM2 (21), AQP2 (25)
rs9652236	13	72688774	5.14E-06	T/G	++++	0.4017	+	N/A	DACH1 (247), MZT1 (594)
rs11081062	18	3662879	2.92E-05	T/C	+0++	0.6274	+	N/A	DLGAP1 (intronic)
rs11663827	18	3663631	2.31E-05	A/G	++++	0.6844	+	N/A	DLGAP1 (intronic)
rs485186	19	49207206	9.94E-06	G/A	++++	0.5067	+	N/A	FUT2 (coding-synon)
rs6131295	20	11996267	3.63E-05	G/A	+	0.5663	–	N/A	BTBD3 (89), SPTLC3 (993)

Table 3

Gene-based results

Genes listed are strongest associated protein-coding genes in the FORGE analysis ($P < 0.001$). Annotation information is taken from ENSEMBL v70 (GENEINFO: CHR, START, STOP), gene names (GENEINFO: GENE) are following HGNC nomenclature. The results (ASSOCIATION: P – gene-based p-value; P_CORR – experiment-wise corrected gene-based p-value; N – number of SNPs annotated to the gene; minP – p-value of SNP with lowest p-value that was annotated to the gene) are based on the Z_FIX statistic as implemented into FORGE. Annotation used a +/− 20kb (from GENEINFO: START and STOP) flanking region for each gene under study.

	GENEINFO				ASSOCIATION			
CHR	START	STOP	GENE	P	P_CORR	N	minP
16	19714902	19729557	C16orf88	9.99E-07	0.019	7	1.94E-05
16	19727778	19868907	IQCK	9.99E-07	0.019	20	1.94E-05
6	9596343	10211841	OFCC1	6.29E-05	1	152	2.86E-04
14	24037244	24048009	JPH4	1.16E-04	1	5	1.49E-05
15	79252289	79383115	RASGRF1	1.18E-04	1	53	6.25E-04
19	2100988	2164464	AP3D1	1.96E-04	1	17	1.93E-03
9	116148597	116163613	ALAD	2.56E-04	1	23	9.83E-03
1	185087220	185126204	TRMT1L	2.93E-04	1	9	1.73E-04
4	75230860	75254468	EREG	3.06E-04	1	19	7.89E-05
20	37590942	37668366	DHX35	3.55E-04	1	7	6.22E-05
6	88117701	88221352	C6orf165	3.55E-04	1	15	1.84E-03
15	41186628	41196173	VPS18	4.36E-04	1	6	1.02E-03
1	185126212	185260897	SWT1	5.00E-04	1	22	8.87E-05
6	88180341	88222054	SLC35A1	6.11E-04	1	10	1.84E-03
1	244571796	244615436	ADSS	6.30E-04	1	14	2.49E-04
9	97418353	97480105	C9orf118	6.32E-04	1	16	4.96E-04
9	71939488	72007371	FAM189A2	6.47E-04	1	18	9.37E-04
3	149191761	149221068	TM4SF4	6.67E-04	1	23	6.00E-04
10	97071528	97321171	SORBS1	6.84E-04	1	86	1.34E-04
16	19566562	19718115	C16orf62	7.54E-04	1	37	1.94E-05
19	19287712	19303425	MEF2BNB	8.09E-04	1	11	8.85E-04
10	90562654	90580303	LIPM	8.46E-04	1	21	2.73E-03

	GENEINFO			ASSOCIATION			
CHR	START	STOP	GENE	P	P_CORR	N	minP
14	24025216	24029480	THTPA	8.90E-04	1	8	1.49E-05
1	204042243	204096863	SOX13	9.17E-04	1	37	1.57E-03
1	231359509	231376933	C1orf131	9.33E-04	1	15	4.05E-03
6	133561736	133853258	EYA4	1.00E-03	1	84	7.44E-05