Published in final edited form as:

Ann Hum Genet. 2009 May ; 73(Pt 3): 274–282. doi:10.1111/j.1469-1809.2009.00513.x.

SNP discovery and haplotype analysis in the segmentally duplicated *DRD5* coding region

D. J. E. HOUSLEY^a, M. NIKOLAS^b, P. J. VENTA^a, K. A. JERNIGAN^a, I. D. WALDMAN^c, J. T. NIGG^b, and K. H. FRIDERICI^a

^a Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI

^b Dept. of Psychology, Michigan State University, East Lansing, MI

^c Dept. of Psychology, Emory University, Atlanta, GA

SUMMARY

The dopamine receptor 5 gene (*DRD5*) holds much promise as a candidate locus for contributing to neuropsychiatric disorders and other diseases influenced by the dopaminergic system, as well as having potential to affect normal behavioral variation. However, detailed analyses of this gene have been complicated by its location within a segmentally duplicated chromosomal region. Microsatellites and SNPs upstream from the coding region have been used for association studies, but we find, using bioinformatics resources, that these markers all lie within a previously unrecognized second segmental duplication (SD). In order to accurately analyze the *DRD5* locus for polymorphisms in the absence of contaminating pseudogene sequences, we developed a fast and reliable method for sequence analysis and genotyping within the *DRD5* coding region. We employed restriction enzyme digestion of genomic DNA to eliminate the pseudogenes prior to PCR amplification of the functional gene. This approach allowed us to determine the *DRD5* haplotype structure using 31 trios and to reveal additional rare variants in 171 unrelated individuals. We clarify the inconsistencies and errors of the recorded SNPs in dbSNP and HapMap and illustrate the importance of using caution when choosing SNPs in regions of suspected duplications. The simple and relatively inexpensive method presented herein allows for convenient analysis of sequence variation in *DRD5* and can be easily adapted to other duplicated genomic regions in order to obtain good quality sequence data.

Keywords

Dopamine receptor; Behavior; ADHD; Genomic rearrangement; Segmental duplication; Paralogous sequence variant

INTRODUCTION

Dopamine receptor genes are key players in regulating neurotransmission in response to dopamine and, therefore, are of central interest in a range of neuropsychiatric diseases, including schizophrenia, substance addiction, and attention deficit hyperactivity disorder (ADHD). Since many behaviors and psychiatric conditions show moderate to high heritability, the study of genetic variation in the dopamine-related genes will contribute to unraveling the molecular mechanisms of the catecholamine pathway. These insights into the

role of genetic variability could contribute to the development of personalized pharmaceutical intervention based upon the genotype of each individual patient.

There are five dopamine receptors belonging to two families, characterized by whether they activate (*DRD1* and *DRD5*) or inhibit (*DRD2*, *DRD3* and *DRD4*) adenylate cyclase in response to ligand binding (Gingrich and Caron 1993). These dopamine receptors have been the subject of numerous studies focusing on the role of dopaminergic genes in a variety of diseases (Wong, Buckle, and Van Tol 2000). *DRD5* has been examined for association primarily with ADHD but also with other disorders such as blepharospasm, cervical dystonia, migraine, smoking initiation and nicotine dependence, schizophrenia, bipolar disorder, and substance abuse (Mill et al. 2004; Barr et al. 2001; Hawi et al. 2003; Daly et al. 1999; Misbahuddin et al. 2002; Placzek et al. 2001; Shepherd et al. 2002; Sullivan et al. 2001; Muir et al. 2001; Williams et al. 1997; Asherson et al. 1998; Vanyukov et al. 1998).

Particular methodological problems confront those attempting to understand the role of the *DRD5* gene because of the presence in the human genome of two segmental duplications each with more than 95% sequence identity to the *DRD5* gene, reported to encompass the transcribed region of the *DRD5* gene and 14 kb of downstream sequence (Grandy et al. 1991; Sunahara et al. 1991). In order to examine the association of the *DRD5* locus with disease status, most studies have used two polymorphic microsatellites generally assumed to be outside the duplication, one located 18.2 kb upstream of the transcription start site and one within the *DRD5* promoter (Daly et al. 1999; Mill et al. 2004). However, in this paper, we call attention to an additional segmental duplication encompassing the entire region upstream of the *DRD5* transcription start site, including the microsatellites, which has the potential to significantly contribute to genotyping errors.

The goal of this study was to characterize the true *DRD5* coding region without interference from the closely related pseudogenes. Although several variable SNPs have already been identified within or directly surrounding the *DRD5* locus, they are not often used for association studies, possibly due to concern about confounding results produced by the presence of the segmental duplications and the technically challenging nature of the assays (Feng et al. 1998; Sobell et al. 1995; Asherson et al. 1998).

Herein we report, for the first time, the haplotype structure of the *DRD5* coding region. We describe a simple and reliable method that allows confident, high volume genotyping of known SNPs within the *DRD5* coding region, while simultaneously allowing the discovery of novel variation. Analysis of the *DRD5* coding region reveals two highly polymorphic SNPs in weak linkage disequilibrium and a low frequency of rare variants.

MATERIALS AND METHODS

DNA samples

Buccal swab samples were obtained from 31 trios consisting of one child and their biological parents and 171 unrelated control individuals. The sample population was drawn from unrelated families and individuals in mid-Michigan, consisting of 71% Caucasians, 16% African American, and 13% other (mixed ethnicity, Latino, Asian, or Pacific Islander). Genomic DNA was prepared from buccal swabs using phenol/chloroform as previously described (Meulenbelt et al. 1995). Informed written parental consent and written assent for minor children was obtained according to University IRB-approved procedures.

Bioinformatics analysis

Human *DRD5* coding region sequences were obtained from the UCSC genome browser (Kent et al. 2002) (<http://genome.ucsc.edu>; March 2006 build). The *DRD5* coding region

and 200 bp of surrounding sequence were compared to the whole human genome sequence using BLAT (Kent 2002). The corresponding homologous pseudogene sequences on HSA1 and HSA2 (human chromosome 1 and human chromosome 2) were obtained from the browser. Each pseudogene sequence was compared to *DRD5* by BLAST (Altschul et al. 1990) (<http://www.ncbi.nlm.nih.gov/BLAST/>) and restriction sites unique to the pseudogenes were identified.

Elimination of pseudogenes, and PCR amplification of *DRD5*

Genomic DNA was digested with *NcoI*, trace amounts of DNA uncut by *NcoI* were removed with T7 endonuclease, and the *DRD5* gene was amplified in two minimally overlapping fragments using the following primers: fragment A (P1: 5'-GAGGGTCCCTGGCTGAG-3' and P2: 5'-CCCTCTCCAGGAGGAAATC-3'); fragment B (P3: 5'-GACTCCAGCCTGAATCGAAC-3' and P4: 5'-GATAAAGGGAGCAGCACTGG-3'). Each primer set was designed such that the amplicon would span the *NcoI* restriction site that is found in the homologous pseudogenes (see online supplement for additional detail).

Genotyping and haplotype analysis

Sequencing was performed using both the forward and reverse primers in separate reactions, on an ABI Prism 3700 DNA Analyzer at the Research Technology Support Facility at Michigan State University and analyzed using SequencherTM v4.7 (Gene Codes Corp., Ann Arbor, MI). Haplotypes for each parent were inferred from parent and offspring genotypes. Allele and haplotype frequencies were estimated using only the 60 parents from the trios and 171 individual controls, who were assumed to represent unrelated individuals in the sample population. The program Haploview (version 3.32) (www.broad.mit.edu/mpg/haploview/) (Barrett et al. 2005) was used to examine the LD structure and to estimate D' and r^2 (see online supplement for additional detail).

Genomic duplication breakpoints and rearrangements

The segmental duplication (SD) track from the UCSC human genome browser was used to identify the extent of the segmental duplications surrounding *DRD5* and to compare the two pseudogenes to one another. Additional genomic sequence from HSA2 was identified from the trace database (BLASTn: Unfinished High Throughput Genomic Sequences) by BLAST (see online supplement for additional details).

RESULTS

Determination of the extent of duplications

A bioinformatics approach was used to analyze the extent and boundaries of the *DRD5* duplications and to determine how these duplications might interfere with genotyping of the *DRD5* gene in humans. This analysis (described briefly below and provided in detail in the online supplementary material) revealed that the first duplication event of the *DRD5* region was from HSA4 to HSA2, followed by an inversion on HSA2, and finally by a second duplication event from HSA2 to HSA1 (figure 1). A partial analysis of the pseudogenes was previously published by another group, but it was limited because, at that time, the complete human genome sequence was not available (Marchese et al. 1995).

Events leading to the segmental duplications containing the *DRD5* pseudogenes

An inferred order of events leading to the creation the *DRD5* pseudogenes is shown in figure 1. Because the *DRD5* pseudogenes are also found on gorilla chromosomes 1 and 2, these events most likely occurred prior to the divergence of the human and gorilla lineages (Marchese et al. 1995). The most likely scenario by which the duplications occurred

involves a segmental duplication of a 32.9 kb region containing the *DRD5* gene from chromosome 4 to the pericentromeric region of an ancestral chromosome 2. A subsequent inversion containing part of the segmental duplication (SD) occurred from the original duplicated region fragmenting it into two pieces (which we refer to as the distal and proximal regions) separated by a 22.2 kb segment of DNA picked up from the ancestral chromosome 2 (shown in pink in figure 1). An even larger segment encompassing this entire region was subsequently duplicated to chromosome 1.

The microsatellite commonly used for psychiatric association studies of *DRD5* (Daly et al. 1999) is present in the proximal SD (encompassing 18.8 kb upstream of *DRD5*) on HSA2. Its presence on HSA1 is likely, given that a clone (AC023576), containing unassembled pieces of HSA1, contains 654 bp of sequence, including the microsatellite, that is 92% identical to the 5' end of the proximal duplication of HSA2 and extends 67 bp 5' to the HSA2/HSA4 breakpoint.

For accurate genotyping, potential PCR products from these segmental duplications must be eliminated for studies that include not only the coding region on HSA4, but also the nearby flanking regions (from 8,933,730 to 9,410,892, NCBI Build 36.1).

Elimination of pseudogenes in the *DRD5* amplicon

Comparison of *DRD5* to the two pseudogenes, *ΨDRD5-1* and *ΨDRD5-2*, revealed several restriction enzyme sites unique to the pseudogenes. The *NcoI* recognition site was chosen because it is present in both pseudogenes directly in the middle of the region homologous to the coding exon (figure 2A). The coding region of *DRD5* is contained within a single exon and is over 1400 bp in length with 60.8% GC content, making it difficult to amplify in one piece. Therefore, the coding region was amplified in two overlapping segments surrounding the *NcoI* site. Using this strategy, only a single restriction enzyme digestion of the genomic DNA was needed prior to amplification with either primer set.

Before proceeding with sequence analysis, we verified that the pseudogene was not present in the PCR product (figure 2B). Genomic DNA not treated with enzyme prior to PCR generated mainly pseudogene amplicons as seen by the presence of primarily *NcoI* digested DNA post-amplification (lane 3; figure 2B). *NcoI* pretreatment eliminated the majority of the pseudogene sequences (lane 4; figure 2B), however, a residual amount of contaminating sequence remained. We reasoned that this might be due to the presence of a small amount of single-stranded DNA at the *NcoI* site which could not be efficiently cut by the restriction enzyme, therefore allowing some pseudogene sequence to still be amplified (Svaren et al. 1987; Belotserkovskii and Johnston 1997). The inclusion of a single-strand specific endonuclease along with the restriction enzyme eliminated the majority of the visual trace of the contaminating pseudogenes (lane 5; figure 2B). Sequence analysis of the resulting PCR products from fragments A and B confirmed the absence of pseudogene sequence in the amplicons.

Sequence analysis, SNP detection, and haplotype structure of *DRD5* after elimination of the pseudogenes

To test for variation and determine *DRD5* haplotype structure, genomic DNA from 31 parent/child trios was treated and sequenced as described above. Four previously described SNPs (L88F, P326P, C335X, and C1481T) were identified and their details are summarized in Table 1. Of the two common SNPs, rs6283 (P326P) is a synonymous variation and rs1967551 (C1481T) exists in the 3'UTR. The two additional variants were each seen once in two separate parents: one change results in a non-synonymous substitution (L88F) and the

other results in a nonsense mutation (C335X). Both of these rare variants were found in a heterozygous state, each in a different parent from unrelated families.

In order to genotype the common SNPs in a large sampling of unrelated individuals, 171 subjects were sequenced for fragment B, which contained both common polymorphisms. Overall minor allele frequencies for rs6283 and rs1967551, calculated using the 171 unrelated individuals and the parents from the trios, were .31 and .44, respectively (Table 1). This re-sequencing effort revealed three additional SNPs, two of which (P330Q and P330P) were previously reported by Sobell and colleagues (Sobell et al. 1995). Allele frequencies for all identified SNPs are reported in Table 1. The allele frequencies for the two common SNPs are also summarized by self-identified ethnicity (Table 2). No individuals homozygous for the rare alleles were identified. One novel non-synonymous SNP, D342E, was discovered in a single individual. The three other rare variants reported by Sobell et al. (1995) were not found in our sample.

Haplotypes for the two common SNPs were assembled using the data from thirty informative trios (Table 3). All four possible haplotypes were found in the population. Haplotype 3 was the most common, occurring on 56% of the chromosomes. The other three haplotypes were each present on 13–15% of the population chromosomes. These two SNPs are in statistically significant linkage disequilibrium (LD) ($D' = 0.34$; $r^2 = 0.11$, $p < 0.002$).

Analysis of gene/pseudogene sequence differences and previously reported putative SNPs

The dbSNP database was queried for SNPs listed within the coding region sequenced in our study in order to compare the SNPs present in the published database with those found by direct sequencing (Table 4). Although 23 SNPs are listed in dbSNP, we observed only two of them in our study population (rs6383 and rs1967551). The other SNPs we identified did not appear in either database. A search of the HapMap database revealed only five SNPs with extremely low reported MAFs within and surrounding *DRD5*. One SNP (rs1967551) represents C1481T with a reported heterozygosity in HapMap of <1%, based upon the report for record ss44517958. In contrast, this SNP had a heterozygosity of 0.43 in our population of mixed ethnicity and is reported to have an MAF of 0.35 in only one of the three dbSNP records (ss2871256). The other SNP that was identified and genotyped in our study, rs6283, does not exist in the Hapmap database but is reported in dbSNP with allele frequencies of 0.325 and 0.16. Comparison of the *DRD5* sequence to that of the pseudogenes revealed that three of the other four SNPs found in HapMap in the *DRD5* gene, all with reported MAFs <1%, probably represent sequence differences between *DRD5* and the two pseudogene sequences.

An analysis comparing all 23 recorded SNPs in the *DRD5* coding region to nucleotide differences with the two pseudogenes revealed that 13 of the SNPs correspond to differences with the pseudogenes (Table 4; Likely PSV). It is possible that some of these and the other non-validated SNPs actually represent polymorphisms on HSA1 or HSA2 or, perhaps more likely, represent differences between duplicated chromosomal segments that are incorrectly assigned as polymorphisms by the computer algorithms used to detect SNPs.

DISCUSSION

The presence of segmental duplications surrounding the *DRD5* gene has been an impediment to accurate genotyping of markers near the locus. These technical complications obscure efforts to evaluate its association with neuropsychiatric disorders. Thus, it is important to fully understand the genomic structure of the gene and to develop an accurate

and convenient method to genotype this locus so its involvement can be properly evaluated. The present paper addresses this problem with a new analysis of the *DRD5* gene.

Previous studies discovered and described the distal duplication of the *DRD5* genomic region (Grandy et al. 1991; Eubanks et al. 1992). However, the publication of the nearly complete human genome sequence allowed us to use bioinformatics to examine the duplication in detail. This approach revealed a previously unreported, proximal duplicated region that includes the *DRD5* basal promoter and 16.8 bp of upstream sequence. It is tempting to speculate that any regulatory variation of *DRD5* affecting phenotypes will lie proximal to the *OTOP1/DRD5* inversion breakpoint inferred from this study (i.e., bases centromeric to 9,373,840 on HSA4; details in the online supplement).

The proximal SD contains 16.8 kb of sequence upstream from the actual *DRD5* transcription start site, as defined by Beischlag et al., and extends 18 bp into the 5' UTR (Beischlag et al. 1995). It should be noted that the TSS reported in the UCSC genome browser is not consistent with the experimentally derived TSS established using RACE (Beischlag et al. 1995). The proximal SD exists in an inverted orientation and is located 22.2 kb away from the distal duplication containing the pseudogene on HSA1 and 2 (figure 1).

The presence of this proximal duplication at two other loci could affect genotyping of the commonly used STR marker, which lies within the 5' end of this SD on HSA1 and HSA2 (figure 1). One of the published PCR primers is a perfect match, while the other primer has five mismatches with the duplication, four of which are at the extreme 5' terminus, so it may still be possible to get co-amplification of unwanted products along with the sequence of interest, complicating genotyping calls. In addition, the other SNP that has been used for association studies, C1481T, lies in the doubly duplicated distal SD and similar complications could arise in genotyping this polymorphism.

Although the original technique used to identify variants in the *DRD5* gene, called REF (restriction endonuclease fingerprinting), was unique, this method is not amenable to contemporary automation and did not incorporate an internal control to ensure lack of pseudogene amplification (Sobell et al. 1995). Our goal of re-sequencing the *DRD5* gene in a large sample required the development of an approach more suited to a larger scale of genotyping and the inclusion of internal controls for specific amplification of the locus of interest without interference from the SDs. Given that roughly 5% of the human genome is segmentally duplicated and that many of these regions contain active genes, we believe that this approach should be of value to other researchers faced with similar predicaments.

A preliminary screen of 31 trios for variants in the *DRD5* coding region facilitated accurate genotyping of two SNPs with minor allele frequencies >0.3 and two previously reported rare variants, one non-synonymous change and one premature stop codon (Feng et al. 1998; Sobell et al. 1995). Haplotype analysis of the association between the two common SNPs, separated by 503 bp, revealed that they are in significant, but weak, LD ($D'=0.34$; $p<0.002$, Fisher's exact test). An additional 171 unrelated individuals were re-sequenced for the fragment in order to determine allele frequencies for each common SNP in a larger population. This re-sequencing approach revealed only three additional low frequency SNPs (P330P, P330Q, and D342E).

Although neither of the two common SNPs are predicted to be functional polymorphisms, since rs6283 (P326P) is synonymous and rs1967551 (C1481T) exists in the 3' UTR (and which does not appear to overlie an miRNA target), they are highly variable markers across populations and will be useful for further association analyses for testing genetic linkage of the *DRD5* coding region to any phenotype of interest, particularly ADHD given the previous results of positive association of the linked microsatellite with the disorder (Lowe et al.

2004). These markers are only 503 bp apart, but the weak degree of LD suggests that they exist in separate haplotype blocks and can both be genotyped using the same sequencing reaction, thus enabling the two LD blocks to be tested for association using a single assay. This re-sequencing method also allows for screening of less frequent mutations, such as the nonsense mutation (C335X), P330Q, P330P, and the newly discovered variant, D342E.

The characterization of the extent of duplications in this region illustrates the importance of examining genomic structure prior to embarking on large genotyping studies and choosing markers which do not lie within duplicated chromosomal regions. Caution should be taken in using non-validated SNPs in regions of duplications due to an inability to determine on which chromosome the actual variation exists. In the case of *DRD5*, choosing a nearby non-duplicated area to examine is not possible because this locus is surrounded by numerous segmental duplications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by RO1MH070004 (KHF) and MSU Office of the Vice President for Research (JTN)

References

- Altschul SF, et al. Basic local alignment search tool. *J Mol Biol.* 1990; 215 (3):403–410. [PubMed: 2231712]
- Asherson P, et al. A study of chromosome 4p markers and dopamine D5 receptor gene in schizophrenia and bipolar disorder. *Mol Psychiatry.* 1998; 3 (4):310–320. [PubMed: 9702739]
- Barr CL, et al. Haplotype study of three polymorphisms at the dopamine transporter locus confirm linkage to attention-deficit/hyperactivity disorder. *Biol Psychiatry.* 2001; 49 (4):333–339. [PubMed: 11239904]
- Barrett JC, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21 (2):263–265. [PubMed: 15297300]
- Beischlag TV, et al. The human dopamine D5 receptor gene: cloning and characterization of the 5'-flanking and promoter region. *Biochemistry.* 1995; 34 (17):5960–5970. [PubMed: 7727453]
- Belotserkovskii BP, Johnston BH. Denaturation and association of DNA sequences by certain polypropylene surfaces. *Anal Biochem.* 1997; 251 (2):251–262. [PubMed: 9299024]
- Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui LC, Scherer SW. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.* 2003; 4:R25. [PubMed: 12702206]
- Daly G, et al. 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 (2):192–196. [PubMed: 10208453]
- Eubanks JH, et al. Localization of the D5 dopamine receptor gene to human chromosome 4p15.1-p15.3, centromeric to the Huntington's disease locus. *Genomics.* 1992; 12 (3):510–516. [PubMed: 1532789]
- Feng J, et al. Scanning of the dopamine D1 and D5 receptor genes by REF in neuropsychiatric patients reveals a novel missense change at a highly conserved amino acid. *Am J Med Genet.* 1998; 81 (2): 172–178. [PubMed: 9613858]
- Gingrich JA, Caron MG. Recent advances in the molecular biology of dopamine receptors. *Annu Rev Neurosci.* 1993; 16:299–321. [PubMed: 8460895]
- Grandy DK, et al. Multiple human D5 dopamine receptor genes: a functional receptor and two pseudogenes. *Proc Natl Acad Sci USA.* 1991; 88 (20):9175–9179. [PubMed: 1833775]

- Hawi Z, et al. Linkage disequilibrium mapping at DAT1, DRD5 and DBH narrows the search for ADHD susceptibility alleles at these loci. *Mol Psychiatry*. 2003; 8 (3):299–308. [PubMed: 12660802]
- Horvath JE, et al. Punctuated duplication seeding events during the evolution of human chromosome 2p11. *Genome Res*. 2005; 15 (7):914–927. [PubMed: 15965031]
- Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res*. 2002; 12 (4):656–664. [PubMed: 11932250]
- Kent WJ, et al. The human genome browser at UCSC. *Genome Res*. 2002; 12 (6):996–1006. [PubMed: 12045153]
- Lowe N, et al. Joint analysis of the DRD5 marker concludes association with attention-deficit/hyperactivity disorder confined to the predominantly inattentive and combined subtypes. *Am J Hum Genet*. 2004; 74 (2):348–356. [PubMed: 14732906]
- Marchese A, et al. Two gene duplication events in the human and primate dopamine D5 receptor gene family. *Gene*. 1995; 154 (2):153–158. [PubMed: 7890157]
- Meulenbelt I, et al. High-yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. *Am J Hum Genet*. 1995; 57 (5):1252–1254. [PubMed: 7485180]
- Mill J, et al. Polymorphisms in the dopamine D5 receptor (DRD5) gene and ADHD. *Am J Med Genet B Neuropsychiatr Genet*. 2004; 125B (1):38–42. [PubMed: 14755441]
- Misbahuddin A, et al. A polymorphism in the dopamine receptor DRD5 is associated with blepharospasm. *Neurology*. 2002; 58 (1):124–126. [PubMed: 11781417]
- Muir WJ, et al. Markers close to the dopamine D5 receptor gene (DRD5) show significant association with schizophrenia but not bipolar disorder. *Am J Med Genet*. 2001; 105 (2):152–158. [PubMed: 11304828]
- Placzek MR, et al. Cervical dystonia is associated with a polymorphism in the dopamine (D5) receptor gene. *J Neurol Neurosurg Psychiatry*. 2001; 71 (2):262–264. [PubMed: 11459908]
- Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*. 2000; 132:365–386. [PubMed: 10547847]
- Shepherd AG, et al. Dopamine receptor genes and migraine with and without aura: an association study. *Headache*. 2002; 42 (5):346–351. [PubMed: 12047334]
- Sobell JL, et al. The D5 dopamine receptor gene in schizophrenia: identification of a nonsense change and multiple missense changes but lack of association with disease. *Hum Mol Genet*. 1995; 4 (4):507–514. [PubMed: 7633397]
- Sullivan PF, et al. An association study of DRD5 with smoking initiation and progression to nicotine dependence. *Am J Med Genet*. 2001; 105 (3):259–265. [PubMed: 11353446]
- Sunahara RK, et al. Cloning of the gene for a human dopamine D5 receptor with higher affinity for dopamine than D1. *Nature*. 1991; 350(6319):614–619. [PubMed: 1826762]
- Svaren J, et al. DNA denatures upon drying after ethanol precipitation. *Nucleic Acids Res*. 1987; 15 (21):8739–8754. [PubMed: 3684571]
- Vanyukov MM, et al. An association between a microsatellite polymorphism at the DRD5 gene and the liability to substance abuse: pilot study. *Behav Genet*. 1998; 28 (2):75–82. [PubMed: 9583233]
- Williams NM, et al. Association between schizophrenia and a microsatellite polymorphism at the dopamine D5 receptor gene. *Psychiatr Genet*. 1997; 7 (2):83–85. [PubMed: 9285966]
- Wong AH, Buckle CE, Van Tol HH. Polymorphisms in dopamine receptors: what do they tell us? *Eur J Pharmacol*. 2000; 410 (2–3):183–203. [PubMed: 11134669]

Figure 1.

Schematic representation of the segmental duplications (SD) surrounding *DRD5*. The large blue arrow represents the *DRD5* coding region, the small blue line is the 5' UTR. The positions of the TSS and the commonly used STR are indicated with arrows. Black lines represent unique sequence on each chromosome. The green box represents the proximal SD. The green hatched box represents homologous sequence between HSA4 and HSA2, however, this region is missing in the assembly of HSA1 (gap). The light blue box represents the distal SD, which exists in an inverted orientation to the proximal SD on HSA1 and 2. The orange box is unique sequence within the 5' UTR which is not found in either SD. The pink box and the dotted gray box represent sequence that is homologous between HSA1 and 2 only. BP indicates the 5' breakpoint between the distal duplication and HSA1 and 2 as described by Beischlag et al. (1995). Numbers represent nucleotide positions for each breakpoint as determined from the UCSC browser (March 2006 build).

Figure 2.

Figure 2A. Schematic of the *DRD5* amplification strategy and detail of restriction enzyme recognition site. The black line represents the HSA4 region containing *DRD5*, the approximate positions of the start codon (T) and stop codon (V) are indicated. The two PCR fragments are shown with forward and reverse primers and locations indicated by arrows. The highlighted box indicates the region surrounding the *NcoI* restriction enzyme site (CCATGG) in the pseudogenes, or lack thereof in *DRD5*.

Figure 2B. Restriction enzyme analysis of the *DRD5* fragment “A” PCR product for coamplification of pseudogene fragments. PCR amplifications were carried out with primers P1 and P2 using genomic DNA template as follows: lanes 2 & 3: no pre-treatment; lane 4: *NcoI* predigest; lane 5: *NcoI* and T7 endonuclease predigest. Amplification results in a PCR product of 905 bp (lane 2). *NcoI* digestion following PCR cleaves 80 bp from the pseudogene amplicons resulting in an 825 bp fragment if coamplification products are present (lanes 3–4). Lane 1: 100 bp ladder (New England Biolabs); lane 6: negative control. The minor products larger than 905 bp in lanes 2, 3, and 4 are assumed to be heteroduplexes between the three coamplified PCR products.

Table 1
Sequence variations discovered in the DRD5 gene using trios and individual subjects.

SNP ID	nucleotide change ^a	codon change	minor allele	MAF ^b	H ^c
rs6283	C262T ^d	L88F		<0.01	
	T978C ^e	P326P	C	0.31	0.43
	C989A ^e	P330Q		0.017	
	G990A ^e	P330P		<0.01	
	C1005A ^e	C335X		<0.01	
rs1967551	C1026A	D342E		<0.01	
	C1481T ^e	3' UTR	T	0.44	0.46

^a numbered from the translation start site
^b MAF: minor allele frequency (determined from parents from trios and/or individual subjects)
^c H: heterozygosity
^d Feng J. et al. 1998
^e Sobell J.L. et al. 1995

Table 2

Allele frequencies by ethnicity for 171 individuals.

ethnicity	n	rs6283				rs1967551			
		allele	allele count	allele freq	H ^a	allele	allele count	allele freq	H ^a
Caucasian		T	169	0.70		T	95	0.39	
		C	73	0.30		C	147	0.61	
	121	total	242		0.39	total	242		0.49
African American		T	32	0.57		T	36	0.64	
		C	24	0.43		C	20	0.36	
	28	total	56		0.57	total	56		0.36
others ^b		T	34	0.77		T	21	0.48	
		C	10	0.23		C	23	0.52	
	22	total	44		0.45	total	44		0.41
overall		T	235	0.69		T	152	0.44	
		C	107	0.31		C	190	0.56	
	171	total	342		0.43	total	342		0.46

^aH: Heterozygosity

^bothers includes: Latinos, Asians, Pacific islanders

DRD5 haplotypes in parents from 30 informative trios.

Table 3

Haplotype ID	rs6283	rs1967551	Observed no. of chromosomes ^a	Observed percentage	Expected no. of chromosomes ^b	Expected percentage ^b
1	C	C	16	13.8	23.7	20.4
2	C	T	18	15.5	10.3	8.9
3	T	C	65	56.0	57.3	49.3
4	T	T	17	14.7	24.7	21.4

^aThe rs1967551 genotype was not obtained from two individuals, so 4 chromosomes are missing from the dataset of 60 individual parents (30 families).

^bBased upon Hardy-Weinberg equilibrium.

Table 4

Information obtained on putative DRD5 coding region SNPs from dbSNP and HapMap.

SNP ref ID	Likely PSV ^a	Het ^b	Alleles	Functional class	Validation method ^c	Method class ^c
rs2227839	yes	0.11	C/T	synonymous	frequency	unknown
rs2227840	yes	0.10	C/G	nonsynonymous	frequency	hybridize, unknown
rs2227848	yes	0.07	C/T	synonymous	frequency	unknown
rs6282	yes	0.00	G/T	nonsynonymous	hapmap	unknown
rs2227849	yes	0.04	A/G	nonsynonymous	frequency	unknown
rs2227841	yes	0.14	C/T	synonymous	frequency	unknown
rs2227845	yes	0.41	G/T	nonsynonymous	frequency	unknown
rs2227846	no	0.22	C/T	synonymous	frequency	unknown
rs2227843	no	0.07	A/G	nonsynonymous	frequency, hapmap	hybridize, unknown
rs2227852	yes	0.38	A/G	nonsynonymous	frequency, hapmap	hybridize, unknown
rs2227847	no	0.24	A/G	nonsynonymous	frequency	unknown
rs2227853	no	0.20	C/T	synonymous	frequency	unknown
rs2227842	yes	0.31	C/T	nonsynonymous	frequency	unknown
rs2227850	yes	0.07	C/T	nonsynonymous	frequency, hapmap	unknown
rs2227851	no	0.00	A/C	nonsynonymous	not validated	unknown
rs2227844	no	0.19	A/G	synonymous	frequency	unknown
rs6283^d	no	0.33	C/T	synonymous	2hit-2allele, frequency	computed, sequence, unknown
rs1800762	no	0.11	A/C	nonsynonymous	frequency	unknown
rs6281	no	0.00	C/T	synonymous	not validated	unknown
rs3205146	yes	0.26	A/G	synonymous	frequency	computed
rs16888557	yes	0.46	A/T	synonymous	frequency	unknown
rs16888561	yes	0.48	A/G	nonsynonymous	frequency	unknown
rs1967551^d	no	0.00	A/G	mRNA-3'UTR	2hit-2allele, hapmap	computed, unknown

^aLikely PSV's (paralogous sequence variants; Cheung et al. 2003); the nucleotide change described by the SNP also represents a sequence difference between the gene and each pseudogene

^bHeterozygosity as reported in dbSNP

^cValidation and Method class are as described in dbSNP

^dboldface type indicates the two SNPs confirmed herein