Molecular Psychiatry (2006) 11, 125–133

www.nature.com/mp

ORIGINAL ARTICLE

# Identification of additional variants within the human dopamine transporter gene provides further evidence for an association with bipolar disorder in two independent samples

TA Greenwood[1], NJ Schork[1], E Eskin[2] and JR Kelsoe[1,3]

[1]*Department of Psychiatry, University of California at San Diego, La Jolla, CA, USA;* [2]*Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA, USA and* [3]*San Diego VA Health Care System, San Diego, CA, USA*

**The dopamine transporter (DAT) is the site of action of stimulants, and variations in the human DAT gene (*DAT1*) have been associated with susceptibility to several psychiatric disorders including attention deficit hyperactivity disorder (ADHD) and bipolar disorder. We have previously reported the association of bipolar disorder to novel SNPs in the 3′ end of *DAT1*. We now report the identification of 20 additional SNPs in *DAT1* for a total of 63 variants. We also report evidence for association to bipolar disorder in a second independent sample of families. Eight newly identified SNPs and 14 previously identified SNPs were analyzed in two independent samples of 50 and 70 families each using the transmission disequilibrium test. Two of the eight new SNPs, one in intron 8 and one in intron 13, were found to be moderately associated with bipolar disorder, each in one of the two independent samples. Analysis of haplotypes comprised of all 22 SNPs in sliding windows of five adjacent SNPs revealed an association to the region near introns 7 and 8 in both samples (empirical *P*-values 0.002 and 0.001, respectively, for the same window). The haplotype block structure observed in the gene in our previous study was confirmed in this sample with greater resolution allowing for discrimination of a third haplotype block in the middle of the gene. Together, these data are consistent with the presence of multiple variants in *DAT1* that convey susceptibility to bipolar disorder.**
*Molecular Psychiatry* (2006) **11**, 125–133. doi:10.1038/sj.mp.4001764; published online 1 November 2005

**Keywords:** genetic variation; single nucleotide polymorphism; haplotype

## Introduction

The human dopamine transporter gene (*DAT1, SLC6A3*) has been implicated as a candidate gene in several neuropsychiatric disorders, including ADHD and bipolar disorder.[1–6] Several lines of pharmacological evidence suggest a possible role for dopamine in bipolar disorder, particularly in mania, and suggestive evidence of linkage to bipolar disorder has been reported for several markers within and near the *DAT1* locus on chromosome 5p15.3.[5,7] We have previously reported that the strongest evidence to date for an association between *DAT1* and bipolar disorder: in a sample of 50 parent–proband trios, transmission/disequilibrium test (TDT) analysis revealed strong statistical evidence for association between a particular haplotype comprised of five SNPs deriving from the 3′ region of *DAT1*, exon 9 through exon 15, and bipolar disorder (allelewise TDT empirical $P=0.001$, genotypewise TDT empirical $P=0.0004$). All additional haplotyped markers comprised of SNPs from the 3′ region of *DAT1* were also found to be associated with bipolar disorder, suggesting the presence of a potential regulatory element within this region of the gene. However, we were unable to further refine and/or localize the region of interest using these SNPs.

We now report the discovery of 20 novel SNPs for a total of 63 variants spanning the *DAT1* gene and the analysis of an additional eight SNPs, in conjunction with the 14 SNPs we have previously reported. Since our initial sample was relatively small, we have attempted to replicate these results in a slightly larger and entirely independent sample of 70 parent–proband trios and further define the region possibly containing a functional variant contributing a susceptibility to bipolar disorder.

## Materials and methods

### *Subjects*

For an evaluation of association between *DAT1* and bipolar disorder, two independent samples of parent–

Correspondence: Dr JR Kelsoe, Department of Psychiatry, 0603, University of California, San Diego, La Jolla, CA 92093, USA.
E-mail: jkelsoe@ucsd.edu
Received 23 February 2005; revised 12 September 2005; accepted 13 September 2005; published online 1 November 2005

proband trios, each consisting of a proband with bipolar disorder (bipolar I or bipolar II) and both parents, were assembled from different sources. An initial sample of 50 trios was selected from the UCSD bipolar consortium family collection (collected at UCSD, U British Columbia, and U Cincinnati) and Old Order Amish pedigree 110 as previously described.[6] This initial sample was later augmented with an additional 70 parent–proband trios selected from waves 1 and 2 of the NIMH Genetics Initiative for Bipolar Disorder collection, for which genomic DNA was obtained.[8] In each sample, families were ascertained through a bipolar proband and for the presence of at least two other affected members. Diagnoses were made using a consensus best estimate process using either the Structured Clinical Interview for DSM-III-R (SCID, UCSD sample) or the Diagnostic Interview for Genetic Studies (DIGS, NIMH sample) as described previously. Written informed consent was obtained for all subjects using procedures approved by each institution's Institutional Review Board (IRB). In each case, one triad was chosen per family for a combined sample of 120 trios deriving from 120 separate families. When available, an affected sibling of the proband was also included as an adjunct to the triad to form a sibpair family. All subjects chosen for these studies were Caucasians of Northern European origin.

For some analyses, the combined set of 120 trios was stratified by linkage to *DAT1* as assessed by nonparametric linkage (NPL) scores $>0.2$ obtained from GENEHUNTER v.2.0.[9] Those 33 sibpair families that showed linkage to this locus, including six UCSD/UBC/UC trios and 27 NIMH trios, were selected for further association studies of bipolar disorder.

*Mutation screening*

Introns 8, 11, 12, 13, and 14 were exhaustively screened in overlapping segments of approximately 500 bp each using methods and samples that have previously been described.[6] Sequence for use as a reference was compiled from the complete *DAT1* genomic sequence[10] and previously obtained sequence from a bacterial artificial chromosome (BAC) containing *DAT1*.[6] All 62 confirmed SNPs were submitted to dbSNP and assigned accession numbers ss16333640–16333700 (http://www.ncbi.nlm.nih.gov/projects/SNP/). A table describing all SNPs and variants identified is available on our website (http://polymorphism.ucsd.edu).

*SNP genotyping and haplotype construction*

The 50 UCSD trios and 70 NIMH trios were genotyped for the original 14 SNPs and VNTR as previously described.[6,11] An additional eight SNPs of those newly identified were chosen for genotyping based on location and genotyping feasibility. Genotyping of these SNPs in the combined sample of 120 trios was performed by Qiagen Genomics with the Masscode System, which uses cleavable mass spectrometry tags in multiplex reactions.[12] PCR reactions with external primers resulted in single-band products for all SNPs, and genotyping quality, as assessed by cluster scores, was high. All SNPs were in Hardy–Weinberg equilibrium.

Following genotyping, complete haplotypes comprised of all 22 SNPs and the VNTR were constructed for each individual using linkage phase data obtained from allele-specific PCR and parent–offspring relationships. These haplotypes could be conclusively determined in most cases. For the few instances in which phase was ambiguous, haplotypes were constructed for the remaining markers, and the phase of the SNP in question was estimated based on the phase observed in the families for which this was not an issue. For calculations of D′, the questionable SNPs were removed from the haplotypes so as not to bias any of the results.

*Statistical analyses*

We used an extended version of the transmission/disequilibrium test (ETDT) that accommodates both biallelic and multiallele marker loci and provides for the analysis of parent–offspring data under two different models, allelewise and genotypewise.[13] The allelewise model attempts to establish a pattern of preferential transmission of certain alleles across all parental genotypes, while the genotypewise model considers every heterozygous parental genotype separately and tests each allele of a particular genotype for an excess of transmission. The empirical significance of each ETDT statistic was evaluated using the Monte Carlo approach (MCETDT) with 1000–10 000 replicates.[14]

Correction for multiple tests in this case is not straightforward. For example, although there were several haplotypes per window in the sliding window haplotype association analyses, these windows were overlapping and therefore not independent. The Linked sample included individuals that were also analyzed in the UCSD and NIMH sample and is thus not an independent sample set. Therefore, we used the method of Nyholt (http://genepi.qimr.edu.au/general/daleN/SNPSpD), which accommodates linkage disequilibrium (LD) between the SNPs, to determine the effective number of independent SNPs and the appropriate correction factor for these analyses.[15] Using this method we determined that there were 19 independent SNPs and a *P*-value of 0.003 needed to be reached to keep the type I error rate at 5%.

Parental gametic haplotypes were used for calculations of LD between the 22 SNPs and the 3′ VNTR. In order to simplify calculations, only the nine and 10-repeat alleles of the VNTR, which were observed in all but three parental haplotypes, were included in the analyses. The standardized, pairwise disequilibrium value D′, the range of which extends from $-1$ to $+1$ with $-1$ and $+1$ representing complete LD and zero representing free association, was chosen for calculations of LD for its reduced dependence on allele frequency.[16] D′ and the significance of LD

were calculated and visualized using the GOLD (Graphical Overview of LD) v.10 software package (http://www.well.ox.ac.uk/asthma/GOLD).[17]

The phylogenetic relationships between haplotypes in the three block regions of *DAT1* were computed by optimizing a probabilistic model that assumes exponential population growth after a population bottleneck. Under this assumption, any variants after the bottleneck occur in haplotypes that are also present in the modern population. Haplotypes that have no intermediates between them (i.e., that vary by $\geqslant 2$ SNPs) are assumed to be before the bottleneck and are therefore 'ancestral' haplotypes. 'Common' haplotypes are defined as being a single mutation or recombination event away from the ancestral haplotypes as such events are likely to have happened after a major population bottleneck. The algorithm identifies 'recent' haplotypes by considering rare ($<5\%$) haplotypes and explaining them with either a single gene conversion, mutation, or recombination event from the 'common' haplotypes. The remaining haplotypes are clustered into ancestral groups and any recombinants (resulting from meiotic crossover events) are identified. The algorithm is implemented as an extension to the HAP program[18] (http://www.calit2.net/compbio/hap).

## Results

### *Mutation screening and genotyping*

Although our previous analyses of the 14 *DAT1* variants indicated an association between bipolar disorder and the 3′ region of the gene, we were unable to further refine the region possibly containing a functional variant contributing a susceptibility to bipolar disorder.[6] As the first mutation screen had identified few variants in the 3′ region of the gene, a more comprehensive evaluation of the existing variation was warranted. Our efforts to further characterize the variation of the 3′ region of *DAT1* entailed a complete screen of all 3′ introns, including 11, 12, 13, and 14. Introns 2, 5, and 8 were also included as sequence was readily available. As in our first study, this work was focused on two subjects from families with evidence of linkage to *DAT1* and one control subject. All 15 exons, as well as introns 1, 7, 9, and 10 had been previously characterized.[6]

Encompassing approximately 16 kb of new sequence, this second round of mutation screening led to the identification of 92 deviations from the reference sequence, which included 33 single insertion/deletion polymorphisms, one of which was polymorphic in the samples sequenced, one 15 bp insertion/deletion polymorphism, and 58 base substitutions, 18 of which were polymorphic in the samples sequenced. The 20 newly identified variants are described on our website (http://polymorphism.ucsd.edu), along with the 43 previously identified SNPs, 14 of which have been characterized and analyzed for association to bipolar disorder.[6,11]

### *Single-SNP association analyses*

As a follow-up to our previous analyses, eight additional variants were chosen for genotyping in our UCSD family sample based on their locations within the gene and genotyping feasibility. In order to replicate and extend our previous results, the original 14 SNPs and the new eight SNPs were also genotyped in a new and independent set of 70 trios and sibpair families from the NIMH Genetics Initiative for Bipolar Disorder. The frequencies and results of the TDT tests for the additional eight SNPs in both samples are shown in Table 1. All of these SNPs had minor allele frequencies greater than 5%, and several were very common in the populations studied. The I13 + 1457 SNP was found to be moderately associated with bipolar disorder in the UCSD sample (empirical $P = 0.024$), while the I8 + 2086 SNP was found to be moderately associated with bipolar disorder in the NIMH sample (empirical $P = 0.036$). In both the cases, the G allele was found to be

**Table 1** Results of the TDT analyses of the eight newly identified SNPs in the UCSD and NIMH

| *SNP* | *Location* | *Variant* | *Heterozygosity* [a] | *Relative frequency* [b] | *UCSD* | | *NIMH* | |
|---|---|---|---|---|---|---|---|---|
| | | | | | *T:NT* | P-*value*[c] | *T:NT* | P-*value*[c] |
| P + 2459 | 1−71 | A→T | 0.510 | 0.576 | 32:39 | 0.490 | 40:41 | 1.000 |
| I5 + 448 | 920 + 448 | G→C | 0.502 | 0.595 | 40:39 | 1.000 | 40:52 | 0.247 |
| I7 + 922 | 1159 + 922 | C→A | 0.186 | 0.893 | 7:9 | 0.793 | 20:13 | 0.294 |
| I8 + 2086 | 1284 + 2086 | G→A | 0.538 | 0.549 | 30:33 | 0.789 | 53:34 | 0.036 |
| I11 + 2478 | 1626 + 2478 | C→T | 0.126 | 0.928 | 15:9 | 0.298 | 11:8 | 0.646 |
| I12 + 268 | 1727 + 267 | G→T | 0.130 | 0.930 | 15:9 | 0.298 | 10:8 | 0.817 |
| I13 + 1457 | 1895 + 1457 | G→A | 0.437 | 0.694 | 39:21 | 0.024 | 40:35 | 0.645 |
| I14 + 4217 | 1967 + 4217 | A→G | 0.146 | 0.918 | 18:11 | 0.243 | 11:8 | 0.646 |

[a]Observed heterozygosity among the parental genotypes of the combined set of 120 trios.
[b]Relative frequency of the common variant among the parental genotypes of the combined sample of 120 trios.
[c]Empirically-derived *P*-values for the allele-wise model of ETDT.

significantly transmitted. Analysis of the original 14 SNPs in the NIMH sample did not reveal an association for any SNP (data not presented).

*LD and cladistic analyses*

Before proceeding with haplotype association studies, we examined the extent of LD between the markers and modeled the phylogenetic relationships between haplotypes. Figure 1 shows the results of the pairwise LD analyses of all 22 SNPs and the 3′ VNTR. This figure illustrates the likely presence of three haplotype blocks within the gene: promoter – intron 2 (block 1), intron 5–7 (block 2), and exon 9 – exon 15 (block 3), which span 4.9, 5.6, and 18.2 kb, respectively. The average D′ in these regions is 0.92, 0.79, and 0.84, respectively. These data appear to indicate two areas of low LD, introns 3–4 and intron 8, possibly due to increased recombination.

The predicted phylogeny of the haplotypes belonging to block 3 is illustrated in Figure 2a. Only haplotypes with frequency greater than 5% are shown due to the existence of a high number of rare haplotypes and, consequently, an unnecessarily complex phylogenetic diagram. Three ancestral groups exist in this region: B ('GGAGCGAACG9C'), D ('AGGATTAGGG10T'), and the group consisting of K ('?TGGCGGAG?10T'). Most of the common variants are from the K group. Of the 29 rare haplotypes that occur in this region (not all shown in figure), 22 are either a single SNP or a recombination away from common haplotypes. Two ancestral groupings, 'AACGAGG' and 'GTAAGA?', which differ by the first 6 SNPs, and 12 rare haplotypes define the haplotype structure of block 1, as we have previously suggested (data not presented; see Figure 2b for haplotypes).[11,19] The haplotypes of block 2 belong to two ancestral groups, 'CCC' and 'GG?', which are differentiated

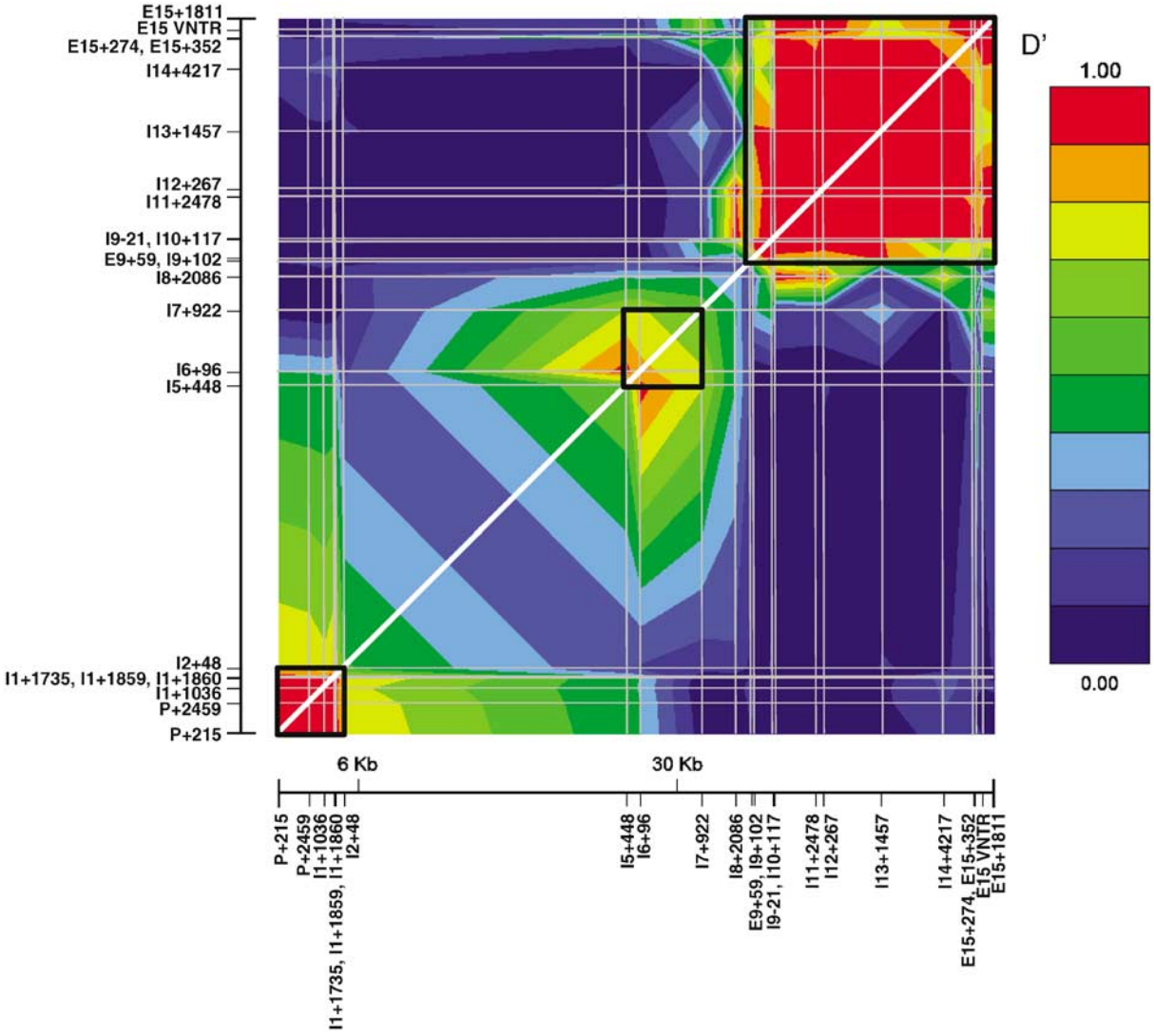


**Figure 1** Scaled GOLD diagram showing distribution of LD across *DAT1* as assessed by the 22 SNPs and 3′ VNTR. The three observed haplotype blocks, promoter – intron2, intron 5 – intron 7, and exon 9 – exon 15, are indicated as boxed regions (D′ > 0.7 for all SNP pairs within the region).

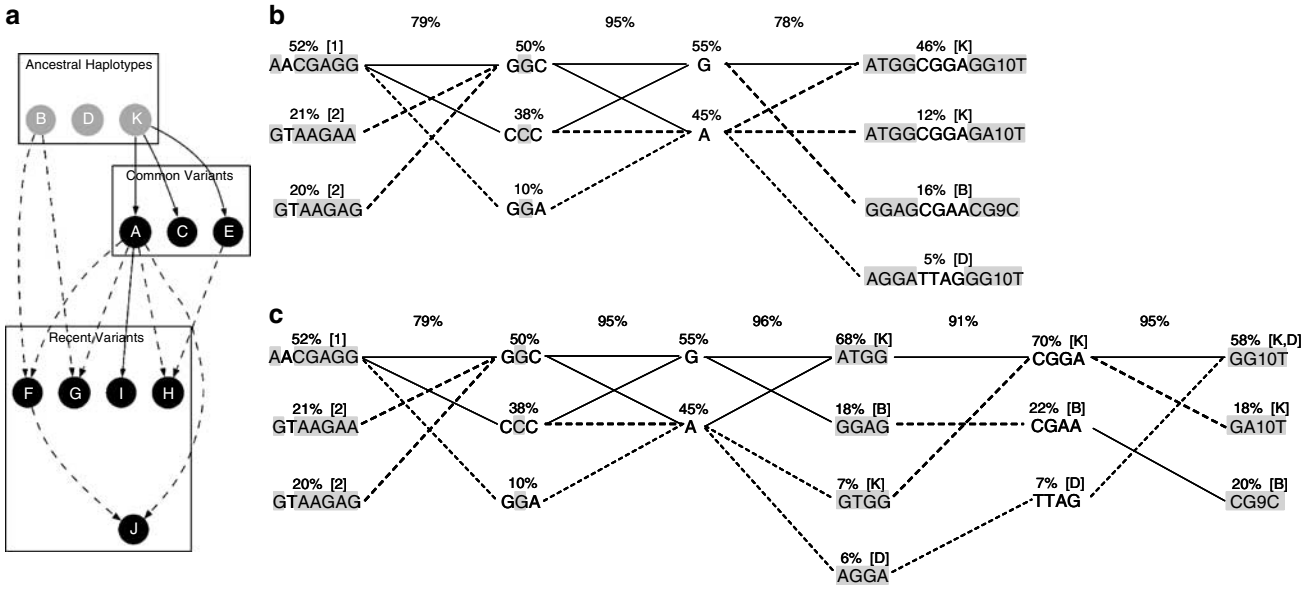


**Figure 2** Visualization of the structure of haplotypes across *DAT1*. (**a**) Predicted phylogeny of the 3′ region (block 3) of *DAT1*, exon 9 – exon 15. 'Ancestral', 'common,' and 'recent' haplotypes are defined in the Materials and methods section. Solid lines indicate single mutation events in a lineage, and dashed lines indicate recombination events in a lineage (there are necessarily two sources for each resulting daughter haplotype). A '?' in an ancestral haplotype represents the precursor to a mutation. This allows us to present the information that the two haplotypes are related, knowing where the mutation occurred but not which haplotype is the parent. The haplotypes represented in the figure are as follows: B = 'GGAGCGAACG9C' (16%), D = 'AGGATTAGGG10T' (5%), K = '?TGGCGGAG?10T'(63%), A = 'ATGGCGGAGG10T' (47%), C = 'ATGGCGGAGA10T' (12%), E = 'GTGGCGGAGA10T' (4%), F = 'ATGGCGAACG9C' (3%), G = 'ATGGCGGAGG9C' (3%), H = 'GTGGCGGAGG10T' (2%), I = 'ATGGCGGAGG10T' (1%), and J = 'ATGGCGAAGG10T' (1%). (**b**) Shown are the SNPs and corresponding haplotypes of the three blocks and the frequencies of the haplotypes within and spanning these blocks. The frequencies of the haplotypes in each block are indicated above the corresponding haplotypes. Solid lines represent haplotypes with >20% frequency and dotted lines represent haplotypes with >5% frequency. The clades to which the haplotypes belong are indicated in brackets as 1 or 2 for the 5′ (block 1) clades and K, B, and D for the 3′ (block 3) clades. The previously analyzed 14 SNPs are shaded. The block regions are indicated as promoter – intron 2, intron 5 – intron 7, (intron 8), and exon 9 – exon 15. (**c**) Details of the haplotypes in the 3′ region with the exon 9 – exon 15 broken up into smaller blocks consisting of exon 9 – intron 10, intron 11 – intron 14, and exon 15.

by the first two SNPs and encompass the majority of haplotypes (data not presented; see Figure 2b for haplotypes). The majority of recent variants are the result of recombination events between these ancestral haplotype groups.

Figure 2b and c illustrate the structure of complete haplotypes across *DAT1*, including the most common ( >5%) connections between the three observed haplotype blocks. Despite the apparently high level of recombination in some regions, there is still observable structure maintained between the haplotype blocks. Many of the common haplotypes in one block are followed by only a single haplotype in the neighboring block. For example, the rarer of the common haplotypes of block 1 ('GTAAGAA' and 'GTAAGAG') are almost always followed by 'GGC' in block 2.

*Sliding window analyses*

Since the TDT only considers heterozygous parents, the use of biallelic markers such as SNPs may result in the loss of a great deal of power to detect LD. The use of multiallelic haplotypes composed of multiple SNPs may help to increase the power of this test by increasing the degree of heterozygosity in the parental genotypes. Although there are several methods of haplotype construction, we used an unbiased sliding window method to construct successive and adjacent haplotypes across *DAT1* in windows of five adjacent SNPs.

Table 2 shows the results of the sliding window association studies. Using the method of Nyholt we determined that the threshold for significance in these analyses was a *P*-value of 0.003.[15] TDT analysis of the sliding window haplotypes in the NIMH sample revealed empirical *P*-values $\leqslant$0.003 for windows 5, 8, 9, and 10 under the genotypewise model. Analysis of the UCSD sample also revealed strong evidence of association for window 9, which spans the region between introns 6 and 9, with an empirical *P*-value $\leqslant$0.003 under the genotypewise model.

Although the results from the sliding window analyses in both samples implicate window 9, the focus of these two samples, considering both the single SNP and haplotype analyses, is on slightly different regions of the gene with the UCSD sample implicating more 3′ regions and NIMH sample implicating more 5′ regions. Furthermore, analysis of the individual haplotype transmissions in these

**Table 2** Sliding window haplotype analysis in the UCSD, NIMH, and Linked samples

| *Window* | *Included SNPs* | *UCSD* | | *NIMH* | | *Linked* | |
|---|---|---|---|---|---|---|---|
| | | *Allele* | *Genotype* | *Allele* | *Genotype* | *Allele* | *Genotype* |
| 1 | P + 215,P + 2459,I1 + 1036,I1 + 1735,I1 + 1859 | 0.187 | 0.227 | 0.381 | 0.408 | **0.039** | **0.045** |
| 2 | P + 2459,I1 + 1036,I1 + 1735,I1 + 1859,I1 + 1860 | 0.261 | 0.513 | 0.236 | 0.267 | **0.047** | 0.060 |
| 3 | I1 + 1036,I1 + 1735,I1 + 1859,I1 + 1860,I2 + 48 | 0.154 | 0.097 | 0.096 | **0.013** | 0.115 | 0.258 |
| 4 | I1 + 1735,I1 + 1859,I1 + 1860,I2 + 48,I5 + 448 | 0.074 | 0.102 | 0.258 | **0.005** | 0.305 | **0.025** |
| 5 | I1 + 1859,I1 + 1860,I2 + 48,I5 + 448,I6 + 96 | 0.109 | 0.244 | **0.048** | **0.001** | 0.120 | **0.001** |
| 6 | I1 + 1860,I2 + 48,I5 + 448,I6 + 96,I7 + 922 | 0.341 | 0.438 | 0.148 | 0.492 | **0.027** | **0.001** |
| 7 | I2 + 48,I5 + 448,I6 + 96,I7 + 922,I8 + 2086 | 0.867 | 0.766 | 0.308 | **0.040** | **0.008** | **0.029** |
| 8 | I5 + 448,I6 + 96,I7 + 922,I8 + 2086,E9 + 59 | 0.732 | 0.286 | 0.149 | **0.0001** | **0.003** | **0.002** |
| 9 | I6 + 96,I7 + 922,I8 + 2086,E9 + 59,I9 + 102 | 0.411 | **0.002** | 0.080 | **0.001** | **0.003** | **<0.00001** |
| 10 | I7 + 922,I8 + 2086,E9 + 59,I9 + 102,I9−21 | 0.277 | **0.040** | 0.057 | **0.003** | **0.0001** | **0.0003** |
| 11 | I8 + 2086,E9 + 59,I9 + 102,I9−21,I10 + 117 | 0.305 | 0.052 | 0.152 | 0.156 | **0.024** | 0.055 |
| 12 | E9 + 59,I9 + 102,I9−21,I10 + 117,I11 + 2078 | **0.046** | **0.020** | 0.631 | 0.574 | **0.003** | **0.003** |
| 13 | I9 + 102,I9−21,I10 + 117,I11 + 2078,I12 + 268 | 0.133 | 0.117 | 0.525 | 0.374 | **0.029** | **0.022** |
| 14 | I9−21,I10 + 117,I11 + 2078,I12 + 268,I13−481 | 0.058 | 0.176 | 0.630 | 0.172 | **0.039** | **0.043** |
| 15 | I10 + 117,I11 + 2078,I12 + 268,I13−481,I14 + 4217 | 0.156 | 0.209 | 0.777 | 0.551 | 0.058 | 0.060 |
| 16 | I11 + 2078,I12 + 268,I13−481,I14 + 4217,E15 + 274 | 0.362 | 0.525 | 0.562 | 0.506 | **0.034** | **0.001** |
| 17 | I12 + 268,I13−481,I14 + 4217,E15 + 274,E15 + 352 | 0.264 | **0.029** | 0.441 | 0.141 | **0.012** | **0.0003** |
| 18 | I13−481,I14 + 4217,E15 + 274,E15 + 352,E15VNTR | 0.182 | **0.016** | 0.134 | 0.165 | **0.011** | **0.004** |
| 19 | I14 + 4217,E15 + 274,E15 + 352,E15VNTR,E15 + 1812 | 0.109 | **0.019** | 0.372 | 0.401 | **0.0001** | **0.0004** |

*Note*: Results presented are empirical *P*-values for the allele-wise (allele) and genotype-wise (genotype) models of the ETDT. All *P* values $<0.05$ are indicated in bold. All *P*-values $\leqslant 0.003$ (shaded) remain significant following a correction for multiple testing.

two samples revealed the 'ATGG' haplotype, corresponding to SNPs from exon 9 through intron 10, to be transmitted 2:1 in the UCSD sample and nontransmitted 2:1 in the NIMH sample. Therefore, a combined analysis of these two samples would likely just dilute the existing associations in each sample due to the allelic heterogeneity observed between them.

We thus attempted to enrich our sample for those families that may segregate a functional polymorphism at the *DAT1* locus contributing to susceptibility to bipolar disorder, and thereby reduce genetic heterogeneity, by stratifying the combined sample by linkage to the *DAT1* region. This resulted in 33 sibpair families (Linked sample) for analysis with a ratio of 4.5 NIMH to 1 UCSD family. Analysis of the sliding windows in the Linked sample revealed empirical *P*-values $\leqslant 0.003$ for windows 5, 6, 8, 10, 12, 16, 17, and 19 under the genotypewise model. The strongest evidence of association was observed for window 9 ($P<0.00001$), the window implicated in the analyses of the UCSD and NIMH samples independently. Similar results were obtained using the allelewise model with empirical *P*-values of $\leqslant 0.003$ observed for windows 8, 9, 10, 12, and 19.

The individual haplotypes that were found to be significant within the windows in the Linked sample are listed in Table 3. It is interesting to note that the 'A' allele of the I8 + 2086 SNP segregates as part of the transmitted haplotypes in the Linked sample, whereas this allele was found to be nontransmitted in the NIMH sample, both as a single SNP and as part of the individual haplotypes (data not presented) segregating in this sample. This indicates that the subset of the NIMH sample that was included in the Linked sample is different than those that contributed to the signal in the NIMH trios as a whole and suggests the potential presence of allelic heterogeneity in this sample.

Haplotypes deriving from ancestral group K ('?TGGCGGAG?10T', see Figure 2), which account for approximately 65% of block 3 haplotypes, were found to be significantly transmitted (windows 12–16) in the Linked sample with a minimum *P*-value of 0.002 observed for window 14, consistent with our previous analyses.[6] The same haplotypes in windows 13–15 were also found to be significantly transmitted in the UCSD sample, with a minimum *P*-value of 0.005 observed for window 14 as well (data not presented). Haplotypes deriving from ancestral group B ('GGAGCGAACG9C'), which account for approximately 16% of block 3 haplotypes, were found to be significantly nontransmitted (windows 12–19), consistent with our previous analyses.[6] The haplotype corresponding to ancestral haplotype D ('AGGAT TAGGG10T'), was also found to be nontransmitted 4:1 in the Linked sample but with only a resultant *P*-value of 0.058, likely due to low sample size (10 transmissions in total). Further sliding window analyses revealed this haplotype to be nontransmitted 5:2 in the UCSD sample set as well with 21 transmissions and a *P*-value of 0.050 (data not presented). The minor alleles of SNPs I11 + 2478, I12 + 267, and I14 + 4217 are unique to this haplotype, as is a I14 + 99 15 bp insertion/deletion polymorphism that we have previously shown to affect

**Table 3** Significantly transmitted (T) and non-transmitted (NT) haplotypes observed in the linked sample set

| *Window* | *Haplotype*[a] | | | | | | | | | | | | | | | | *T* | *NT* | P-*Value* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | A | G | G | **C** | **A** | | | | | | | | | | | | 10 | 3 | **0.052** |
| 8 | G | G | C | **A** | G | | | | | | | | | | | | 11 | 3 | **0.033** |
| 9 | | G | C | **A** | G | T | | | | | | | | | | | 11 | 3 | **0.033** |
| 10 | | | C | **A** | G | T | G | | | | | | | | | | 11 | 3 | **0.033** |
| 11 | | | | **A** | G | T | G | G | | | | | | | | | 13 | 3 | **0.013** |
| 12 | | | | | G | T | G | G | C | | | | | | | | 13 | 3 | **0.013** |
| 13 | | | | | | T | G | G | C | G | | | | | | | 33 | 16 | **0.015** |
| 14 | | | | | | | G | G | C | G | **G** | | | | | | 40 | 17 | **0.002** |
| 15 | | | | | | | | G | C | G | **G** | A | | | | | 38 | 17 | **0.005** |
| 16 | | | | | | | | | C | G | **G** | A | G | | | | 40 | 15 | **0.008** |
| 17 | | | | | | | | | | G | **G** | A | G | G | | | 38 | 23 | 0.055 |
| 10 | | | | **C** | **G** | G | G | A | | | | | | | | | 11 | 26 | **0.014** |
| 11 | | | | | **G** | G | G | A | G | | | | | | | | 13 | 26 | **0.037** |
| 12 | | | | | | G | G | A | G | C | | | | | | | 13 | 26 | **0.037** |
| 13 | | | | | | | G | A | G | C | G | | | | | | 13 | 26 | **0.037** |
| 14 | | | | | | | | A | G | C | G | **A** | | | | | 13 | 26 | **0.037** |
| 15 | | | | | | | | | G | C | G | **A** | A | | | | 15 | 30 | **0.025** |
| 16 | | | | | | | | | | C | G | **A** | A | C | | | 12 | 29 | **0.008** |
| 17 | | | | | | | | | | | G | **A** | A | C | G | | 15 | 30 | **0.025** |
| 19 | | | | | | | | | | | | | A | C | G | 9 | 8 | 26 | **0.002** |
| 14 | | | | | | | | G | A | T | T | **A** | | | | | 2 | 8 | 0.058 |
| 15 | | | | | | | | | A | T | T | **A** | G | | | | 2 | 8 | 0.058 |
| 16 | | | | | | | | | | T | T | **A** | G | G | | | 2 | 8 | 0.058 |
| 17 | | | | | | | | | | | T | **A** | G | G | G | | 2 | 8 | 0.058 |
| | I2 + 48 | I5 + 448 | I6 + 96 | I7 + 922 | I8 + 2086 | E9 + 59 | I9 + 102 | I9−21 | I10 + 117 | I11 + 2478 | I12 + 268 | I13 + 1457 | I14 + 4217 | E15 + 274 | E15 + 352 | E15 VNTR | E15 + 1812 | | | |

[a]The previously analyzed 14 SNPs (Greenwood *et al*., 2001; 2002) and their haplotypes are indicated in the shaded regions. The two SNPs that were individually significant in the UCSD and NIMH samples are indicated in bold.

*DAT1* expression.[19] Therefore, it appears to be a distinct third 3′ clade accounting for approximately 6% of 3′ haplotypes, as we have previously suggested.[19] Together, these results suggest the possible presence of multiple susceptibility mutations in the gene.

## Discussion

These data provide further support for an association of *DAT1* with bipolar disorder in a second independent sample using a denser map of SNPs, as well as a higher level of resolution of the structure of the *DAT1* gene. We observed a similar block structure as that previously described for the *DAT1* gene, although we provide an increased resolution in the intron 2 – exon 9 region, allowing for the discrimination of a previously unidentified haplotype block in this region between introns 5 and 7. These studies show increased recombination in this region compared to the rest of the gene, with the majority of the recombination appearing to have occurred in or near introns 3, 4, and 8. The results of our single SNP analyses suggest that the UCSD and NIMH samples show association to slightly different regions of *DAT1*. The UCSD association appears to center around the I13 + 1457 SNP, consistent with our previous studies implicating the region between exons 9 and 15, whereas the NIMH association appears to center around the I8 + 2086 SNP.[6] However, the sliding window haplotype association studies revealed an overlap between these samples with the evidence for association in each sample maximizing near intron 8 in a region between haplotype blocks.

These analyses provide further evidence that haplotypes of the 'ATGGCGGAGG10T' clade of the 3′ haplotype block segregate with bipolar disorder, while haplotypes of the 'GGAGCGAACG9C' and 'AGGATTAGGG10T' clades do not. The I13 + 1457 SNP serves as a marker to distinguish between these three 3′ clades. The 'G' allele of this SNP, which is associated with haplotypes of the 'ATGGCG GAGG10T' clade, was not only significantly transmitted in both the UCSD and Linked samples, but was also present in the individual 'windows' of the region with haplotypes found to be significantly transmitted these samples. The 'A' allele of this SNP segregates with haplotypes of the two nontransmitted 3′ clades.

Although we have previously reported a difference in expression between clades of the 5′ region and have implicated the P + 2459 SNP as potentially the cause of this functional difference, we did not observe any association of bipolar disorder to this individual SNP, nor to any of the 5′ haplotypes, consistent with our previous association results in the UCSD sample.[6,19] However, the 'T' variant of this SNP was recently observed to be associated with schizophrenia in a sample of Iranian males.[20] It is this variant that was observed to increase *DAT1* gene expression twofold in our functional analyses.[18] Thus, it is possible that increased levels of expression of *DAT1* driven by variation in the core promoter of this gene contribute a predisposition to psychiatric illness. Our previous report of functional differences in transcriptional regulation suggested the presence of multiple functional variants in the gene. The present, more detailed, association analyses may be consistent with this functional analysis in suggesting possible allelic heterogeneity.

A 15 bp insertion in intron 14 (I14 + 99) that has been previously shown to correlate with decreased functional activity of the *DAT1* gene was found to segregate exclusively with the 'AGGATTAGGG10T' haplotype, which forms a third clade in the 3′ region of the gene.[19] Haplotypes of this clade were found to be nontransmitted 4:1 in the Linked families, although this observation was not statistically significant due to low sample size and allele frequency. Haplotypes of this clade that included this insertion were also found to be nontransmitted in the UCSD sample ($P = 0.050$, data not presented). It could thus be argued that a variant resulting in decreased *DAT1* expression may exhibit some sort of protective effect against psychiatric illness.

There has been much focus on the 3′ VNTR within exon 15 of *DAT1* as a possible functional element. Association studies have implicated variants of this VNTR in the causation of several psychiatric disorders, most notably ADHD.[1–3,20,21] Functional analyses of the two most common variants, the 10- and 9-repeat alleles have, however, been somewhat contradictory with different groups reaching different conclusions about which of these two alleles is associated with increased DAT gene expression.[22–27] In our previous studies, we found no evidence that either of the common alleles had any effect on DAT gene expression, nor did we find evidence for an association of this VNTR to bipolar disorder.[6,19] These association analyses in a second independent sample replicate our previous results and suggest that variants of the VNTR may not play a role in bipolar disorder or, at the very least, argue against this being the sole functional variant.

This study is limited primarily by the relatively small sample size. Our ability to resolve possible allelic heterogeneity resulting from functional mutations in rarer haplotypes is thus likely to be limited. The use of an even denser map of SNPs, particularly in the region of association, would not only improve the informativity but may also help us to further resolve the region of association. Therefore, replication in a larger sample with a denser map is necessary to confirm these results.

In summary, we report the replication of our previous observation of association of bipolar disorder in a second independent family set using a denser map of SNPs. These data localize a susceptibility variant to a region near intron 8 and suggest the possibility of multiple functional susceptibility variants. We do not find support for association to the frequently studied 3′ VNTR. This study complements our previous functional study of transcriptional

regulation in suggesting that *DAT1* contains multiple sequence variants, which impact transcriptional regulation and together contribute towards disease susceptibility.

## Acknowledgments

We are grateful to the family members who participated in this study. This work was supported by grants to JRK from NIH (MH47612, MH59567, MH68503, DA13769), the Department of Veterans Affairs and by the UCSD General Clinical Research Center (M01 RR00827). JRK is a founder and holds equity in Psynomics Inc. We would like to thank A Dessa Sadovnick, Ronald Remick, Susan McElroy, and Paul Keck for contributing a portion of the samples analyzed. Data and biomaterials were collected in four projects that participated in the NIMH Bipolar Disorder Genetics Initiative. From 1991–1998, the Principal Investigators and Co-investigators were: Indiana University, Indianapolis, IN, UO1 MH46282, John Nurnberger, MD, PhD, Marvin Miller, MD, and Elizabeth Bowman, MD; Washington University, St Louis, MO, UO1 MH46280, Theodore Reich, MD, Allison Goate, PhD, and John Rice, PhD; Johns Hopkins University, Baltimore, MD UO1 MH46274, J Raymond DePaulo, Jr, MD, Sylvia Simpson, MD, MPH, and Cohn Stine, PhD; NIMH Intramural Research Program, Clinical Neurogenetics Branch, Bethesda, MD, Elliot Gershon, MD, Diane Kazuba, BA, and Elizabeth Maxwell, MSW.

## References

1 Cook Jr EH, Stein MA, Krasowski MD, Cox NJ, Olkon DM, Kieffer JE *et al.* Association of attention-deficit disorder and the dopamine transporter gene. *Am J Hum Genet* 1995; **56**: 993–998.
2 Gill M, Daly G, Heron S, Hawi Z, Fitzgerald M. Confirmation of association between attention deficit hyperactivity disorder and a dopamine transporter polymorphism. *Mol Psychiatry* 1997; **2**: 311–313.
3 Waldman ID, Rowe DC, Abramowitz A, Kozel ST, Mohr JH, Sherman SL *et al.* Association and linkage of the dopamine transporter gene and attention-deficit hyperactivity disorder in children: heterogeneity owing to diagnostic subtype and severity. *Am J Hum Genet* 1998; **63**: 1767–1776.
4 Daly G, Hawi Z, Fitzgerald M, Gill M. Mapping susceptibility loci in attention deficit hyperactivity disorder: preferential transmission of parental alleles at DAT1, DBH and DRD5 to affected children. *Mol Psychiatry* 1999; **4**: 192–196.
5 Kelsoe JR, Sadovnick AD, Kristbjarnarson H, Bergesch P, Mroczkowski-Parker Z, Drennan M *et al.* Possible locus for bipolar disorder near the dopamine transporter on chromosome 5. *Am J Med Genet* 1996; **67**: 533–540.
6 Greenwood TA, Alexander M, Keck PE, McElroy S, Sadovnick AD, Remick RA *et al.* Evidence for linkage disequilibrium between the dopamine transporter and bipolar disorder. *Am J Med Genet* 2001; **105**: 145–151.
7 Kelsoe JR, Spence MA, Loetscher E, Foguet M, Sadovnick AD, Remick RA *et al.* A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22. *Proc Natl Acad Sci USA* 2001; **98**: 585–590.
8 Dick DM, Foroud T, Edenberg HJ, Miller M, Bowman E, Rau NL *et al.* Apparent replication of suggestive linkage on chromosome 16 in the NIMH genetics initiative bipolar pedigrees. *Am J Med Genet* 2002; **114**: 407–412.
9 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
10 Vandenbergh DJ, Thompson MD, Cook EH, Bendahhou E, Nguyen T, Krasowski MD *et al.* Human dopamine transporter gene: coding region conservation among normal, Tourette's disorder, alcohol dependence and attention-deficit hyperactivity disorder populations. *Mol Psychiatry* 2000; **5**: 283–292.
11 Greenwood TA, Alexander M, Keck PE, McElroy S, Sadovnick AD, Remick RA *et al.* Segmental linkage disequilibrium within the dopamine transporter gene. *Mol Psychiatry* 2002; **7**: 165–173.
12 Kokoris M, Dix K, Moynihan K, Mathis J, Erwin B, Grass P *et al.* High-throughput SNP genotyping with the Masscode system. *Mol Diagn* 2000; **5**: 329–340.
13 Sham PC, Curtis D. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann Hum Genet* 1995; **59**: 323–336.
14 Sham PC, Curtis D. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet* 1995; **59**: 97–105.
15 Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004; **74**: 765–769.
16 Lewontin RC. On measures of gametic disequilibrium. *Genetics* 1988; **120**: 849–852.
17 Abecasis GR, Cookson WO. GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* 2000; **16**: 182–183.
18 Halperin E, Eskin E. Haplotype reconstruction from genotype data using Imperfect Phylogeny. *Bioinformatics* 2004; **20**: 1842–1849.
19 Greenwood TA, Kelsoe JR. Promoter and intronic variants affect the transcriptional regulation of the human dopamine transporter gene. *Genomics* 2003; **82**: 511–520.
20 Khodayari N, Garshasbi M, Fadai F, Rahimi A, Hafizi L, Ebrahimi A *et al.* Association of the dopamine transporter gene (DAT1) core promoter polymorphism −67T variant with schizophrenia. *Am J Med Genet* 2004; **129B**: 10–12.
21 Curran S, Mill J, Tahir E, Kent L, Richards S, Gould A *et al.* Association study of a dopamine transporter polymorphism and attention deficit hyperactivity disorder in UK and Turkish samples. *Mol Psychiatry* 2001; **6**: 425–428.
22 Heinz A, Goldman D, Jones DW, Palmour R, Hommer D, Gorey JG *et al.* Genotype influences *in vivo* dopamine transporter availability in human striatum. *Neuropsychopharmacology* 2000; **22**: 133–139.
23 Jacobsen LK, Staley JK, Zoghbi SS, Seibyl JP, Kosten TR, Innis RB *et al.* Prediction of dopamine transporter binding availability by genotype: a preliminary report. *Am J Psychiatry* 2000; **157**: 1700–1703.
24 Michelhaugh SK, Fiskerstrand C, Lovejoy E, Bannon MJ, Quinn JP. The dopamine transporter gene (SLC6A3) variable number of tandem repeats domain enhances transcription in dopamine neurons. *J Neurochem* 2001; **79**: 1033–1038.
25 Fuke S, Suo S, Takahashi N, Koike H, Sasagawa N, Ishiura S. The VNTR polymorphism of the human dopamine transporter (DAT1) gene affects gene expression. *Pharmacogenomics J* 2001; **1**: 152–156.
26 Mill J, Asherson P, Browes C, D'Souza U, Craig I. Expression of the dopamine transporter gene is regulated by the 3′ UTR VNTR: Evidence from brain and lymphocytes using quantitative RT-PCR. *Am J Med Genet (Neuropsychiatric Genetics)* 2002; **114**: 975–979.
27 Miller GM, Madras BK. Polymorphisms in the 3′-untranslated region of human and monkey dopamine transporter genes affect reporter gene expression. *Mol Psychiatry* 2002; **7**: 44–55.