Molecular Psychiatry (2002) 7, 165–173
© 2002 Nature Publishing Group All rights reserved 1359-4184/02 $25.00
www.nature.com/mp

ORIGINAL RESEARCH ARTICLE

Segmental linkage disequilibrium within the dopamine transporter gene

TA Greenwood[1], M Alexander[1], PE Keck[2], S McElroy[2], AD Sadovnick[3], RA Remick[4], SH Shaw[1] and JR Kelsoe[1]

[1]*Depts of Psychiatry, University of California, San Diego and San Diego VA Health Care System, San Diego, CA, USA;* [2]*Dept of Psychiatry, University of Cincinnati, Cincinnati, OH, USA;* [3]*Dept of Medical Genetics, University of British Columbia, Vancouver, Canada;* [4]*Dept of Psychiatry, St Paul's Hospital, Vancouver, Canada*

The dopamine transporter gene (DAT) has been implicated in a variety of disorders, including bipolar disorder, attention-deficit hyperactivity disorder, cocaine-induced paranoia, Tourette's syndrome, and Parkinson's disease. As no clear functional polymorphism has been identified to date, studies rely on linkage disequilibrium (LD) to assess the possible genetic contribution of DAT to the various disorders. A better understanding of the complex structure of LD across the gene is thus critical for an accurate interpretation of the results of such studies, and may facilitate the mapping of the actual functional variants. In the process of characterizing the extent of variation within the DAT gene, we have identified a number of single nucleotide polymorphisms (SNPs) suitable for LD studies, 14 of which have been analyzed, along with a 3′ repeat polymorphism, in a sample of 120 parent-proband triads. Calculations of pairwise LD between the SNPs in the parental haplotypes revealed a high degree of LD ($P < 0.00001$) in the 5′ (distal promoter through intron 6) and 3′ (exon 9 through exon 15) regions of DAT. This segmental LD pattern is maintained over approximately 27 kb and 20 kb in these two regions, respectively, with very little significant LD between them, possibly due to the presence of a recombination hotspot located near the middle of the gene. These analyses of the DAT gene thus reveal a complex structure resulting from both recombination and mutation, knowledge of which may be invaluable to the design of future studies.
Molecular Psychiatry (2002) **7,** 165–173. DOI: 10.1038/sj/mp/4000958

Keywords: genetic variation; single nucleotide polymorphism; linkage disequilibrium; intragenic recombination

Introduction

Disturbances in dopaminergic systems have been implicated in the etiology of several neuropsychiatric disorders, including bipolar disorder, attention-deficit hyperactivity disorder (ADHD), schizophrenia, and Tourette's syndrome. The dopamine transporter (DAT) mediates the active reuptake of dopamine, a key neurotransmitter in the regulation of mood and movement, from the synapse into the presynaptic terminal.[1] DAT is the principal regulator of synaptic dopamine concentration, as well as the duration of dopamine activity, and thus plays a critical role in the regulation of dopaminergic transmission. As the site of action of cocaine and amphetamine, which inhibit dopamine reuptake, DAT may be an important element in psychostimulant reward, leading to abuse of such substances.[2] DAT is also involved in the accumulation of certain neurotoxins into dopaminergic neurons leading to Parkinson's disease.[3] The DAT gene has been mapped to chromosome 5p15.3,[4] and its entire sequence, spanning nearly 60 kb, has recently been elucidated.[5]

DAT has been pursued as a candidate gene in numerous disorders, and DAT sequence variants have been analyzed in attempts to determine its possible genetic contribution to these disorders. Studies of a 40-bp variable number of tandem repeats (VNTR) in the 3′ untranslated region have revealed an allelic or genotypic association with bipolar disorder,[6] ADHD,[7,8] schizophrenia,[9] cocaine-induced paranoia,[10] alcoholism,[11] the severity of alcohol withdrawal,[12,13] and Parkinson's disease.[14,15] A recent study has shown an association between the 9-repeat allele of this VNTR and reduced DAT protein availability *in vivo* as measured by SPECT imaging.[16] Although these studies have employed the use of the VNTR polymorphism within exon 15 of DAT, there are no data to indicate an actual effect of the VNTR on gene function. Thus, it is likely that this polymorphism is not the functional variant itself, rather that it is in linkage disequilibrium with the actual pathological variant that lies elsewhere within the gene.

Linkage disequilibrium (LD) occurs in populations

Correspondence: JR Kelsoe, Dept of Psychiatry, 0603, UCSD, La Jolla, CA 92093, USA. E-mail: jkelsoe@ucsd.edu
Received 8 December 2000; revised 30 April 2001; accepted 22 May 2001

as a consequence of mutation, selection of single or linked alleles, random genetic drift, and population admixture, and it decays at a rate proportional to the recombination fraction between the two loci in LD and the number of generations since the establishment of LD.[17] The patterns of variation thus created across the genome, and even within a particular gene, can be used to map genes contributing to complex genetic disorders. Studies of LD using single nucleotide polymorphisms (SNPs) may provide a powerful alternative to linkage analysis using microsatellite markers in determining the significance of a candidate gene in a particular disorder, since SNPs are more abundant and much less prone to mutation than microsatellites. Multi-allelic haplotypes comprised of SNPs are more informative than individual SNPs, comparable to microsatellite markers, and may prove to be extremely useful in determining the significance of a candidate gene in a particular disorder, as well as helping to narrow down the region of interest at a disease locus. However, the utility of such haplotypes may only be maximized when the LD relationships between the SNPs are taken into account.

Several attempts have been made to characterize the variation within the DAT gene and to use the resultant SNPs to assess association with disease.[5,18,19] We have previously reported the analysis of 14 SNPs that span the gene from the distal promoter through the 3′ UTR in a study of LD between DAT and bipolar disorder, which implicates the 3′ end of DAT as the region of interest.[19] This study illustrates the inherent strengths and pitfalls of such LD studies using SNPs and SNP-based haplotypes and highlights the utility of characterizing the LD relationships between SNPs prior to assessing association with disease. We now report the analysis of these same 14 SNPs and the 3′ VNTR to assess their LD relationships and to characterize the pattern of LD across the gene. We have also included in these analyses seven microsatellite markers in the vicinity of DAT to further extend the regional pattern of LD. Taken together, these data define a region within the 3′ end of DAT possibly containing a regulatory variant that may play a role in multiple disorders. Knowledge of this LD structure across the gene may be essential for the successful identification of this variant.

Materials and methods

Subjects

A sample of 50 parent-offspring triads was chosen from the UCSD/UBC/UC family collection and Old Order Amish pedigree 110. Each triad was a subset of a more extended family that was ascertained as part of a multisite linkage study of bipolar disorder, which included collaborative sites at the University of California at San Diego, the University of British Columbia, and the University of Cincinnati. All subjects provided informed consent per local IRB approved procedures prior to participation. Blood was obtained from all subjects for the immortalization of lymphoblastoid cell lines, and DNA was prepared by phenol/chloroform extraction from cultured cells. An additional 70 triads were chosen from the NIMH Genetics Initiative for Bipolar Disorder collection. One triad was selected from each family for a total of 120 triads deriving from 120 separate families. All subjects for this study were Caucasians of northern European descent.

Genotyping and haplotype determination

Genotyping of the triads for these SNPs was accomplished by allele-specific PCR (AS-PCR), in which primers were designed to specifically amplify the reference allele or its variant in separate PCR reactions. Whenever possible, SNPs were combined in double-ended AS-PCR reactions to generate haplotyped markers, as was the case for the following SNPs: I1+1036/I1+1735, I1+1859/I1+1860/I2+28, and E9+59/I9+102. The triads were also genotyped for the 40-bp VNTR in the 3′ untranslated region of exon 15.[20] The primer sequences, methods and specific reaction conditions for SNP and VNTR genotyping have been previously reported.[19]

Seven microsatellite markers in the vicinity of the DAT gene were chosen for LD analyses: D5S2005, D5S678, D5S2488, D5S392, D5S417, D5S1980, and D5S2849. D5S678 has been localized to approximately 25 kb 3′ of DAT.[5] NCBI map viewer coordinates (www.ncbi.nlm.nih.gov) indicate that these markers span a 2.5-Mb region of chromosome 5pter in the following order, telomeric to centromeric: D5S2005 – D5S678 – DAT 3′ → 5′ – D5S2488 – D5S392 – D5S417 – D5S1980 – D5S2849. The PCR reactions for the genotyping of these markers were assembled as follows: 50 ng of genomic DNA, 0.5 μM of each primer, 250 μM dNTPs, 33.5 mM Tris pH 8, 25 mM KCl, 2.5 mM $MgCl_2$, 8.3 mM $(NH_4)_2SO_4$, 0.85 mg ml^{-1} BSA, and 0.05 or 0.1 units of AmpliTaq Gold (Perkin-Elmer Applied Biosystems, Foster City, CA, USA) in a total volume of 20 μl. Forward primers were labeled with one of three fluors. Thermocycling was performed using a PTC-200 thermocycler from MJ Research (Waltham, MA, USA) with one of two cycling conditions. A modified touchdown protocol was used for D5S2488, D5S392, D5S678, and D5S417 with an initial 10 min denaturation at 95°C followed by 20 cycles of 1 min denaturation at 94°C , 1 min primer annealing (temperature starting at 65°C and decreased by 1°C every two cycles), and 1 min primer extension at 72°C with a final extension of 30 min at 72°C.[21] The remaining three markers were amplified using the following conditions: initial denaturation at 96°C for 10 min followed by 35 cycles of 1 min denaturation at 96°C, 1 min primer annealing at 55°C, and 1 min primer extension at 72°C with a final extension at 72°C for 30 min. PCR products were separated by electrophoresis and detected using an ABI 377 and Genescan software 3.0 (Perkin Elmer Applied Biosystems). Multiple markers with different fluors in different molecular weight ranges were pooled, along with a molecular weight standard, for multiplex detection of two to four markers per lane. All genotypes were read in a machine-

assisted fashion using Genotyper software 2.0 (Perkin Elmer Applied Biosystems). Standard DNA samples were used for consistency of correct allele assignment. An average of ten alleles was observed for these markers, of which, on average, four had a parental frequency greater than 10%.

Following genotyping, complete haplotypes comprised of all 14 SNPs and the VNTR were constructed for each individual using linkage phase data obtained from the AS-PCR and parent-offspring relationships. These haplotypes could be conclusively determined in most cases. In ten of the families, however, phase could not be unequivocally inferred for one SNP (I6+96). Therefore, haplotypes were constructed for the remaining 13 SNPs and the VNTR, and the I6+96 SNP was excluded from analyses in these cases. Haplotypes were also determined for the microsatellite markers using parent-offspring relationships. On average, linkage phase could not be resolved for one family per marker. All of the SNPs and markers were found to be in Hardy–Weinberg equilibrium in the population studied.

Linkage disequilibrium analyses

Parental gametic haplotypes were used for calculations of LD for the 14 SNPs, the 3′ VNTR, and the seven microsatellites. In order to simplify calculations, only the 9 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′ was chosen for calculations of LD for its independence of allele frequency.[22] The widely used coefficient of disequilibrium, D, is the difference between the observed haplotype frequency and the frequency expected under statistical independence. The D′ measure is a proportion of the maximum value of D, whose range extends from −1 to +1, with −1 and +1 representing complete LD and 0 representing free association. D′ and the significance of LD were calculated using the GOLD software package.[23]

Results

Degree of genotypic variation

We have previously reported our efforts to detect, quantify, and establish the pattern of sequence variation within the DAT gene.[19] This entailed a screen of the entire coding sequence and the immediately adjacent non-coding sequence, including both 5′ and 3′ UTRs, 9.5 kb of flanking intronic sequence, and 2.5 kb of promoter sequence. A reference sequence for comparison purposes was compiled from published sequence and supplemented with sequence generated from a BAC clone of DAT (BACH-278o18, Genome Systems, St Louis, MO, USA). This mutation screen led to the identification of 92 deviations from the reference sequence, which included 25 insertion-deletion polymorphisms and 67 base substitutions, 39 of which were polymorphic among the samples sequenced. The one coding sequence variant identified, an A to G substitution in exon 9, was found to be synonymous, as previously reported.[5]

Of the 39 SNPs identified by the mutation screen, those for further analysis were chosen on the basis of their locations within the gene, the availability of at least one known heterozygote, and the feasibility of unequivocal allele detection by allele-specific PCR (AS-PCR). The 14 SNPs thus chosen for an examination of LD span the gene from the distal promoter through exon 15. The relative positions of these SNPs within the DAT gene are shown in Figure 1. Their specific positions, as well as their frequencies and observed heterozygosities among the 240 parental genotypes, are indicated in Table 1. All of the SNP variants were quite frequent within the population studied, with only the I10+117 SNP exhibiting a minor allele frequency less than 15%.

Haplotypic variation and intragenic recombination

Linkage phase data from AS-PCR and parent-offspring relationships were used to construct haplotypes comprised of all 14 SNPs and the VNTR. Figure 2 illustrates the 72 distinct DAT haplotypes observed for the 14 SNPs in 480 chromosomes. The haplotypes are arranged in order of sequence similarity as they appeared in a tree generated by PAUP using the neighbor-joining method, which clusters together the haplotypes that are most similar to each other.[24] The frequencies of the 21 haplotypes that were observed in at least 1% of the chromosomes studied are indicated.

Several features of these haplotypes are evident from this plot. A cluster of sites in the 5′ region of the gene (promoter through intron 2) partitions the haplotypes into two major clades. Another set of sites in the 3′ region of the gene (exon 9 through exon 15) further partitions the haplotypes into minor clades within the two major clades. The presence of both minor clades within each of the two major clades reflects a history of intragenic recombination. There also appears to be a greater degree of diversity in the 3′ region as compared to the 5′ region, reflecting either more abundant mutation or a greater degree of intragenic recombination in this region.

Of the many haplotypes that could possibly be observed with 14 segregating sites, only a fraction would actually be observed in any given population. In the absence of intragenic recombination or repeated mutation, the maximum number of haplotypes expected for *s* sites is *s+1*. The observation of more than *s+1* haplotypes for a particular region is an indication of recombination in the history of that section of the gene. Thus, one way to demonstrate intragenic recombination is through the use of a graph with a sliding window to show five consecutive sites at a time and the number of observed haplotypes for each set of sites (Figure 3). In this case, more than six haplotypes were observed for all sets of five sites, indicating intragenic recombination has occurred throughout the DAT gene. However, fewer haplotypes were observed for windows 1 through 3, corresponding to the 5′ region of DAT (promoter through intron 6), as compared to the rest of the gene, indicating a relatively lower level of recombination within this region. The 3′ region of

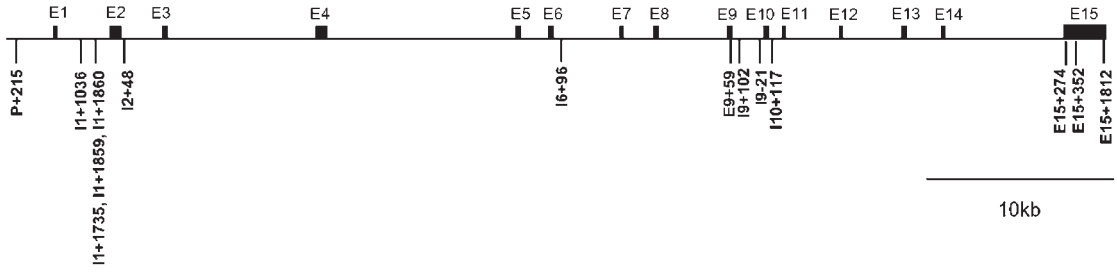

Figure 1 Schematic diagram of the DAT gene indicating the relative positions of the SNPs.

Table 1 Selection of polymorphic sites observed in DAT

Site	*Location*	*Variant*	*Heterozygosity*[a]	*Relative frequency*[b]
P+215	2315 bp 5′ of exon 1	A → G	0.510	0.566
I1+1036	1036 bp 3′ of exon 1	C → A	0.485	0.583
I1+1735	1735 bp 3′ of exon 1	G → A	0.502	0.573
I1+1859	1859 bp 3′ of exon 1	A → G	0.502	0.571
I1+1860	1860 bp 3′ of exon 1	G → A	0.502	0.571
I2+48	48 bp 3′ of exon 2	G → A	0.361	0.751
I6+96	96 bp 3′ of exon 6	G → C	0.552	0.602
E9+59	59 bp within exon 9	A → G	0.386	0.751
I9+102	102 bp 3′ of exon 9	T → G	0.386	0.753
I9–21	21 bp 5′ of exon 10	G → A	0.311	0.820
I10+117	117 bp 3′ of exon 10	G→ A	0.104	0.938
E15+274	274 bp within exon 15	G → C	0.332	0.790
E15+352	352 bp within exon 15	G → A	0.261	0.815
E15 VNTR	exon 15	10R → 9R[c]	0.352	0.761
E15+1812	1812 bp within exon 15	T → C	0.332	0.770

[a]Observed degree of heterozygosity among parental genotypes.
[b]Relative frequency among parental genotypes of the more common allele.
[c]Only the two most common alleles of 9 and 10 repeats were included in analyses.

the gene (exon 9 through exon 15) exhibits a greater degree of diversity and thus a greater number of haplotypes for windows 8 through 10, which is indicative of a higher recombination rate in this region.

The dramatic increase in haplotypes observed for windows 4 through 7, which are comprised of both 5′ and 3′ SNPs, is suggestive of a recombination hotspot somewhere in the middle of the gene. The observed increase in haplotypes for windows 4 through 6, each of which span approximately 30 kb, may also reflect the relatively large regions encompassed by these windows compared to those comprised of SNPs deriving from the same region of the gene. However, this explanation does not account for window 7, which spans only 10 kb and exhibits 15 haplotypes for the five SNPs from intron 6 through intron 10. Window 3, on the other hand, spans a region twice as large as that of window 7, and exhibits slightly more than half the number of haplotypes for the five SNPs from intron 1 through intron 6. This observation that the I6+96 SNP in combination with the 3′ SNPs results in substantially more haplotypes than it does in combination with the 5′ SNPs, despite the reduced area encompassed, is consistent with a recombination hotspot 3′ of this SNP.

Another way to infer that at least one recombination event took place between two sites in the history of the sample is to use the 'four-gamete' test, a two-site analog to Figure 3.[25] According to the infinite-sites model, the mutation rate for any site is infinitesimal, so at most one mutation event can occur at that site in the history of the sample. Therefore, for a given pair of sites, there can be at most four gametic types in the population. Each site may mutate once to generate three gametic types, but since the model does not allow for back or recurrent mutation, the fourth gametic type can only be present in the population if at least one recombination event has occurred between the two sites. Figure 4 shows the site pairs of SNPs for which all four gametes were present in the sample of 480 chromosomes. The distribution of such site pairs indicates that intragenic recombination is likely to have occurred throughout the DAT gene. In this sample, only nine of the 91 site pairs were found to be in complete LD with fewer than four gametes present, five of which sites involved the I10+117 SNP, which exhibited a minor allele frequency of approximately 6%.

Linkage disequilibrium

Despite the apparently abundant intragenic recombination, a great deal of linkage disequilibrium exists

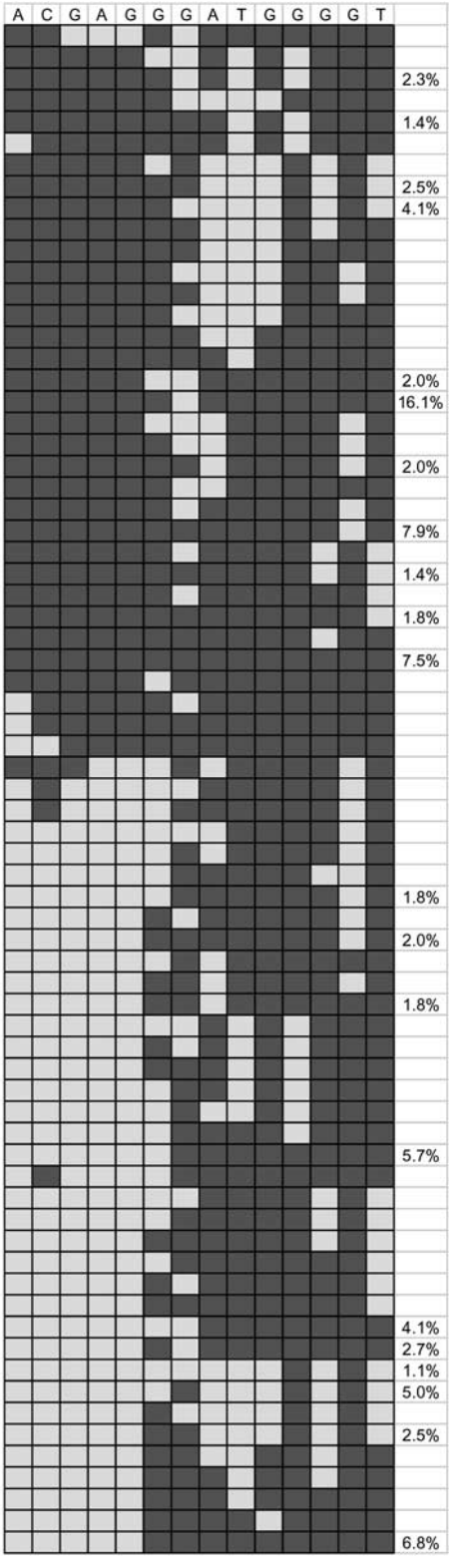

Figure 2 Illustration of the 72 observed DAT haplotypes. For each site, a dark square represents the common nucleotide indicated at the top of the figure and a light square represents the alternative nucleotide. Haplotypes are arranged in order of the tips of a neighboring-joining tree such that more-similar haplotypes are clustered. Frequencies of those haplotypes observed in at least 1% of chromosomes are indicated on the right.

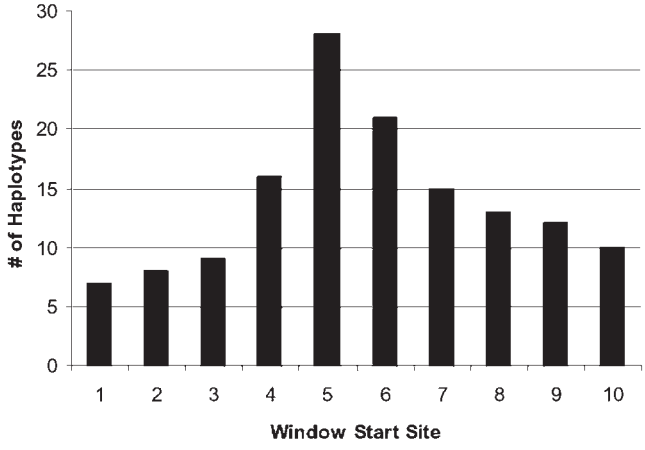

Figure 3 Number of observed distinct haplotypes in a sliding window of five sites. For *s* sites, a maximum of *s+1* haplotypes is expected in the absence of intragenic recombination or repeat mutation. The observation of more than *s+1* haplotypes indicates recombination in that region of the gene.

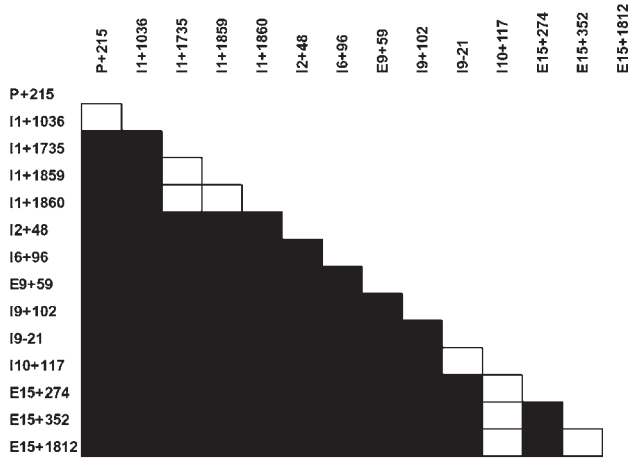

Figure 4 Plot of the four-gamete test with a dark square indicating that all four possible gametic phases were present for a given site pair.

among the site pairs within the DAT gene. Figure 5 shows the results of the pairwise tests in a plot of D′ *vs* distance for pairwise comparisons between all sites, indicating those site pairs that exhibited significant disequilibrium. These analyses revealed 23 site pairs within DAT to be in complete disequilibrium with D′ values greater than 0.95. In general, D′ values greater than 0.3 were observed between the seven SNPs in the 5′ region of DAT (promoter through intron 6), as well as between the seven SNPs and VNTR in the 3′ region (exon 9 through exon 15). With one exception, D′ values over 0.3 were not observed between site pairs comprised of SNPs from the two different regions.

Calculations of the significance of LD revealed a similar pattern with a high level of significance ($P < 0.00001$) observed for nearly all site pairs with D′ values greater than 0.3 (data not presented). The only exceptions were the same five site pairs involving the

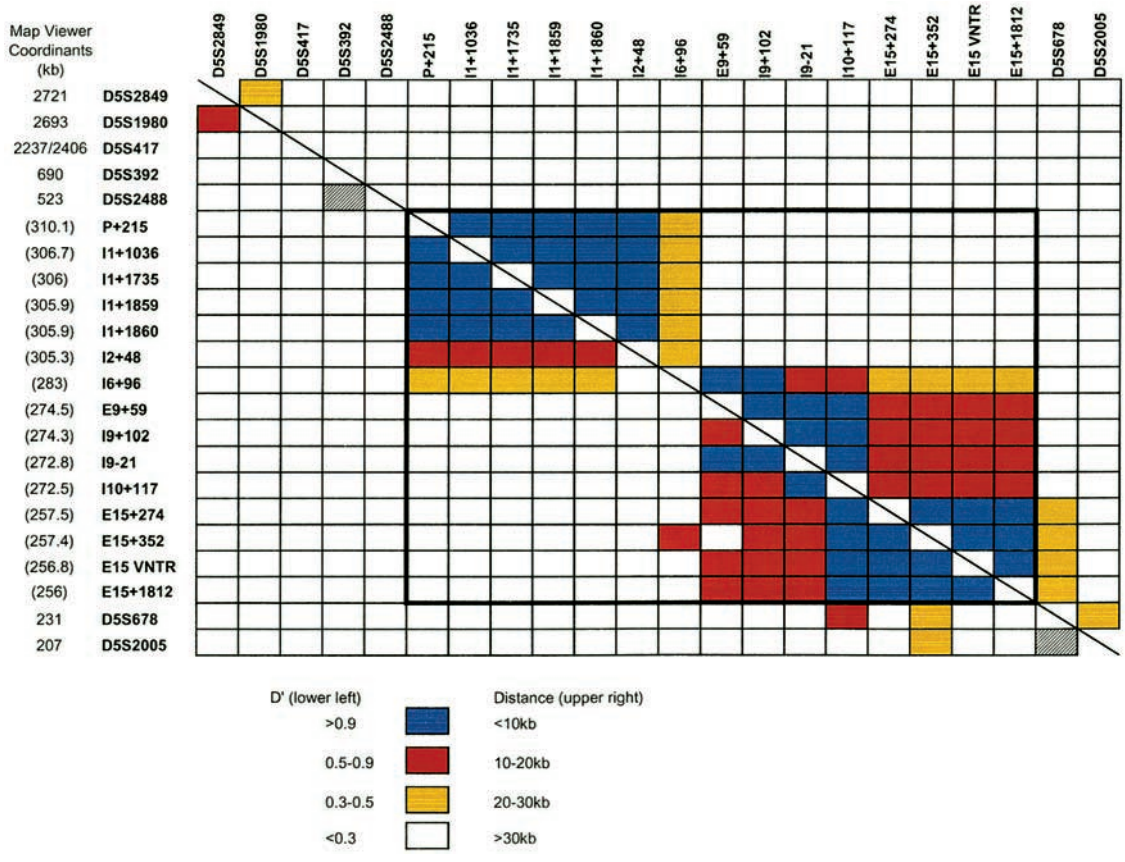

Figure 5 Results of pairwise disequilibrium tests between all site pairs, with D′ values indicated below the diagonal and distances above the diagonal. Cross-hatched squares represent the two site pairs that were found to be highly significant ($P <$ 0.00001) with D′ values below 0.3. Map viewer coordinants for all microsatellite markers, arranged centromeric to telomeric, are indicated, as well as the estimated positions of the DAT markers.

I10+117 SNP that were previously shown to be in complete LD by the four-gamete test. These site pairs show a reduced degree of LD as compared to the surrounding sites, with *P* values less than 0.001, which likely reflects the lack of power of this SNP to test for LD.

These pairwise LD analyses thus revealed a high level of disequilibrium between SNPs in the 5′ region of DAT, as well as between SNPs in the 3′ region, with very little significant LD maintained between these two regions. The segmental pattern of LD thus created extends over approximately 27 kb in the 5′ region and 20 kb in the 3′ region. The partitioning of the gene, based on these LD data, seems to be between intron 6 and exon 9, which is consistent with the data presented in Figure 3.

Within the DAT gene, pairwise D′ values over 0.5 are observed primarily at distances less than 20 kb. Although significant LD is maintained at distances over 20 kb in some cases, mainly between SNPs deriving from the same region of the gene, it is never observed at distances greater than 30 kb. Thus, LD within DAT appears to drop off rapidly at distances greater than 20 kb, which may reflect the fact that, in many cases, SNPs greater than 20 kb apart derive from the two distinct regions of the gene. It is interesting to note, however, that D′ values greater than 0.5 are observed for nearly all site pairs separated by less than 20 kb, except site pairs involving the I6 SNP and SNPs from the exon 9/intron 10 region, for which significant LD is not observed. This observation is consistent with a recombination hotspot in the middle of the gene.

Pairwise LD analyses between the markers within DAT and the seven microsatellites revealed D′ values greater than 0.3 for three site pairs, D5S678/I10+117, D5S678/E15+352, and D5S2005/E15+352, for which *P* values less than 0.001 were observed. Calculations of the significance of LD also revealed a low level of significance ($P < 0.01$) between D5S678 and the seven SNPs in the 3′ region of DAT, and *P* values less than 0.05 were observed for comparisons between D5S678 and seven SNPs in the 5′ region of DAT, as well as between D5S2005 and two additional SNPs in exon 15 (data not presented). Pairwise LD analyses between the microsatellite markers revealed significance only for those markers that map to the same position on the Marshfield genetic map, D5S2849/D5S1980, D5SD5S2488/D5S392, and D5S678/D5S2005, for which *P* values less than 0.00001 were observed (data not presented). Of these three pairs, only D5S2489/D5S1980 gave a relatively high D′ value of 0.65, with the other two site pairs exhibiting D′ values of approximately 0.25.

Discussion

These data shed light on several aspects of the structure and ancestral history of the DAT gene. Analysis of the observed DAT haplotypes reveals a complex structure resulting from both mutation and recombination. A prominent feature is the abundant intragenic recombination that appears to have occurred throughout the gene. The degree of recombination may be increased in DAT relative to other genes due to its telomeric position on chromosome 5. In the absence of recombination or recurrent mutation, gene trees can be drawn to describe the sequence similarities of a series of haplotypes and to delineate the cladistic relationships that reflect the history of the mutations that are likely to have occurred. However, intragenic recombination results in a swapping of gene segments between alleles, which introduces loops into the tree such that it no longer maintains a branched cladistic structure. Thus, the high degree of recombination observed within the DAT gene precludes any efforts at meaningful gene tree construction.

Despite the apparently abundant recombination, a high degree of linkage disequilibrium persists within the gene. This preservation of LD is best exemplified by the 5′ region of the gene, for which the raw haplotype data reveal the presence of two distinct clades of haplotypes with little recombination between them. Since the majority of SNPs in this region derive from the known regulatory regions of the promoter and intron 1, it is interesting to speculate whether a functional difference exists between these two clades that, through the action of selection, has served to maintain them in the population.[26] Substantial LD is also present within the 3′ region of DAT, however, more recombination has occurred in this region to generate a greater degree of diversity. Even in the presence of greater diversity, two clades still exist in this region, once again raising the possibility of a functional correlation and the action of selection in preserving these clades in the population. It is noteworthy that one of these clades segregates with the 10-repeat allele of the 3′ VNTR that has been implicated most notably in ADHD and bipolar disorder, while the other clade segregates with the 9-repeat allele, which has been associated with reduced DAT protein availability, cocaine-induced paranoia, and increased severity of alcohol withdrawal. This further suggests the presence of a functional element in the 3′ region of the gene, alternative variants of which segregate with each of the clades and may contribute to different disorders.

Another striking feature of DAT is the relative lack of LD between SNPs in the 5′ region and SNPs in the 3′ region, for which there are several possible explanations. This observed segmental distribution of LD may reflect selection for two functionally distinct regions in the gene. Alternatively, this pattern of LD may have been generated by a recombination hotspot located in the middle of the gene, at which numerous recombinations in the history of DAT have served to effectively unlink the two end regions. As VNTRs have been speculated to be hotspots for recombination, such recombination events may have occurred within intron 8 where a highly polymorphic, large-allele VNTR has been tentatively mapped.[27]

The suggestion of a midgene recombination hotspot is supported by several lines of data indicating increased recombination between SNPs deriving from intron 6 and exon 9. Although the region spanned by these two SNPs is relatively small, less than 10 kb, significant LD is not maintained. This is in contrast to pairwise comparisons between the I6+96 SNP and those deriving from the promoter and intron 1, which exhibit a high degree of LD ($P < 0.00001$) over distances of 23–27 kb. These observations, however, are highly dependent on one SNP in intron 6, which, despite its relatively high frequency in the population, may not be entirely representative of other SNPs in the region.

A final possible explanation for this segmental LD pattern, that LD simply decreases with increasing distance across the DAT gene, likely oversimplifies the actual situation. Although LD is observed to drop off rapidly at distances greater than 20 kb, these distances are mainly observed between SNPs deriving from the two distinct regions of the gene. However, these results, as with the results of any LD study, are dependent upon the markers chosen for analysis. One factor that may have contributed to these results is the somewhat uneven distribution of the SNPs in this study arising from several factors, the foremost being the availability of SNPs, which were found clustered in certain regions of the gene. SNPs for analysis were chosen based on certain criteria that were met by only 14 of the 39 available SNPs. Thus, the identification and analysis of additional SNPs, especially in the middle region of the gene, may be necessary to more finely resolve the LD pattern across the gene, as well as to delineate its cause.

Although these data clearly indicate the presence of considerable population structure within the DAT gene, these results must be qualified, since the parents of affected offspring were used for the analyses of LD. However, similar results would likely be obtained from parent-offspring triads that did not segregate bipolar disorder. The parents are potential carriers of an allele associated with bipolar disorder, but the nontransmitted alleles should be representative of the general population. In addition, if a mutation that predisposes to bipolar disorder is present in the DAT gene, it will most likely be associated with a subset of one or more common haplotypes that are present in the population and shared by many normal, as well as affected, individuals. Since most of the SNPs have allele frequencies of greater than 15% in the sample surveyed, they are likely to be very ancient and thus represent common alleles in the population. Therefore, the use of parents of affected offspring to evaluate the degree of linkage disequilibrium within DAT should give a valid description of the structure of the gene.

The complex pattern of variation observed for the DAT gene suggests that association studies using mark-

ers chosen at random from polymorphic sites even within the gene may not be reliable for the detection of nearby causal variation. These data highlight the necessity of characterizing the gene of interest in a representative group of individuals prior to the beginning of an association study, and choosing a subset of sites that are representative of the disequilibrium structure of variation in the gene. The presence of two distinct regions within DAT suggests that studies of this gene would best be conducted using haplotypes comprised of SNPs deriving from each region of the gene separately, as haplotypes spanning these regions may be uninformative. This haplotype-based LD strategy has allowed us to restrict the region of association to bipolar disorder to the 3′ end of DAT.[19]

The use of high density SNP maps and haplotypes in LD studies has been the focus of much recent attention, and these data illustrate an important point regarding these methods. SNPs may be superior markers for the detection of LD because of their lower mutation rate and greater abundance as compared to microsatellites. However, the magnitude of LD may vary greatly from marker to marker in the vicinity of a disease locus according to a range of factors, including the proximity of the marker to the disease locus, early recombination events, the distribution of marker allele frequencies in the founder population, selection, and random genetic drift. Thus, LD relationships between SNPs may be quite complex, and their utility for disease mapping may only be fully realized when this is taken into account.

A comparison of the recently elucidated LD structures of the LPL and ATM genes supports the notion that LD varies greatly across the genome. The LPL gene exhibits an even higher level of intragenic recombination than DAT and a subsequently complex structure over a 9.7-kb region.[28] This is in stark contrast to the ATM locus, which displays a remarkable lack of recombination with a pattern of LD that extends over 142 kb.[29] Although the use of such candidate genes in association studies may be an important strategy in the detection of genes contributing only a portion of the overall susceptibility to the development of complex diseases, knowledge of the LD structure and intermarker relationships is essential to the design and efficacy of future studies of these genes.

Acknowledgements

We would like to thank the family members who participated in this study, without whom this work would not be possible. This work was supported by Novartis Pharma AG and grants to JRK from the Department of Veterans Affairs and the NIMH (MH47612, MH59567). Support was also provided by the UCSD Mental Health Clinical Research Center (MH30914) and the UCSD General Clinical Research Center (M01 RR00827). TAG was supported by training grants to the UCSD Biomedical Sciences and Genetics graduate programs (5T32CA67754, 1T32GM08666). Data and biomaterials were collected in four projects that participated in the National Institute of Mental Health (NIMH) Bipolar Disorder Genetics Initiative. From 1991–98, 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 Giros B, Caron MG. Molecular characterization of the dopamine transporter. *Trends in Pharmacol Sci* 1993; **14**: 43–49.
2 Ritz MC, Lamb RJ, Goldberg SR, Kukor MJ. Cocaine receptors on dopamine transporters are related to self-administration of cocaine. *Science* 1987; **237**: 1219–1223.
3 Kitayama S, Wang J-B, Uhl GR. Dopamine transporter mutants selectively enhance MPP^+ transport. *Synapse* 1993; **15**: 58–62.
4 Giros B, el Mestikawy S, Godinot N, Zheng K, Han H, Yang-Feng T *et al*. Cloning, pharmacological characterization, and chromosome assignment of the human dopamine transporter. *Mol Pharmacol* 1992; **42**: 383–390.
5 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.
6 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 (Neuropsychiat Genet)* 1996; **67**: 533–540.
7 Cook 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.
8 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.
9 Persico AM, Macciardi F. Genotypic association between dopamine transporter gene polymorphisms and schizophrenia. *Am J Med Genet (Neuropsychiatr Genet)* 1997; **74**: 53–57.
10 Gelernter J, Kranzler HR, Satel SL, Rao PA. Genetic association between dopamine transporter protein alleles and cocaine-induced paranoia. *Neuropsychopharmacology* 1994; **11**: 195–200.
11 Muramatsu T, Higuchi S. Dopamine transporter gene polymorphism and alcoholism. *Biochem Biophys Res Commun* 1995; **211**: 28–32.
12 Sander T, Harms H, Podschus J, Finckh U, Nickel B, Rolfs A *et al*. Allelic association of dopamine transporter gene polymorphism in alcohol dependence with withdrawal seizures or delirium. *Biol Psychiatry* 1997; **41**: 299–304.
13 Schmidt LG, Harms H, Kuhn S, Rommelspacher H, Sander T. Modification of alcohol withdrawal by the A9 allele of the dopamine transporter gene. *Am J Psychiatry* 1998; **155**: 474–478.
14 Higuchi S, Muramatsu T, Arai H, Hayashida M, Sasaki H, Trojanowski JQ. Polymorphisms of dopamine receptor and transporter genes and Parkinson's disease. *J Neural Transm* 1995; **10**: 107–113.
15 Le Couteur DG, Leighton PW, McCann SJ, Pond SM. Association of a polymorphism in the dopamine transporter gene with Parkinson's disease. *Mov Disord* 1997; **12**: 760–763.
16 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.

17 Hartl DL, Clark AG. *Principles of Population Genetics.* Sinauer Associates: Massachusetts, 1990.
18 Grunehage F, Schulze TG, Muller DJ, Lanczik M, Franzek E, Albus M *et al*. Systematic screening for DNA sequence variation in the coding region of the human dopamine transporter gene (DAT1). *Mol Psychiatry* 2000; **5**: 275–282.
19 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 (Neuropsychiatr Genet)* 2001; **105**: 145–151.
20 Vandenbergh DJ, Persico AM, Uhl GR. A human dopamine transporter cDNA predicts reduced glycosylation, displays a novel repetitive element and provides racially-dimorphic TaqI RFLPs. *Mol Brain Res* 1992; **15**: 161–166.
21 Don RH, Cox PT, Wainwright BJ, Baker K, Matick JS. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* 1991; **19**: 4008.
22 Lewontin RC. On measures of gametic disequilibrium. *Genetics* 1988; **120**: 849–852.
23 Abecasis GR, Cookson WO. GOLD – Graphical overview of linkage disequilibrium. *Bioinformatics* 2000;**16**: 182–183.
24 Swofford DL. *PAUP.* Phylogenetic Analysis Using Parsimony (* and Other Methods)*. Version 4.0. Sinauer Associates: Massachusetts, 1998.
25 Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 1985; **111**: 147–164.
26 Kouzmenko AP, Pereira A, Singh BS. Intronic sequences are involved in neural targeting of human dopamine transporter gene expression. *Biochem Biophys Res Comm* 1997; **240**: 807–811.
27 Byerley W, Hoff M, Holik J, Caron MG, Giros B. VNTR polymorphism for the human dopamine transporter gene (DAT1). *Hum Mol Genet* 1993; **2**: 335.
28 Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J *et al*. Haplotype structure and population genetic inferences from the nucleotide sequence variation in human lipoprotein lipase. *Am J Hum Genet* 1998; **63**: 595–612.
29 Bonnen PE, Story MD, Ashorn CL, Buchholz TA, Weil MM, Nelson DL. Haplotypes at ATM identify coding-sequence variation and indicate a region of extensive linkage disequilibrium. *Am J Hum Genet* 2000; **67**: 1437–1451.