A Haplotype Containing Quantitative Trait Loci for SLC1A1 Gene Expression and Its Association With Obsessive-Compulsive Disorder

Jens R. Wendland, MD; Pablo R. Moya, PhD; Kiara R. Timpano, MS; Adriana P. Anavitarte, BS; Matthew R. Kruse, BA; Michael G. Wheaton, BA; Renee F. Ren-Patterson, MD, PhD; Dennis L. Murphy, MD

Context: Recent evidence from linkage analyses and follow-up candidate gene studies supports the involvement of SLC1A1, which encodes the neuronal glutamate transporter, in the development of obsessive-compulsive disorder (OCD).

Objectives: To determine the role of genetic variation of SLC1A1 in OCD in a large case-control study and to better understand how SLC1A1 variation affects functionality.

Design: A case-control study.

Setting: Publicly accessible SLC1A1 expression and genotype data.

Patients: Three hundred twenty-five OCD probands and 662 ethnically and sex-matched controls.

Interventions: Probands were assessed with the Structured Clinical Interview for DSM-IV, the Yale-Brown Obsessive Compulsive Scale, and the Saving Inventory–Revised. Six single-nucleotide polymorphisms (SNPs) were genotyped. Multiple testing corrections for single-marker and haplotype analyses were performed by permutation.

Results: Gene expression of SLC1A1 is heritable in lymphoblastoid cell lines. We identified 3 SNPs in or near SLC1A1 that correlated with gene expression levels, 1 of which had previously been associated with OCD. Two of these SNPs also predicted expression levels in human brain tissue, and 1 SNP was further functional in reporter gene studies. Two haplotypes at 3 SNPs, rs3087879, rs301430, and rs7858819, were significantly associated with OCD after multiple-testing correction and contained 2 SNPs associated with expression levels. In addition, another SNP correlating with SLC1A1 gene expression, rs3933331, was associated with an OCD-hoarding subphenotype as assessed by 2 independent, validated scales.

Conclusions: Our case-control data corroborate previous smaller family-based studies that indicated that SLC1A1 is a susceptibility locus for OCD. The expression and genotype database-mining approach we used provides a potentially useful complementary approach to strengthen future candidate gene studies in neuropsychiatric and other disorders.
phisms (SNPs) nor haplotypes closely corresponded.\textsuperscript{16-18} These 3 family-based studies motivated us to analyze several \textit{SLC1A1} polymorphisms in our large and phenotypically well-characterized collection of OCD probands using a case-control study design. Case-control studies, while susceptible to population stratification, provide greater statistical power than family-based designs. We focused on \textit{SLC1A1} SNPs that were significant in at least 1 previous study and we included additional SNPs that predicted gene expression levels as determined by analyzing deposited expression and genotype data from nonpatient sources. We present evidence for a susceptibility haplotype spanning large portions of \textit{SLC1A1} and also for a SNP located 100 kilobases (kb) upstream of \textit{SLC1A1} that was associated with an OCD subphenotype, compulsive hoarding, which has some genetic linkage and clinical differences from the more common OCD phenotype.\textsuperscript{19,21}

### METHODS

#### \textbf{SLC1A1 GENE EXPRESSION ANALYSES}

For heritability calculations, we extracted lymphoblastoid cell line \textit{SLC1A1} gene expression values (probe No. 213664\_at) from the Gene Expression Omnibus data set GSE1485.\textsuperscript{22} All individuals had a “present” call for this probe; technical replicates were averaged and then log\textsubscript{2}-transformed. Family structures for the 14 three-generational families represented in this data set were obtained from the Centre d’Etude du Polymorphisme Humain database (http://www.cephb.fr/en/cephdb/browser.php). Pedigree and expression data were then analyzed with the Sequential Oligogenic Linkage Analysis Routines software package by performing a standard quantitative genetic analysis.\textsuperscript{23} The residual kurtosis of 0.125 was within the normal range.

For expression quantitative trait locus (eQTL) analyses, we retrieved lymphoblastoid cell line \textit{SLC1A1} expression values for the 60 HapMap Centre d’Etude du Polymorphisme Humain founders with northern and western European ancestry from 2 independent platforms: the Gene Expression Omnibus data set GSE5859 (Affymetrix platform, probe No. 213664\_at; replicates were averaged and transformed as above)\textsuperscript{24} and from the Gene Expression Variation Web site hosted by the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/humgen\_genevar; Illumina platform, probe No. GI\_31343623-S).\textsuperscript{25} Genotypes for all 90 individuals from the Centre d’Etude du Polymorphisme Humain HapMap data were downloaded from the HapMap Web site (http://www.hapmap.org, release 22) for the entire \textit{SLC1A1} region plus 1 megabase (Mb) flanking both sides (corresponding to position 3 480 444 to 5 577 469 on chromosome 9 in the University of California–Santa Cruz Genome Browser Human, March 2006 assembly). Markers inconsistent with mendelian inheritance patterns, with minor allele frequencies of less than 1%, or with more than 10% missing genotypes were removed, resulting in a total of 2467 polymorphisms that were analyzed. These SNPs were then tested for quantitative association with \textit{SLC1A1} expression using PLINK,\textsuperscript{26} Microsoft Excel (Microsoft, Redmond, Washington), and Prism 4 (GraphPad Software, San Diego, California).

#### \textbf{REAL-TIME REVERSE TRANSCRIPTASE–POLYMERASE CHAIN REACTION QUANTITATION OF SLC1A1 IN POSTMORTEM BRAIN TISSUE}

We obtained RNA and genomic DNA from 105 postmortem dorsolateral prefrontal cortex samples through the Stanley Medical Research Institute, Chevy Chase, Maryland. This collection is composed of 3 diagnostic groups (35 normal, 35 bipolar, and 35 schizophrenic donors). With the exception of 2, all samples were from white donors. Detailed demographics of this sample are available at http://www.stanleyresearch.org. Six of these 105 samples did not provide sufficient RNA for quantitation, and the 2 nonwhite samples were quantitated but not analyzed, resulting in a total of 97 samples, which are presented herein. All samples were successfully genotyped for rs3933331, rs7858819, and rs301430. We reverse-transcribed 1 μg of total RNA with oligo\textsubscript{dT} primers using the Transcriptor First Strand cDNA Synthesis kit according to the manufacturer’s instructions (Roche Applied Science, Indianapolis, Indiana). Next, we amplified a total of 5 ng of reverse-transcribed RNA in triplicate reactions with 1 assay for \textit{SLC1A1} and 1 endogenous control assay for \textit{PGK1}. The \textit{SLC1A1} assay was designed with the Universal Probe Library Assay Design Center, version 2.4 (Roche Applied Science) using primers GCT GTC GAC TGG CTG CG and GGA GAG CTT TTC CAC AAT GC (Operon Biotechnologies, Huntsville, Alabama) and Universal Probe Library probe No. 9 (Roche Applied Science). The \textit{PGK1} assay was predesigned and purchased directly from the vendor (Universal Probe Library Human PGK1 Gene Assay, Roche Applied Science). Both assays span introns and do not contain known SNPs in primers or probes. Amplification was carried out in a 10-μL total reaction volume with 300 nM of each primer and 100 nM of fluorescent probe using PerfeCta qPCR FastMix (Quanta BioSciences, Gaithersburg, Maryland). Thermocycling consisted of 2 minutes at 45°C for uracil-N-glycosylase incubation, 30 seconds at 95°C for hot-start activation, and 40 amplification cycles of 2 seconds at 95°C and 20 seconds at 60°C followed by fluorescence acquisition in a LightCycler 480 instrument (Roche Applied Science). Crossing-point determination by second derivative maximum was carried out with LightCycler 480 software, release 1.3. Relative abundances of \textit{SLC1A1} messenger RNA were then calculated using the efficiency-corrected comparative threshold method and log\textsubscript{2}-transformed.\textsuperscript{27} Efficiency calculations were performed with 2-fold serially diluted pooled complementary DNA (range, 30-0.47 ng per reaction). Amplicon size and assay specificity were verified by agarose electrophoresis. No-template controls did not display detectable amplification (data not shown). All experiments were done after the specimen code was broken and they were thus unblinded. The entire data set has been uploaded to the Stanley Medical Research Institute data bank (http://www.stanleyresearch.org/brain).

### REPORTER GENE EXPERIMENTS

We performed luciferase assay experiments for 3 SNPs within or near \textit{SLC1A1} as recently described.\textsuperscript{28} Briefly, genomic DNA was amplified with primers flanking rs3933331 (TTC TGC CCA AGA TCA TCA and CAG TGG TTG CAT CCA), rs7858819 (TGC AAA TGG TTA TGG GAC AA and TCA AAT AGA CAA TCA CA and TCT GGG TTT AGA CTG CCA CA), and rs301430 (CAT GTC TTC AGG CAG GGA CT and CAC TGC GAC ATT CTG GTG) by agarose electrophoresis. No-template controls did not contain known SNPs in primers or probes. Amplification was carried out in a 20-μL total reaction volume with 100 nM of each primer and 100 nM of fluorescent probe using PerfeCta qPCR FastMix (Quanta BioSciences, Gaithersburg, Maryland). Thermocycling consisted of 9 minutes at 95°C for uracil-N-glycosylase incubation, 30 seconds at 95°C for hot-start activation, and 40 amplification cycles of 3 seconds at 95°C and 20 seconds at 60°C followed by fluorescence acquisition in a LightCycler 480 instrument (Roche Applied Science). Crosspoint determination by second derivative maximum was carried out with LightCycler 480 software, release 1.3. Relative abundances of \textit{SLC1A1} messenger RNA were then calculated using the efficiency-corrected comparative threshold method and log\textsubscript{2}-transformed.\textsuperscript{27} Efficiency calculations were performed with 2-fold serially diluted pooled complementary DNA (range, 30-0.47 ng per reaction). Amplicon size and assay specificity were verified by agarose electrophoresis. No-template controls did not display detectable amplification (data not shown). All experiments were done after the specimen code was broken and they were thus unblinded. The entire data set has been uploaded to the Stanley Medical Research Institute data bank (http://www.stanleyresearch.org/brain).
standard conditions. For each SNP, data were normalized to the geometric mean of 1 allele and tested against the other allele with a 2-sided, unpaired t-test using Prism 4 (GraphPad Software).

SUBJECTS

Adult OCD probands (325 white subjects) were recruited through an ongoing outpatient OCD program at the National Institute of Mental Health, Bethesda, Maryland. All probands were aged at least 18 years at the time of inclusion (mean age at time of interview, 38 years; standard deviation [SD], 16 years) and gave written informed consent. The diagnosis was made on the basis of the Structured Clinical Interview for DSM-IV. Exclusion criteria were active psychosis or schizophrenia, severe mental retardation that would prohibit an evaluation of OCD, and OCD symptoms that occurred exclusively in the context of major depression.

The sample consisted of 186 women (37%) and 139 men (43%), with a mean age at OCD onset of 14 years (SD, 9.6 years), which was similar to that in the SLC1A1 OCD study by Arnold et al (14.4 years) and higher than those in the studies by Dickel et al (8 years) and Stewart et al (7.9 years). As reported previously in our sample, mood (81%) and anxiety (54%) disorders were most frequently represented among comorbid disorders, with alcohol abuse disorder (24%), substance abuse disorders (14%), and tics (6%) less common. Hoarding, as we will define, was present in 27% of this patient sample. Total mean Yale-Brown Obsessive Compulsive Severity Scale (Y-BOCS) lifetime severity scale score was 29.6 (SD, 15.4), which is approximately equally contributed to by obsession and compulsion scores.

Our control samples (662 white subjects [46% male and 54% female]) originated from 3 independent sources: (1) undergraduate students (270 self-declared healthy white students) from a large southeastern university who participated in a separate study of genes and personality in return for partial course credit; (2) Coriell Cell Repository, Camden, New Jersey (200 self-declared healthy white Americans); and (3) European Collection of Cell Cultures (Sigma-Aldrich, St Louis, Missouri; 192 apparently healthy white blood donors in the United Kingdom). The male to female ratio did not significantly differ between cases and controls (p < 0.05). We cannot rule out the occurrence of OCD within the second and third control groups, as self-report and standard scales were used for evaluation of only the first group. Genotypic and allelic frequencies did not significantly differ between these 3 control groups (data not shown). All studies were conducted under protocols approved by the institutional review board at the National Institute of Mental Health Division of Intramural Research Programs and by the Human Subjects Committee at Florida State University. Additional details on proband and control samples have been described previously.

GENOTYPING AND STATISTICAL GENETICS

Genomic DNA was extracted from whole blood obtained through venipuncture or from saliva samples (Oragene discs; DNA Genotek, Ottawa, Ontario, Canada). All SNPs were genotyped with 5’ exonuclease assays (Applied Biosystems, TaqMan predesigned or custom assays; Applied Biosystems, Foster City, California) ( assay identification numbers and primer/probe sequences are available from the corresponding author) under standard conditions described previously. The overall genotyping completion rate exceeded 98% for each assay; the per-individual genotyping completion rate was at least 95% of 6 SNPs. None of the SNPs deviated from Hardy-Weinberg equilibrium as determined by contingency-table statistics (nominal p > 0.05; data not shown). Duplicate samples (at least 10% of all samples, randomly chosen, for each of the 6 SNPs) and no-template controls consistently yielded expected results.

Genotype data from probands and controls were analyzed for single-locus and haplotype associations with the software packages PLINK and WHAP. Conditional haplotype analyses were carried out as recently described after detection of a significant omnibus association. Calculation of intermarker linkage disequilibrium and correction for multiple testing with 100,000 permutations for the 6 individual markers as well as for the 3-locus haplotype was conducted using haploview, version 4.0.

Using the Genetic Power Calculator, we determined that our sample had 78% power at p = 0.05 with unscreened controls (risk allele frequency = 0.3; disease prevalence = 0.025; genotypic relative risk AA = 1.4 and AA = 1.8; marker allele frequency = 0.3 at D’ = 0.9 with risk variant). When we selected screened controls and used otherwise identical parameters, power rose only slightly, to 80%. We conclude from these calculations that our partially unscreened control sample did not significantly affect our study results.

SUBPHENOTYPE ANALYSES

We examined 2 subphenotypes in this article: (1) we measured lifetime OCD symptom severity using Y-BOCS, and (2) we further considered hoarding as a second subphenotype based on 2 separate, validated measures of the hoarding construct. We focused on hoarding as a subphenotype, since its familiality is stronger than other OCD subphenotypes, which is further reflected by 2 recent significant linkage analyses. The Y-BOCS symptom checklist was used to derive factor scores for the hoarding symptom dimension using principal component analysis and the Varimax procedure, as previously reported. The factor analysis generated a factor score for each subject representing the correlation of their symptom profile with the hoarding factor, factor IV, as reported previously. Because the hoarding factor was essentially independent (factor loading = 0.85) from the other 3 factors identified in our sample (factor loading on the hoarding factor by other factors, 0.01-0.16) and likewise by others, the hoarding factor scores were standardized, with a mean of 0 and an SD of 1. Subjects were dichotomized on the basis of scores greater than 0 or less than 0 on the hoarding factor. Details of this procedure have been presented elsewhere. We also assessed the severity of hoarding symptoms with the Saving Inventory-Revised (SIR). This measure is a validated tool for examining the magnitude of hoarding symptoms and includes facets of this phenomenon not considered in the Y-BOCS symptom checklist. Both of these measures of hoarding have been previously characterized and evaluated in our clinical OCD sample.

Statistical differences in these subphenotypes according to genotype were ascertained using linear regression and SPSS, version 14.0 (SPSS Inc, Chicago, Illinois), for Microsoft Windows XP (Microsoft). All tests were performed with 2-sided p < .05. Given missing data, the populations for each of the analyses varied slightly (SIR, n = 198; hoarding factor, n = 216).

RESULTS

SLC1A1 GENE EXPRESSION ANALYSES IN LYMPHOBlastOID CELL LINES

Using publicly available gene expression data, we observed that SLC1A1 is robustly expressed in lymphoblastoid cell lines. This prompted us to investigate whether SLC1A1 expression constitutes a heritable trait. By analyzing 194 lymphoblastoid cell lines from 14 three-generation families, we determined an additive (narrow-sense or h2) heritability of 0.23 (p < .001) for SLC1A1 gene expression. 

©2009 American Medical Association. All rights reserved.
expression, with no significant sex effects when sex was included as a covariate. We then sought to identify eQTLs for SLC1A1 acting in cis within 1 Mb of the SLC1A1 genomic locus by performing a linear regression analysis of expression data on genotypes recoded as 0, 1, or 2 on the basis of minor allele occurrences. Of 2467 markers tested, 3 SNPs that were not in complete linkage disequilibrium scored a nominal $P = .02$ and had the same direction of effect in both data sets (ie, positive or negative regression coefficient; $r^2$ value, 9%-14%) (Figure 1). One of these SNPs, rs3933331, is located about 100 kb upstream of the SLC1A1 transcription initiation site, while the 2 other eQTL SNPs, rs7858819 and rs301430, are in the 3' end of SLC1A1. Remarkably, rs301430 has recently been associated with early-onset OCD in a collection of 71 trios both individually and as part of a haplotype, and the haplotypic association has been replicated subsequently.

**SLC1A1 GENE EXPRESSION ANALYSES IN HUMAN POSTMORTEM BRAIN TISSUE**

As a next step, we analyzed whether the 3 eQTLs we identified were also associated with SLC1A1 gene expression in neuronal tissue. To this end, we determined genotypes and relative SLC1A1 messenger RNA levels by quantitative polymerase chain reaction in 97 postmortem brain samples from normal, bipolar, and schizophrenic donors, all derived from the dorsolateral prefrontal cortex. There were no significant differences in SLC1A1 expression levels among the 3 diagnosis groups as calculated by analysis of variance (data not shown). We then analyzed all 3 groups jointly for association with eQTLs to have increased power. As shown in Figure 2, rs7858819 and rs301430 were nominally significantly associated with SLC1A1 messenger RNA levels in human brain tissue ($P = .02$ each). Moreover, the direction of effect (increased expression with increasing number of T alleles at rs7858819 and decreased expression at rs301430) was identical to the data from lymphoblastoid cell line data (Figure 1 and Figure 2).

**IN VITRO ANALYSES OF SLC1A1 EXPRESSION–ASSOCIATED SNPS WITH REPORTER GENE CONSTRUCTS**

Following the replication of 2 of 3 eQTLs in neuronal tissue, we sought to directly assess functionality of these SNPs in vitro through the use of luciferase reporter gene assays. We generated constructs with both alleles at each of the 3 SNPs, rs3933331, rs7858819, and rs301430, and used human neuroblastoma (SH-SY5Y) and rat pheochromocytoma...
cytoma (PC12) cell lines for transfection. As depicted in Figure 3, the 2 alleles at rs301430 had a highly significant differential effect on reporter gene expression in both cell lines ($P < .001$ for both). In line with the T allele–associated decrease in SLC1A1 expression in lymphoblastoid cell lines and neuronal tissue, the rs301430 construct with the T allele showed about 50% decreased expression compared with the C allele (Figure 3).

**ASSOCIATION ANALYSES OF SLC1A1 AND OCD**

We then genotyped the 3 eQTL SNPs we described and 3 additional SNPs within SLC1A1 that have been individually associated with OCD in prior family-based studies. In our collection of 325 probands and 662 controls matched for ethnicity and sex, Table 1 details genomic location and linkage disequilibrium of the 6 analyzed markers. As presented in Table 1, none of these 6 SNPs were individually associated with OCD after permutation-based multiple-testing correction. We also performed these analyses separately for males and females and did not observe significant results (data not shown).

Haplotype analyses, in contrast, revealed a highly significant omnibus association of 3 of these markers, rs3087879, rs301430, and rs7858819, with OCD (Table 2). In haplotype-specific tests, when tested individually against all other haplotypes, with $df = 1$, 3 of the 4 common haplotypes were nominally significant at $P < .05$. Two of these haplotypes remained significant after multiple-testing correction by permutation (Table 2).

One haplotype, H4 in Table 2, was highly significant and almost twice as common in OCD than in controls (haplotype frequency, 9.4% vs 5.1%, respectively; odds ratio, 1.89; $P = .001$, family-wise permuted). We obtained similar—though not always as highly significant—results when we conducted these haplotype analyses stratified by sex (Table 2).

**SUBPHENOTYPE ANALYSES OF OCD**

Lastly, we set out to determine whether or not SLC1A1 is also associated with differential OCD symptom presentation, given the growing appreciation of OCD as a clinically, and possibly etiologically, heterogeneous disorder. We performed subphenotype analyses on the basis of Y-BOCS–ascertained OCD symptom severity and the hoarding phenotype evaluated by 2 measures: (1) a Y-BOCS factor analysis–derived hoarding symptom dimension, and (2) the magnitude of hoarding symptoms, as indexed by the SIR. We observed a nominally significant association of 1 eQTL SNP, rs3933331, with the hoarding phenotype. Specifically, this SNP was significantly associated with hoarding as assessed by both the Y-BOCS hoarding symptom dimension ($\beta = .21, t_{350} = 2.9, P = .005$) and total SIR scores ($\beta = .19, t_{350} = 2.6, P = .009$). That is, greater scores on the hoarding dimension, as well as more severe SIR symptom scores, were strongly correlated with the number of G alleles at rs3933331 (Figure 2). The association with SIR scores appears to be largely driven by the acquisitioning subscale ($\beta = .22, P = .002$), as rs3933331...
We identified an SLC1A1 haplotype that was strongly associated with OCD that consists of 3 markers between intron 2 and the 3′ untranslated region of SLC1A1, 2 of which we found to be SNPs that predict gene expression levels (ie, eQTLs) in lymphoblastoid cell lines. In addition, another eQTL polymorphism, located somewhat upstream from SLC1A1, was strongly correlated with an OCD subphenotype (hoarding), as evaluated by 2 validated scales. Our data corroborate the hypothesis of dys-functional glutamatergic neurotransmission in OCD and hoarding, mediated in part through SLC1A1.

Three recently published, family-based association studies have suggested SLC1A1 to be a candidate gene for OCD.16-18 However, no given SNP within SLC1A1 in these 3 studies was consistently positively associated with OCD across the 3 studies when tested individually, with the exception of rs3780412, which was considered significant in 2 studies: the overall and males-only study by Dickel and coworkers17 and in the males-only sample by Stewart and colleagues.18 However, our results do not support a specific involvement of this marker in OCD, nor did we observe any preferential or stronger association of single variants or haplotypes in males compared with females in our case-control sample (Table 1 and Table 2). Speculatively, this observation might be related to the increased power of our case-control design compared with the family-based studies published previously.16-18 Clearly, additional studies are warranted to clarify the discrepancy for pref-
ential association in males observed in prior family-based studies, though the overall finding of involvement of SLC1A1 in OCD now appears more clearly replicable.

One of the most important common findings in these 3 previous SLC1A1 studies was a haplotypic association in the 3' region of the gene. Our results are in line with this observation, as 2 of the 3 SNPs of the associated haplotype we identified are located in exon 10 and the 3' untranslated region. The third SNP of this haplotype, however, lies somewhat upstream (intron 2) of this region, and although all 3 SNPs (rs7858819, rs301430, and rs3087879) were in considerable linkage disequilibrium with each other (Table 1), we speculatively interpret this observation as evidence for more than 1 functional locus within SLC1A1, as has been suggested previously. Additional support for this tentative conclusion stems from another SNP, rs3933331, located farther upstream and clearly not in linkage disequilibrium with the previously mentioned haplotype markers that we found to be associated with gene expression functionality and that predicted an OCD hoarding subphenotype in our sample (Figure 1 and Figure 2). The possibility of 2 or more functional (possibly causal) variants of differential effect size provides 1 plausible explanation for the molecular discrepancies between present and antis of differential effect size provides 1 plausible explanation for the molecular discrepancies between present and

The second complementary study, ie, luciferase reporter gene constructs, was aimed at direct functional assessment of individual SNPs. While association analyses with tissue samples or cell lines have the advantage of more closely resembling in vivo conditions, the eQTLs do not necessarily have to directly confer functionality, as they may merely be in linkage disequilibrium with the actual functional polymorphism. We addressed this uncertainty by examining individual polymorphisms in vitro with reporter gene assays. Our luciferase data for rs301430 suggest that this variant is indeed directly functional. Additional studies are now needed to elucidate the molecular mechanisms through which rs301430 affects gene expression.

The H4 haplotype, as detailed in Table 2, which was almost twice as common in those with OCD as in controls (odds ratio, 1.89), represents the strongest haplotypic association of SLC1A1 with OCD published to date. It illustrates the usefulness of our complementary expression–mining approach, because 2 of its markers were eQTLs, and conditional haplotypic association testing with the remaining markers failed to yield positive results (see the "Methods" section). The individual alleles in H4, however, are not fully compatible with the notion of increased or decreased SLC1A1 gene expression in OCD, as the C allele at rs7858819 was the lower-expressing allele in the regression analyses (Figure 1B), while the C allele at rs301430 in H4 correlated with increased expression (Table 2 and Figure 1). On the other hand, the mildly protective haplotype H2 in Table 2 (odds ratio, 0.78), which narrowly survived correction for multiple testing in individual haplotype analysis, contained the lower-expressing allele at both eQTLs. Here, too, it will be essential to conduct gene expression analyses in larger samples and with more densely spaced markers.

The major limitation of our study is the potential susceptibility of the case-control design to population stratification, especially because our controls originated from 3 different groups that did not closely correspond in their geographic origin to our proband sample. While we addressed this issue by matching self-reported ethnicities of probands and controls, spurious effects in case-control studies (which are regarded by some authors as relatively uncommon) can be ruled out by genotyping multiple ancestry-informative markers, which was not feasible for this work. Dense genetic information on SLC1A1 markers from different ethnicities and geographic locations will increasingly become available from genome-wide association studies. Such data will allow a better judgment of how strongly (and if at all) population stratification is relevant to the SLC1A1 markers analyzed herein.
Another aspect of genetic studies such as ours with several investigated phenotypes that requires discussion is the multiple comparison problem. In these studies, the multiple statistical tests are clearly not independent of one another given linkage disequilibrium between variants (Table 1) and the high correlation between subphenotypes (hoarding in this case) and the general diagnosis of OCD. Therefore, methods that correct by the total number of tests (eg, the Bonferroni procedure) would be more conservative than appropriate in our view. We conducted empirical family-wise error correction for the categorical single-marker and haplotype analyses by permutation and determined that 2 haplotypes remain significant (H4 and H2) (Table 2). The significance values for our subphenotype analyses are nominally robust (for association of rs3933331 with Y-BOCS score, \( P = .005 \), and with SIR-assessed hoarding, \( P = .009 \)) (Figure 4). They would, however, not withstand Bonferroni correction and therefore need to be interpreted with caution and observed in larger samples in future studies.

In summary, we have identified a strong association of OCD with a 3-marker haplotype composed of eQTLs for SLC1A1, which in turn were determined by exploiting expression and genotype data repositories and further validated with brain expression studies and reporter assays. Our complementary expression analysis approach might prove to be of use in future genetic studies for genes expressed at quantifiable levels in lymphoblastoid cell lines and other population-based tissue collections. Our haplotypic association results call for continued appreciation of glutamatergic neurotransmission abnormalities in OCD research.

Submitted for Publication: February 29, 2008; final revision received July 5, 2008; accepted August 23, 2008.

Correspondence: Jens R. Wendland, MD, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 35 Convent Dr, Bldg 35, Porter Bldg, Room 1B209, Bethesda, MD 20892-3717 (wendlandj@mail.nih.gov).

Author Contributions: Drs Wendland and Murphy had full access to all of the data in the study and take responsibility for the integrity of the data and accuracy of the data analysis.

Financial Disclosure: None reported.

Funding/Support: This research was supported by the Intramural Research Program of the National Institute of Mental Health, National Institutes of Health, and by a NARSAD 2007 Young Investigator Award and a National Institute of Mental Health Division of Intramural Research Programs Julius Axelrod Memorial Fellowship Training Award to Dr Wendland.

Additional Contributions: We are indebted to all the participants in this study. Diane Kazuba, BS, and Brenda Justement, RN, BSN, MA, conducted patient interviews; Teresa Tolliver, MS, provided technical assistance in DNA extraction; and James Nagle, MS, and Debbie Kauffman, BS, of the National Institute of Neurological Disorders and Stroke DNA sequencing facility provided technical assistance. Postmortem brain tissue was supplied by the Stanley Medical Research Institute.

Figure 4. Two linear regression analyses of rs3933331 for the hoarding subphenotype, both factor analysis–derived (A) and on the basis of the Saving Inventory–Revised (SIR) score (B), which were nominally significant. A, \( P = .005 \); B, \( P = .009 \).

REFERENCES


