Polymorphisms in the *Trace Amine Receptor 4 (TRAR4)* Gene on Chromosome 6q23.2 Are Associated with Susceptibility to Schizophrenia

Jubao Duan,¹ Maria Martinez,² Alan R. Sanders,¹ Cuiping Hou,¹ Naruya Saitou,³ Takashi Kitano,³ Bryan J. Mowry,⁴,⁵ Raymond R. Crowe,⁶ Jeremy M. Silverman,⁷ Douglas F. Levinson,⁸ and Pablo V. Gejman¹

¹Center for Psychiatric Genetics, Department of Psychiatry and Behavioral Sciences, Evanston Northwestern Healthcare and Feinberg School of Medicine, Northwestern University, Evanston, IL; ²Méthodologie Statistique et Épidémiologie Génétique des Maladies Multifactorielles, INSERM, Evry, France; ³Division of Population Genetics, National Institute of Genetics, Mishima, Japan; ⁴Queensland Centre for Schizophrenia Research, The Park, Centre for Mental Health, Wacol, Australia; ⁵Department of Psychiatry, University of Queensland, Brisbane, Australia; ⁶Mental Health Clinical Research Center and Department of Psychiatry, University of Iowa College of Medicine, Iowa City; ⁷Department of Psychiatry, Mount Sinai School of Medicine, New York; and ⁸Department of Psychiatry, University of Pennsylvania, Philadelphia

Several linkage studies across multiple population groups provide convergent support for a susceptibility locus for schizophrenia—and, more recently, for bipolar disorder—on chromosome 6q13-q26. We genotyped 192 European-ancestry and African American (AA) pedigrees with schizophrenia from samples that previously showed linkage evidence to 6q13-q26, focusing on the ancestry and African American (AA) pedigrees with schizophrenia from samples that previously showed linkage evidence to 6q13-q26, focusing on the *MOXD1-STX7-TRARs* gene cluster at 6q23.2, which contains a number of prime candidate genes for schizophrenia. Thirty-one screening single-nucleotide polymorphisms (SNPs) were selected, providing a minimum coverage of at least 1 SNP/20 kb. The association observed with rs4305745 (P = .0014) within the *TRAR4* (trace amine receptor 4) gene remained significant after correction for multiple testing. Evidence for association was proportionally stronger in the smaller AA sample. We performed database searches and sequenced genomic DNA in a 30-proband subsample to obtain a high-density map of 23 SNPs spanning 21.6 kb of this gene. Single-SNP analyses and also haplotype analyses revealed that rs4305745 and/or two other polymorphisms in perfect linkage disequilibrium (LD) with rs4305745 appear to be the most likely variants underlying the association of the *TRAR4* region with schizophrenia. Comparative genomic analyses further revealed that rs4305745 and/or the associated polymorphisms in complete LD with rs4305745 could potentially affect gene expression. Moreover, RT-PCR studies of various human tissues, including brain, confirm that *TRAR4* is preferentially expressed in those brain regions that have been implicated in the pathophysiology of schizophrenia. These data provide strong preliminary evidence that *TRAR4* is a candidate gene for schizophrenia; replication is currently being attempted in additional clinical samples.

Introduction

Schizophrenia is a frequently chronic and devastating brain disorder that affects ~1% of the population worldwide (Jablensky et al. 1992). Typically, it presents in adolescence or young adulthood and is characterized by major disruptions of thinking (delusions and/or disorganization), perception (hallucinations), mood, and behavior (Gottesman and Shields 1982). Schizophrenia is strongly familial, with a heritability of ~80%, but its etiology is hypothesized to involve both genetic and environmental factors (Sanders and Gejman 2001). Recently, encouraging evidence for several genes potentially involved in the etiology of schizophrenia has been reported—namely, *dysbindin* (*DTNBP1*) (Straub et al. 2002; Schwab et al. 2003), *neuregulin 1* (*NRG1*) (Stefansson et al. 2002, 2003; Williams et al. 2003), *proline dehydrogenase (oxidase) 1* (*PRODH*) (Jacquet et al. 2002; Liu et al. 2002), *catechol-O-methyltransferase* (*COMT*) (Li et al. 2000; Egan et al. 2001; Shifman et al. 2002), *regulator of G-protein signaling 4* (*RGS4*) (Chowdari et al. 2002; Morris et al. 2004; Williams et al. 2004), *D-amino-acid oxidase activator* (*DAOA* [previously called *G72*) (Chumakov et al. 2002; Schumacher et al. 2004), and *D-amino-acid oxidase* (*DAO*) (Chumakov et al. 2002; Schumacher et al. 2004). Additionally, Brzustowicz et al. (2004) reported linkage disequilibrium between *C-terminal PDZ domain ligand of neuronal nitric oxide synthase* (*CAPON*) and schizo-
phenia, perhaps related to a previous association signal at 6q13-q26 ([MIM 603175]) (Cao et al. 1997; Martinez et al. 1999; Kauffman et al. 1998; Kaufmann et al. 1999; Bailer et al. 2000; Levinson et al. 2000; Lindholm et al. 2001; Lerer et al. 2003; Lewis et al. 2003). In the first report of linkage to 6q (Cao et al. 1997), support for linkage was observed in the region from D6S301 (located at 111 cm) to D6S305 (located at 170 cm). Linkage to this region was confirmed in subsequent studies (Kaufmann et al. 1998; Martinez et al. 1999; Bailer et al. 2000; Levinson et al. 2000). All the families used in this study are from three data sets, which we call “NIMH-IRP” (National Institute of Mental Health Intramural Research Program), “NIMH-GI” (NIMH Genetics Initiative), and “AU/US” (Australia/United States), that previously have been shown to yield evidence for linkage to chromosome 6q13-q26 ([MIM 603175]) (Cao et al. 1997; Martinez et al. 1999; Levinson et al. 2000). An analysis of 12 microsatellite markers in a 50-cM region was performed by Martinez et al. (1999); nonparametric affected sibling pair methods yielded P values of .00018, .00095, and .013 for the NIMH-IRP, NIMH-GI, and AU/US data sets, respectively. In a Palestinian/Israeli pedigree sample (Lerer et al. 2003), the linkage peak with a nonparametric linkage (NPL) score of 4.61 was at D6S292 (~137 cm), and the region from ~131 cm to ~144 cm contained the 1-NPL decrease portion of the linkage peak. Furthermore, another group found evidence for linkage in a Swedish pedigree to the region from ~170 cm to ~180 cm (Lindholm et al. 2001). Whether these different results are better explained by the presence of more than one schizophrenia susceptibility gene in 6q or reflect typical peak variability in complex disorders (Hauser and Boehnke 1997; Hsueh et al. 2001) is currently unknown. Recently, bipolar disorder was reported to map to 6q, with one study yielding a maximum LOD score of 2.2 at 113 cm near D6S1021 (Dick et al. 2003) and another study reporting a maximum LOD score of 3.56 at ~124–126 cm near D6S1632 (Middleton et al. 2004), raising the possibility that a common gene for schizophrenia and bipolar disorder may be located in 6q.

We were interested in a MOXD1-STX7-TRAR genes cluster at 6q23.2 (132.8 cm) that harbors prime candidates for schizophrenia (fig. 1: MOXD1 (monooxygenase, dopamine-β-hydroxylase-like 1) (Chambers et al. 1998), STX7 (syntaxin 7) (Wang et al. 1997), and all known human trace amine receptor (TRAR) genes—namely, TRAR1, TRAR3, TRAR4, TRAR5, PNR (putative neurotransmitter receptor gene) (Zeng et al. 1998; Borowsky et al. 2001; Bunzow et al. 2001; Lee et al. 2001), and three TRAR pseudogenes (TRAR2, GPR57, and GPR58 [GPR57 and GPR58 are G-protein-coupled receptor pseudogenes]) (Liu et al. 1998; Borowsky et al. 2001; Bunzow et al. 2001; Lee et al. 2001).

Trace amines (TAs) are endogenous amine compounds that are chemically similar to classic biogenic amines like dopamine (DA), norepinephrine, serotonin, and histamine. Abnormalities that involve the classic biogenic amines are the basis for biological hypotheses for a wide variety of disorders, including dystonias, Parkinson disease, schizophrenia, drug addiction, and mood disorders. In mammals, TAs are present at low levels, with no apparent dedicated synapses, but a blockage of amine degradation leads to significant accumulations of TAs, suggesting high synthesis and turnover, as reviewed elsewhere (Premont et al. 2001). TAs in mammals include tyramine (TYR), tryptamine, β-phenylethylamine (β-PEA), and octopamine (OCT) (Brancheck and Blackburn 2003), and all are synthesized from amino acid precursors by the aromatic amino acid decarboxylase.

TAs were thought to be “false transmitters,” which displace classic biogenic amines from their storage and act on transporters in a fashion similar to the amphetamine (Parker and Cubeddu 1986), but the identification of brain receptors specific to TAs indicates that they also have effects of their own (Borowsky et al. 2001). This might explain the fact that, although TYR, β-PEA, OCT, and amphetamine would require the integrity of vesicular stores of DA if displacement of DA were their only mechanism of action, they (except OCT) are still active when DA is depleted (Baud et al. 1985). TRARs bind amphetamine, MDMA (3,4-methylenedioxymethamphetamine, known as “ecstasy”), and LSD (D-lysodergic acid diethylamide) with high affinity. This suggests a direct link between TRARs and mechanisms of psychosis, because the administration of amphetamine can induce a schizophrenia-like psychosis (Connell 1958; Snyder et al. 1967; Angrist et al. 1974; Laruelle and Abi-Dargham 1999). In addition, psychedelic experiences induced by LSD can have a remarkable similarity to schizophrenia (Vardy and Kay 1983; Gouzoulis et al. 1994). Furthermore, LSD can induce habituation deficits (the normal decrease in response magnitude to repeated stimuli over time) that are similar to those exhibited by patients with schizophrenia (Geyer and Braff 1987; Braff et al. 1992).

MOXD1 is a homologue of dopamine-β-hydroxylase that is potentially involved in the biosynthesis of norepinephrine from DA (Chambers et al. 1998). Syntaxin 7 (STX7) is a critical component of the synaptic protein complex SNARE (receptor for soluble N-ethylmaleim-
ide-sensitive factor attachment proteins), which is involved in NMDA receptor and dopaminergic receptor function (Pei et al. 2004); the dysfunction of SNARE has been suggested in schizophrenia (Honer et al. 2002). Specifically, syntaxins mediate vesicle fusion in vesicular transport processes (Teng et al. 2001). We investigated DNA polymorphisms in the MOXD1-STX7-TRARs cluster by use of family-based association methods, and here we present preliminary evidence of association between TRAR4 and schizophrenia.

Material and Methods

Subjects and Phenotypes

Three samples were studied, which we call the “NIMH-IRP,” “NIMH-GI,” and “AU/US” collections. Ascertainment of the NIMH-IRP sample was initially described by Gershon et al. (1988). The full sample—from which the present sample of 67 pedigrees was drawn—was described later (Cao et al. 1997; Gejman et al. 2001). The collection of the NIMH-GI sample was described in a report of a genome scan of 71 pedigrees (Cloninger et al. 1998), and additional NIMH-GI families were subsequently included in the repository-based data set (see the “Electronic-Database Information” section). Of these NIMH-GI pedigrees, 69 were used in the present analysis and for two previous ones (Cao et al. 1997; Martinez et al. 1999). The AU/US sample was described initially in a report of a genome scan of 43 pedigrees (Levinson et al. 1998); full or partial trios for the present study were from 56 of the 71 pedigrees in

Figure 1 Genomic structure of the 6q23.2 gene cluster and association mapping of the initial screening. The genomic positions are based on the UCSC July 2003 assembly of the human genome (see the UCSC Genome Bioinformatics Web site). a, The relative position of the 6q23.2 gene cluster to the peak markers from various linkage studies: D6S424 (Cao et al. 1997; Martinez et al. 1999), D6S416 (Cao et al. 1997), D6S292 (Lerer et al. 2003), and D6S264 (Lindholm et al. 2001). b, Genes in the 6q23.2 gene cluster. c, The log transformation of the FBAT P value for the 31 SNP markers analyzed in the initial association screening. Each data point of the markers points to its relative position in the gene cluster shown in panel b. The most significant marker is rs4305745, with a P value of .0014 (see table A5 [online only], for FBAT P values and other detailed information for all the initially selected markers, and table 1, for the single-marker association results for all the additional TRAR4 markers examined in the dense mapping effort).
the expanded sample used in linkage fine-mapping studies (Mowry et al. 2000) and in additional analyses of this data set (Martinez et al. 1999; Levinson et al. 2000). For the present study, we genotyped a total of 827 individuals from 192 families (67 NIMH-IRP families, 69 NIMH-GI families, and 56 AU/US families). Details are provided in table A1 (online only) and in the “Electronic-Database Information” section. Schizophrenia and schizoaffective disorder were diagnosed using the criteria of the DSM-IIIR (American Psychiatric Association 1987). The institutional review board of the Evanston Northwestern Healthcare Research Institute approved this study.

SNP Selection and Genotyping

SNPs were selected from public databases with the help of a bioinformatics tool, SNPper (Riva and Kohane 2002), and novel TRAR4 SNPs were identified by direct sequencing. The DNA samples were genotyped using one of two methods: (1) template-directed dye-terminator incorporation with fluorescence-polarization detection (FP-TDI) (Chen et al. 1999) or (2) the TaqMan assay developed by Applied Biosystems. For the FP-TDI assays, after PCR amplification of genomic DNA, the AcycloPrime-FP SNP detection kit (PerkinElmer) was used for post-PCR cleanup and the single-base extension reaction. We detected FP by use of either an Analyst fluorescence reader (LJL Biosystems) or a Wallac Victor3 (PerkinElmer), and FP data were converted to genotypes with the assistance of an automated genotype-calling spreadsheet (Akula et al. 2002). PCR primers and probes for the FP-TDI assays were designed using Primer3 (Rozen and Skaletsky 2000). For the TaqMan assays, the genomic sequence flanking the SNP was submitted to Applied Biosystems for development of an assay-by-design. Each unique TaqMan minor-groove–binding (MGB) allele-specific probe was labeled with either a 5′-FAM or a 5′-VIC reporter dye. PCR amplification of genomic DNA was performed in a 384-well plate in an ABI Prism 7900 or a DNA Engine Tetrad 2 (MJ Research). After PCR, the allele discrimination was performed on an ABI Prism 7900 Sequence Detection System by use of Sequence Detector Software (SDS), version 2.0. Standard genotype calling was converted by a customized spreadsheet. Nucleotide sequences for the PCR primers, the FP-TDI and TaqMan probes, and related information for each marker can be found in table A2 (online only).

The average completion rate of our experiments was 96%. To empirically check for errors in the genotyping method, we compared the results for marker rs4305745 from both methods; the difference rate between FP-TDI and the TaqMan assay was ~0.25% (data not shown). We used MERLIN (Multipoint Engine for Rapid Like-

hood Inference) (Abecasis et al. 2002) (with all the SNPs at once) to check for Mendelian inconsistencies, blanked them as described below (sometimes for individuals and sometimes for the family, when the error could not be traced to a particular individual), and then addressed all unlikely recombinants. Genotyping errors were detected for 0.17% of genotypes (MERLIN) (95 errors/54,611 nonzero genotypes), including 26 Mendelian inconsistencies (0.047%) and 69 unlikely recombinants (0.12%). We did not change genotypes for unlikely recombinants unless (1) MERLIN estimated a high probability of an individual genotype error, compared with other possible errors (see MERLIN documentation for details), and/or (2) manual rereading of each genotype tracing or other raw genotyping output for the family/marker in question pinpointed a specific error for a particular individual. All genotype errors (Mendelian inconsistencies or specific errors that resulted in unlikely recombinants) were blanked (zeroed) for the involved individuals, and, given our high genotyping completion rate and low genotyping error, we did not perform a second-pass genotyping procedure for these individuals. Genotypes were read blind to psychiatric status. We checked Hardy-Weinberg equilibrium (HWE) for family founders (at least 200 individuals) for all 55 SNPs. Three MOXD1 SNPs were found not to be in HWE (rs2206064, rs1981187, and rs2275394 [although the HWE P value for rs2206064 did not remain significant after the number of markers examined for HWE was taken into account]). The minor-allele frequency of rs2206064 was only 2%, which might explain the lack of HWE. The frequencies of homozygotes for both rs1981187 and rs2275394 were higher than expected (this was not a bias introduced by cleaning—only a handful genotypes were blanked, and they were not primarily heterozygous genotypes). We suspected that our mixed-pedigree sample might have contributed to the absence of HWE, and, indeed, we found that these MOXD1 SNPs were all in HWE when the European-ancestry (EA) founders and the African American (AA) founders were analyzed separately (data not shown).

Intermarker Linkage Disequilibrium (LD) Analysis

LD between the SNPs was estimated with the program ldmax from the GOLD (Graphical Overview of Linkage Disequilibrium) package (Abecasis and Cookson 2000) by use of the genotypes from unrelated founders. The program ldmax estimates haplotype frequencies from genotype data by employing an expectation-maximization algorithm (Excoffier and Slatkin 1995). The standard and normalized Lewontin’s disequilibrium coefficients (D, D′) are derived. Association significance is assessed from a χ² distribution with (n₁ − 1)(n₂ − 1) de-
Association Analysis

To detect LD with illness, we used the transmission/disequilibrium test (TDT), as implemented in the Family-Based Association Test (FBAT) program, version 1.5 (Laird et al. 2000; Rabinowitz and Laird 2000). The null hypothesis of interest is the absence of association in the presence of linkage. We thus employed the empirical-variance estimator (the \(-e\) flag option in the FBAT program) to account for SNP-genotype correlations among affected siblings that result from linkage. The FBAT test statistic uses a score function, \[ Z = S_j - \frac{E(S_j)}{\text{Var}(S_j)}, \] where \( S_j \) is the observed number of transmitted marker alleles, \( j \), to affected offspring, and \( E(S_j) \) and \( \text{Var}(S_j) \) are the expected and variance values of \( S_j \) under the null hypothesis. Asymptotically, \( Z \) is assumed to follow a normal distribution, with a mean and a variance equal to 0 and 1, respectively. The test statistic can also be expressed as \( Z^2 \), which follows a \( \chi^2 \) distribution, with 1 df. The FBAT program is able to deal with the transmission of multilocus haplotypes, even when phase is unknown and parental genotypes may be missing. It can use pedigrees as well as nuclear families, but pedigrees are broken down into all individual nuclear families; only informative families (i.e., those contributing to the test statistic) are included. For the analyses of the screening SNPs, alleles and haplotypes were tested for association if they were present in at least 10 informative families; in our data, this corresponds to not testing alleles and haplotypes rarer than 3%. This restriction, however, was not used when the investigation was limited to specific subsets of families in the secondary analyses. For multilocus association analyses, FBAT provides global \( P \) values, which assess the significance of transmission distortion for all the tested haplotypes. In the present analyses, we limited the number of multilocus systems tested by use of a stepwise procedure and by restriction of multilocus tests to combinations that included the SNP with highest single \( Z \) score value, as further detailed in the “Results” section below. For the FBAT analyses, we assumed an additive model for each SNP, and only one affection status model was used (affected subjects had schizophrenia or schizoaffective disorder). The additive model is expected to perform well even when the true model is nonadditive.

Linkage Analysis

Model-free linkage analyses with the MOXD1-STX7-TRARs gene cluster were performed using the LOD score test from the affecteds-only sharing method (Kong and Cox 1997), as implemented in the MERLIN program (Abecasis et al. 2002). The likelihood of the observed marker information among affected relatives is maximized as a function of the marker-allele–sharing parameter and is compared, using a likelihood-ratio test, with the likelihood of the marker data under the null hypothesis of no linkage. The resulting distribution of the allele-sharing test (\( T \)) is \( \chi^2 \), with 1 df, and the statistic can also be reported as a LOD score of \( T/2 \ln(10) \).

We performed additional analyses, by race, to account for putative genetic and/or allelic heterogeneity within our family samples. Association and linkage tests were evaluated separately in AA and EA families.

Mutation Detection

Sequencing of TRAR4 was performed on an ABI 3100 Genetic Analyzer. Purified PCR products from various amplicons of relevant genomic DNA fragments were used as templates in sequencing reactions with the chemistry of BigDye 3.1 (Applied Biosystems). PCR primers were designed by Primer3 (Rozen and Skaletsky 2000) and were also used as sequencing primers for forward and reverse sequencing. The primer sequences and product sizes are given in table A3 (online only). We used the software SeqScape, version 2.1 (Applied Biosystems), to assist in mutation detection, and we visually verified each mutation. The reference sequence of TRAR4 used in the analysis was from the University of California–Santa Cruz (UCSC) human genome draft, July 2003 freeze (see the UCSC Genome Bioinformatics Web site).

For the nonhuman primates, DNAs were extracted from peripheral blood samples of two different chimpanzees (PTR-S109 and PTR-S286) from West Africa and from tissue samples of two different lowland gorillas (GGO-S110 and GGO-S249). The forward primer for amplicon 1 and the reverse primer for amplicon 7 were used to PCR-amplify the entire DNA segment by standard methods, with an annealing temperature of 60°C. This product was then sequenced bidirectionally with the seven primer pairs that are detailed in table A3 (online only). PCR products were confirmed by 1.5% agarose gel electrophoresis and were purified using MicroSpin Columns (Amersham Biosciences). The purified PCR products were sequenced using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) on an ABI Prism 377/3100 DNA sequencer. Sequence data were assembled by the Phred/Phrap program (Ewing et al. 1998) and also were checked manually using the Consed program (Gordon et al. 1998). Sequence data with reads for both strands and/or with high quality (\( >30 \) quality value) were used and deposited into the DDBJ/EMBL/GenBank International Nucleotide Sequence Database (accession numbers AB180397–AB180400).

RT-PCR and Real-Time PCR

Total mRNAs from various brain tissues were purchased from either BD Biosciences or Ambion. Gene expression of TRAR4 was first confirmed with general RT-
PCR with primer pairs used previously for amplification of segment 4 of TRAR4 (table A3 [online only]). In brief, total mRNA was reverse-transcribed with TaqMan Reverse Transcription Reagents (Applied Biosystems), and the synthesized first-strand cDNAs were then used as templates to amplify TRAR4 with HotStarTaq polymerase (Qiagen). β-actin was used as an internal control in the RT-PCR.

Reverse-transcribed cDNAs were also used in real-time PCR on an ABI Prism 7900 Sequence Detection System, in accordance with the manufacturer’s protocol. The TaqMan MGB probes and PCR primer pairs for the gene-expression assay of TRAR4, GAPD, (glycer-aldehyde-3-phosphate dehydrogenase), and/or TRAR1 were purchased as an Assay-On-Demand from Applied Biosystems. The relative gene expression in different brain tissues was normalized to GAPD expression by use of the standard-curve method, as described by Applied Biosystems.

Bioinformatics Tools for Prediction of Functional Effects of Genetic Polymorphisms

We used SIFT and PolyPhen to predict the potential functional effect of missense polymorphisms (Ramensky et al. 2002; Ng and Henikoff 2003). We used VISTA, which defines the conserved region among genomic sequences from different species (Couronne et al. 2003), to predict the potential regulatory sequences.

Results

We have studied 192 families with previous evidence of linkage. Of the 33 SNPs initially selected for study, 31 were accepted for analysis in the screening experiment, since 2 of the MOXD1 SNPs, rs2206064 and rs7751860, had minor-allele frequencies resulting in <3% informative families. The screening SNPs spanned ∼500 kb of the MOXD1-STX7-TRARs gene cluster, which contains a prime set of positional and pathophysiological candidates for schizophrenia susceptibility. We selected at least one common SNP for each gene, with a minimum coverage of /H11091/1 SNP/20 kb (the screening set of SNPs). Linkage analyses confirmed the presence of excess allele sharing in this region with individual SNPs from the MOXD1-STX7-TRARs gene cluster. Nine SNPs showed linkage P values <.05 (table A4 [online only]), with the most significant one being rs6937506 (TRAR4), with a LOD score of 1.76 (P <.002).

Association results are presented in table 1 and can also be seen in table A5 (online only). Of 31 screening
SNPs, 4 SNPs spanning 106 kb, 2 located in TRAR4 and 1 each in STX7 and GPR57, yielded a P value <0.05. The most significant one, rs4305745, was found in TRAR4 (P = .0014) (fig. 1c). This SNP was the only one that remained significant after Bonferroni correction for 31 tests (it should be noted that the question of whether Bonferroni corrections are appropriate or overly conservative is open to different interpretations; however, given that LD is likely to be present over short physical distances, the tests cannot be considered independent). The SNP rs4305745 is located 1,214 bp downstream from the stop codon of TRAR4. We concentrated further laboratory efforts on TRAR4. We aimed for a high-density map of >1 SNP/2 kb by searching public SNP databases and sequencing genomic DNA. We sequenced the TRAR4 gene (~1 kb of the 5’ region, the 1,038-bp CDS (coding DNA sequence), and ~1.5 kb of the 3’ UTR, which spans rs4305745) in 30 probands selected from the NIMH-GI families: 16 EA families and 14 AA families. We found 10 coding variants (26 total variants [table 2 and online-only table A6]) by sequencing TRAR4. Three of these variants had previously been found in 96 healthy EA individuals (Freudenberg-Hua et al. 2003). Of the seven novel variants, five are missense and are present only in AA individuals, as shown in table 2. From this combined bioinformatics and sequencing effort, 20 SNPs spanning 21.6 kb of the TRAR4 gene (>1 SNP/2 kb) were identified for genotyping. Two additional markers 3’ to rs4305745 showed association with schizophrenia in the whole sample (table 1 and online-only table A7): rs6903874 (P = .0026) and rs6937506 (P = .0052). Two other additional markers (ss28447873 [an insertion/deletion polymorphism] and rs7452939) were found to be in perfect LD with rs4305745 (they span 53 bp) and, hence, were also associated with illness to the same degree as rs4305745.

Although similar association trends were observed in EA and AA individuals, table A7 (online only) shows that the evidence for association was proportionally stronger in the AA sample. The evidence for association with rs4305745 and rs6937506 was significant in both samples (P = .035 in the AA sample and P = .015 in the EA sample for rs4305745; P = .025 in the AA sample and P = .035 in the EA sample for rs6937506). However, a cluster of three SNPs in the promoter region of TRAR4 yielded significant association in the AA sample—rs4473885 (P = .032), rs4085406 (P = .047), and rs6907909 (P = .019)—but not in the EA sample (table A7 [online only]). These differences raise the possibility of allelic heterogeneity specific to the AA population. All new missense mutations detected by sequencing were also exclusive to the AA sample (table 2). Two mutations, A518G (Tyr173Cys) and C683T (Ala228Val), were predicted by either PolyPhen or SIFT (Ramensky et al. 2002; Couronne et al. 2003; Ng and Henikoff 2003) to be nonconservative (table 2). How- ever, after genotyping the whole sample, we noted that none of the newly found missense SNPs cosegregated with disease in a specific manner (data not shown).

<table>
<thead>
<tr>
<th>dbSNP Accession No.</th>
<th>Position (bp)</th>
<th>SNP</th>
<th>Amino Acid Change</th>
<th>Type of Change</th>
<th>Function</th>
<th>Location in Protein</th>
<th>Minor-Allele Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs8192622</td>
<td>132,872,108</td>
<td>C78T</td>
<td>Pro26Pro</td>
<td>Synonymous</td>
<td>NA</td>
<td>Extracellular</td>
<td>.05 .09 .00</td>
</tr>
<tr>
<td>ss28447862</td>
<td>132,872,140</td>
<td>T110C</td>
<td>Ile37Thr</td>
<td>Conservative</td>
<td>Tolerated</td>
<td>TMD 1</td>
<td>.03 .00 .07</td>
</tr>
<tr>
<td>ss28447876</td>
<td>132,872,523</td>
<td>G493A</td>
<td>Gly165Ser</td>
<td>Nonconservative</td>
<td>Tolerated</td>
<td>TMD 4</td>
<td>.02 .00 .04</td>
</tr>
<tr>
<td>ss28447863</td>
<td>132,872,548</td>
<td>A518G</td>
<td>Tyr173Cys</td>
<td>Nonconservative</td>
<td>Damaging</td>
<td>Extracellular</td>
<td>.03 .00 .07</td>
</tr>
<tr>
<td>ss28447864</td>
<td>132,872,660</td>
<td>C630G</td>
<td>Thr210Thr</td>
<td>Synonymous</td>
<td>NA</td>
<td>TMD 5</td>
<td>.03 .03 .04</td>
</tr>
<tr>
<td>ss28447865</td>
<td>132,872,713</td>
<td>C683T</td>
<td>Ala228Val</td>
<td>Conservative</td>
<td>Damaging</td>
<td>Cytoplasmic</td>
<td>.10 .00 .21</td>
</tr>
<tr>
<td>ss28447867</td>
<td>132,872,774</td>
<td>A744G</td>
<td>Arg248Arg</td>
<td>Synonymous</td>
<td>NA</td>
<td>Cytoplasmic</td>
<td>.02 .00 .04</td>
</tr>
<tr>
<td>rs8192624</td>
<td>132,872,823</td>
<td>G793A</td>
<td>Val265Ile</td>
<td>Conservative</td>
<td>Tolerated</td>
<td>TMD 6</td>
<td>.10 .16 .04</td>
</tr>
<tr>
<td>rs8192625</td>
<td>132,872,902</td>
<td>G872A</td>
<td>Cys291Tyr</td>
<td>Nonconservative</td>
<td>Tolerated</td>
<td>TMD 7</td>
<td>.13 .16 .11</td>
</tr>
<tr>
<td>ss28447866</td>
<td>132,873,006</td>
<td>G976A</td>
<td>Val326Ile</td>
<td>Conservative</td>
<td>Tolerated</td>
<td>Cytoplasmic</td>
<td>.03 .00 .07</td>
</tr>
</tbody>
</table>

A | previously known SNPs are indicated by an rs number; novel SNPs are indicated by an NCBI Assay ID number (ss number) that we obtained (data to be released in the next build of dbSNP).
B | Position in bp was derived from the UCSC July 2003 freeze of chromosome 6 (see the UCSC Genome Bioinformatics Web site).
C | SNPs were named according to their relative position to the first letter of start codon ATG and with the first base as the major allele.
D | Function determined by SIFT. NA = not applicable.
E | The position in the protein was predicted on the basis of 7-transmembrane GPCR structures, as depicted in the GPCR database (see GPCRDB Web site). TMD = transmembrane domain.
F | EA = European ancestry.
G | AA = African American.
thermore, all the missense variants, except for A518G (Tyr173Cys), were also found in a set of 48 AA subjects from the Coriell Human Variation AA DNA panel (table A8 [online only]). Additionally, some of these missense variants were homozygous in some control individuals. It is interesting to note that the ratio of missense to synonymous mutations (9:3) is close to what is predicted under neutral expectations (i.e., a pseudogene that has an expected ratio close to 4:1). Indeed, in the gorilla, \( \text{TRAR4} \) has already become a pseudogene (with a nonsense mutation at codon 15: Tyr in chimpanzee and human and STOP in gorilla), although the human and chimpanzee \( \text{TRAR4} \) versions have not yet become pseudogenes.

The \( \text{TRAR4} \) region was found to have two LD blocks, which are depicted in figure 2. The SNP rs4305745 (marker 16 in fig. 2) is in the LD block constituted by 3’-flanking SNPs. The pattern suggests that association for \( \text{TRAR4} \) originates from rs4305745. None of the 5’-flanking SNPs are in LD with rs4305745, which instead is in strong LD with markers 19 (rs6903874) and 21 (rs6937506) from the 3’ LD block (and also shows a trend with marker 12 [rs8192625]). The LD pattern generated from the 31 initial screening markers indicated that the whole region of the \( \text{MOXD1-} \text{STX7-} \text{TRARs} \) gene cluster is separated into four major strong-LD blocks, whereas the \( \text{TRAR4} \) region represented by rs4305745 is not in strong LD with any of the major LD blocks (data not shown).

We conducted haplotype association analyses with all \( \text{TRAR4} \) two-locus systems (\( n = 17 \), after exclusion of 5 markers with minor-allele frequencies <3%) derived from rs4305745, which was chosen as the anchor since it had the most significant single-locus association (\( P = .0014 \)). For each two-locus system, we derived the global \( \chi^2 \) value by use of only those haplotypes with frequencies >3% (table A9 [online only]). All 17 of these two-locus systems exhibited \( P \) values <.05 (not corrected by multiple testing), and all harbored the “A” allele of rs4305745. None of the two-locus systems

**Figure 2**  Pairwise LD in 192 founders for the \( \text{TRAR4} \) region. \( a \). Relative physical position of 23 markers in the \( \text{TRAR4} \) region (shown in table 1). The SNPs ss28447862, ss28447876, ss28447865, rs8192624, and ss28447866 were excluded in the following LD measurements because of their low minor-allele frequencies. \( b \). LD pattern in AA subjects (left panel) and EA subjects (right panel). The graph was generated by GOLD (Abecasis and Cookson 2000). In each LD pattern, the \( D' \) values (upper left diagonal) and the \( P \) values (lower right diagonal, converted to log \( P \) values) are calculated from the program ldmax of the GOLD package (Abecasis and Cookson 2000).
showed association stronger than that of rs4305745 alone (table A9 [online only]), and the same results were obtained with up to five multilocus systems, with each haplotype system extended stepwise to contain the most significant previous smaller haplotype (data not shown). This suggests that rs4305745 and/or the other two nearby polymorphisms, ss28447873 and rs7452939, in perfect LD (further confirmed by genotyping the whole sample) with rs4305745 (table A6 [online only]) are the most likely mutations underlying the association of the TRAR4 region with schizophrenia susceptibility.

To explore the possible functional effects of associated SNPs and their haplotypes, we first defined the conserved noncoding sequence (considered as a potential functional region) by comparative genomic analysis of TRAR4 genomic sequences of human, mouse, and rat by use of VISTA (Couronne et al. 2003). The cluster of three polymorphisms (rs4305745, ss28447873, and rs7452939—all equally implicated as candidates by the association analysis) that exhibit the most significant association is very close to two conserved regions right after the stop codon (sequence identity >70% among human, mouse, and rat genomes). The sequence identity immediately around this SNP (rs4305745) is ~50% (fig. A1 [online only]), suggesting that this SNP and/or other polymorphisms in perfect LD with rs4305745 may ultimately affect gene expression, a hypothesis we are currently testing. RT-PCR from various brain tissues also confirmed that rs4305745 was flanked by the 3′ UTR of TRAR4 (data not shown), suggesting that rs4305745 or one of its haplotypes may affect gene expression at the posttranscriptional level. Another significant SNP, rs6903874, is also within a conserved region (fig. A1 [online only]), and it is possible that this SNP and/or other SNPs nearby might be functional, although this SNP is much farther than rs4305745 from the stop codon of TRAR4.

A comprehensive gene-expression analysis of TRAR4 will help to elucidate its potential functional roles in the pathophysiology and pharmacology of schizophrenia. We investigated TRAR4 expression in various human tissues by RT-PCR and found that TRAR4 was expressed at low abundance in various human brain tissues as well as in human fetal liver but not in the cerebellum or placenta (fig. 3a). A quantitative real-time PCR, shown in figure 3b, further revealed that TRAR4 has comparable levels of expression in basal ganglia, frontal cortex, substantia nigra, amygdala, and hippocampus, with the highest expression in hippocampus and the lowest expression in basal ganglia. These results are consistent with a previous expression study that included TRAR4 (Borowsky et al. 2001). These regions have been implicated in the pathophysiology and pharmacology of schizophrenia (Grossberg 2000; Freedman 2003). The tissue distribution of TRAR4 gene expression is similar to that of the only well-characterized TRAR, TRAR1 (Borowsky et al. 2001; Bunzow et al. 2001). However, further comparison of gene expression of TRAR4 and TRAR1 indicated that TRAR4 is, overall, more abundant than TRAR1, particularly in basal ganglia (~14-fold), frontal cortex (~21-fold), and substantia nigra (~14-fold) (fig. 3c), which suggests that...
**Discussion**

TA receptor genes have been proposed as candidate genes for schizophrenia on the basis of highly provocative pathophysiological evidence (Boulton 1980; Preumont et al. 2001; Branchek and Blackburn 2003) as well as linkage mapping data (Cao et al. 1997). Here, we present preliminary evidence that TRAR4, a gene that belongs to the TRAR family, contributes to susceptibility to schizophrenia in three data sets, with evidence of genetic linkage to 6q. Furthermore, the TRARs gene cluster at chromosome 6q23 is contained within a wide area of linkage detected in several other clinical samples (Bailer et al. 2000; Levinson et al. 2000; Lindholm et al. 2001; Lerer et al. 2003; Lewis et al. 2003). The linkage evidence for schizophrenia in 6q is not population specific—it has been gathered from multiple population groups: AA individuals, EA individuals, and Jews and Arabs from Israel. However, the evidence for association of TRAR4 in our samples, although present in both EA and AA individuals, appears higher in AA subjects.

The most significantly associated SNPs within TRAR4 are located in the 3′ UTR of TRAR4, and these polymorphisms may affect the gene expression at the posttranscriptional level. There are several possible mechanisms by which TRAR4 could contribute to the pathogenesis of schizophrenia. First, the to-be-determined ligand(s) of TRAR4 may be neurotransmitter(s) specifically implicated—by a new signaling pathway—in schizophrenia. Second, TRAR4 may be activated strongly by β-PEA, an endogenous chemical analogue of amphetamine—or perhaps by amphetamine itself. It has been documented that amphetamine can produce a paranoid schizophrenia syndrome in humans by inducing DA release in the striatum, and the DA release induced by amphetamine (or β-PEA) is increased in patients with schizophrenia (Laruelle et al. 2003). Pharmacological profiling has shown that β-PEA and amphetamine can directly activate TRAR1 (Borowsky et al. 2001; Bunzow et al. 2001). Finally, TRAR4 may crosstalk with the COMT system, which has been implicated in the pathogenesis and etiology of schizophrenia (Li et al. 2000; Egan et al. 2001; Shifman et al. 2002). It has been shown that rat TRAR1 can be activated with higher potency and efficacy by 3-methyltyramine, a major metabolite of DA produced by COMT, than by the precursor catecholamines (Bunzow et al. 2001). If TRAR4 were shown to act like TRAR1 in this regard, some “inactive” catecholamine metabolites might act as endogenous agonists of TRAR4, which could predispose to psychosis. Moreover, we have shown that TRAR4 is more abundant than TRAR1 (fig. 3), which suggests that TRAR4 may play a more important role in areas of the brain that are relevant to the dopaminergic system. It can be seen that the plausible contributions of TRAR4 to schizophrenia susceptibility are closely related to the DA hypothesis of schizophrenia, a major hypothesis that has been intensively researched over the past 50 years. Given the increased understanding about the interactions of the DA system with the glutamate system in the pathogenesis of schizophrenia (Laruelle et al. 2003), our finding about TRAR4 may prompt new investigations of the interactions among these neurotransmitter systems.

Regulatory sequence disruption can affect protein expression and cause disease (Mitchison 2001). The associated SNPs in the 3′ UTR of TRAR4 may contribute to the susceptibility for schizophrenia by affecting the gene expression at the posttranscriptional level. Our RT-PCR experiment indicated that the TRAR4 3′ UTR spanned the most associated SNP, rs4305745; therefore, it is possible that TRAR4 gene expression was affected at the posttranscriptional level by these 3′ UTR SNPs (rs4305745 and/or ss28447873 and rs7452939 [two SNPs in perfect LD with rs4305745]). The chimpanzee and gorilla sequencing results indicated that the ancestral allele for rs4305745 is A (table A6 [online only]). Because of the perfect LD among rs4305745, ss28447873, and rs7452939, the associated allele A of rs4305745 actually represents a human haplotype of A-A-A, spanning rs4305745, ss28447873, and rs7452939 (fig. A1c [online only]). (Also note that the ancestral haplotype appears to be A-A-G, as seen in table A6 [online only].) It is interesting that the predicted TRAR4 mRNA structure exhibited a significant change for the overtransmitted haplotype A-A-A, compared with haplotype G-del-G; the same structure change can be generated by allele A of rs4305745 alone (data not shown), suggesting that rs4305745 is most likely the causative SNP.

We have found that the mutation rate in the coding region for TRAR4 (1 mutation/100 bp) is well above the average (1 mutation/346 bp) (Cargill et al. 1999). In addition, there are more missense mutations in TRAR4 than synonymous mutations (nine vs. three) (table 2 and online-only table A8), suggesting that TRAR4 may be becoming a pseudogene. However, an expressed pseudogene sometimes regulates the mRNA stability of a homologous gene (Hirotsune et al. 2003). We did not find evidence of association with schizophrenia for missense SNPs; however, some missense mutations—particularly A518G (Tyr173Cys)—may be pharmacologically important. A518G is located in the putative extracellular domain of the receptor and, hence, may affect ligand binding (table 2). Besides the changes in protein structure, these missense mutations may also alter the gene expression by affecting mRNA folding structures, as described for dopamine D2 receptor (DRD2) (Duan et al. 2003). Actually, A518G
was predicted to have a remarkable effect on TRAR4 mRNA folding, as predicted in silico by Mfold (Zuker et al. 1999) (data not shown). It is also notable that this missense SNP and several others were only found in AA subjects, and 173Cys was only detected in AA probands with schizophrenia. It will be important to examine the functional effects of some of these missense SNPs and to test the association with schizophrenia in additional samples.

Molecular genetics studies of schizophrenia have found several replicated linkages to various chromosomal regions (see review by Owen et al. [2004]), and association studies have recently pointed to several genes at some of those linkage regions with independent confirmations, including NRG1 at 8p21-p12 (Stefansson et al. 2002, 2003; Williams et al. 2003), DTNBP1 at 6p22.3 (Straub et al. 2002; Schwab et al. 2003), COMT at 22q11.21 (Li et al. 2000; Egan et al. 2001; Shifman et al. 2002), RGS4 at 1q23.3 (Chowdari et al. 2002; Morris et al. 2004; Williams et al. 2004), and DAOA at 13q33.2 (Chumakov et al. 2002; Schumacher et al. 2004). Although nonreplications have been reported, it would be extremely unlikely that all the aforementioned results will end as false positives. We now propose that TRAR4 is also a susceptibility gene for schizophrenia. There are also two other weak association peaks at the MOXD1-STX7-TRARs gene cluster, in addition to the TRAR4 peak shown in figure 1: one is in STX7 and the other is in the pseudogene, GPR5. Therefore, we cannot exclude the possibility that those loci may also contribute to susceptibility for schizophrenia. The identification of TRAR4 as a susceptibility gene for schizophrenia, which is consistent with human and animal models of toxic psychosis and is in agreement with the expression pattern of TRAR4 (expressed in frontal cortex, amygdala, and hippocampus), appears to substantiate the dopaminergic hypothesis of schizophrenia, but the exact mechanisms of disease mediated by TRAR4 remain to be elucidated. The two reports suggesting linkage of the same 6q chromosomal area to bipolar disorder raise the possibility that TRAR4 might be involved in the pathophysiology of both schizophrenia and bipolar disorder, and there is a precedent for a gene potentially involved in both disorders (Chumakov et al. 2002; Hattori et al. 2003; Chen et al. 2004; Schumacher et al. 2003). We would like to note a couple of caveats: first, our combined sample has a rather modest size in the context of complex genetics, and, second, additional gene(s) in the region might account for our association signal. However, although it is certainly true that gene-prediction programs have their limitations and different methods sometimes produce different results, at least one of them has been able to pick up all the TRARs genes and pseudogenes we know about (see the UCSC Genome Bioinformatics Web site [July 2003 freeze]). Although it is possible that gene-prediction programs have missed a G-protein-coupled receptor (GPCR) pseudogene, there is little chance they have missed an actual GPCR gene in this cluster since the GPCR motif is so robust. Other non-GPCR genes, however, might have been missed. Although the identification of TRAR4 as a susceptibility gene for schizophrenia is encouraging, it is important that our observations should be considered as preliminary until replication has been completed in additional samples.

Acknowledgments

We thank the patients and families for their participation. We thank Dr. Josep M. Comeron for helpful comments and suggestions regarding the evolutionary context of TRAR4. The data and biomaterials for the NIMH-GI families were collected in three projects. From 1991 to 1997, the Principal Investigators and Co-Investigators were as follows: Ming T. Tsuang, M.D., Ph.D., D.Sc., Stephen Faraone, Ph.D., and John Peppe, Ph.D. (Harvard University, Boston [U01 MH46318]); C. Robert Cloninger, M.D., Theodore Reich, M.D., and Dragan Svrakic, M.D., Ph.D. (Washington University, St. Louis [U01 MH46276]); Charles Kaufmann, M.D., Dolores Malaspina, M.D., and Jill Harkavy Friedman, Ph.D. (Columbia University, New York [U01 MH46289]). We thank Dr. Wendy Cao for providing the β-actin primers. We also acknowledge Drs. Elliot S. Gershon and Lynn E. DeLisi for work on recruitment of part of the sample. We thank Kumamoto Primate Park and Ueno Zoo in Tokyo for chimpanzee and gorilla samples, respectively. Research Career Development Awards (to J.D. and A.R.S.) at the Evanston Northwestern Healthcare Research Institute (Evanston, IL) supported this work.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

- dbSNP Home Page, http://www.ncbi.nlm.nih.gov/SNP/ (for the 18 novel SNPs [accession numbers ss28447859–ss28447876]; data will become available to the public when the National Center for Biotechnological Information releases the latest dbSNP build, and, at that time, the accession numbers will be converted into reference cluster ID numbers)
- G-protein-coupled receptors database (GPCRDB), http://www.gpcr.org/
- Mfold, http://www.bioinfo.rpi.edu/applications/mfold/ (for RNA structure prediction)
- National Institute of Mental Health (NIMH) Schizophrenia Genetics Initiative Data Archive, http://zork.wustl.edu/nimh/sz.html (families used in this study were 30101, 30103, 30104, 30106, 30108, 30110, 30111, 30112, 30113, 30114, 30116, 30119, 30122, 30123, 30124, 30126,
References


analysis of schizophrenia and bipolar disorder, part II: schizophrenia. Am J Hum Genet 73:34–48
Sanders AR, Gejman PV (2001) Influential ideas and experimental progress in schizophrenia genetics research. JAMA 285:2831–2833
regulin 1 and susceptibility to schizophrenia. Am J Hum Genet 71:877–892