High prevalence of rare dopamine receptor D4 alleles in children diagnosed with attention-deficit hyperactivity disorder

DL Grady¹, H-C Chi¹, Y-C Ding¹, M Smith², E Wang¹, S Schuck³, P Flodman², MA Spence², JM Swanson³ and RK Moyzis¹,³

¹Department of Biological Chemistry, College of Medicine, University of California, Irvine, CA, USA; ²Department of Pediatrics, University of California, Irvine, Medical Center, Orange, CA, USA; ³Child Development Center, University of California, Irvine, CA, USA

Associations have been reported of the 7-repeat (7R) allele of the human dopamine receptor D4 (DRD4) gene with both the personality trait of novelty seeking and attention-deficit/hyperactivity disorder (ADHD). The increased prevalence of the 7R allele in ADHD probands is consistent with the common variant–common disorder hypothesis, which proposes that the high frequency of many complex genetic disorders is related to common DNA variants. Recently, based on the unusual DNA sequence organization and strong linkage disequilibrium surrounding the DRD4 7R allele, we proposed that this allele originated as a rare mutational event, which nevertheless increased to high prevalence in human populations by positive selection. We have now determined, by DNA resequencing of 250 DRD4 alleles obtained from 132 ADHD probands, that most ADHD 7R alleles are of the conserved haplotype found in our previous 600 allele worldwide DNA sample. Interestingly, however, half of the 24 haplotypes uncovered in ADHD probands were novel (not one of the 56 haplotypes found in our prior population studies). Over 10 percent of the ADHD probands had these novel haplotypes, most of which were 7R allele derived. The probability that this high incidence of novel alleles occurred by chance in our ADHD sample is much less than 0.0001. These results suggest that allelic heterogeneity at the DRD4 locus may also contribute to the observed association with ADHD.

Molecular Psychiatry (2003) 8, 536–545. doi:10.1038/sj.mp.4001350

Keywords: DRD4; ADHD; cSNPs; allelic heterogeneity; common variant-common disorder (CVCD) hypothesis

Attention-deficit hyperactivity disorder (ADHD) is a neurobehavioral disorder defined by symptoms of developmentally inappropriate inattention, impulsivity, and hyperactivity with early onset.¹ Current estimates indicate that 3–6% of school age children are diagnosed with ADHD, making it the most prevalent disorder of childhood.² While the broad DSM–IV³ phenotype of ADHD almost certainly has multiple biological etiologies,¹ numerous family, twin and adoption studies have documented a strong genetic basis.⁴,⁵ However, given high cross-national variation in the recognition and treatment of ADHD,²,⁶ we proposed that the ADHD-combined type (DSM-IV) without serious comorbidity should be used as a ‘refined’ phenotype in biological and genetic research.⁶

Despite the high heritability of ADHD, initial genome scan studies have failed to identify genes of major effect,⁷ although a region on chromosome 16p13 has been implicated in subsequent studies by the same group.⁸ Such negative results are not unexpected for a complex genetic disorder like ADHD, where phenotypic heterogeneity is likely, and the practical but (to date) restricted sample sizes limit statistical power.⁹–¹³ Candidate gene studies, on the other hand, require much smaller sample sizes to achieve the same statistical power. The efficacy of a dopamine agonist drug, methylphenidate, in the treatment of ADHD has suggested that genes in the dopamine pathway may be involved in the disorder’s etiology.¹²–¹⁴ This dopamine hypothesis of ADHD suggests a number of candidate genes that could logically be tested for their association with the disorder. The draft human genome sequence¹⁵–¹⁷ has provided information sufficient to examine multiple candidate genes in parallel, often representing most of the proteins in a relevant biochemical pathway.
One of these candidate genes, DRD4, located near the telomere of chromosome 11p, is one of the most variable human genes known. Most of this diversity is the result of length and single-nucleotide polymorphism (cSNP) variation in a 48 bp tandem repeat (VNTR) in exon 3, encoding the third intracellular loop of this dopamine receptor. Variant alleles containing two (2R) to 11 (11R) repeats are found, with the resulting proteins having 32–176 amino acids at this position. A number of investigations have found associations between particular alleles of this highly variable gene and behavioral phenotypes. While some studies have suggested that the 7R allele of the DRD4 gene might be associated with the personality trait of novelty seeking, the most reproduced association is between the 7R allele and ADHD. Recently, we showed by DNA resequencing/haplotyping of 600 DRD4 alleles, representing a worldwide population sample, that the origin of 2R through 6R alleles can be explained by simple one-step recombination/mutation events. In contrast, the 7R allele is not simply related to the other common alleles, differing by greater than six recombinations/mutations. Strong linkage disequilibrium (LD) was found between the 7R allele and surrounding DRD4 polymorphisms, suggesting that this allele is at least 5–10-fold ‘younger’ than the common 4R allele. Based on an observed bias towards nonsynonymous amino-acid changes, the unusual DNA sequence organization, and the strong LD surrounding the DRD4 7R allele, we proposed that this allele originated as a rare mutational event, which nevertheless increased to high frequency in human populations by positive selection.

Why is the DRD4 7R allele, which arose recently and underwent strong positive selection, nevertheless now disproportionately represented in individuals diagnosed with ADHD? We suggested that selection for an adjacent polymorphism was unlikely, given the distinct and unusual DNA sequence organization of the DRD4 7R allele itself. The DRD4 7R allele is at moderate prevalence in most populations that have been examined for ADHD (approximately 10–15%). Therefore, the approximate two-fold increase in DRD4 7R allele frequency in ADHD probands, calculated from a recent meta-analysis, is consistent with the common variant–common disorder (CVCD) hypothesis (also called the common disease–common variant hypothesis). In the CVCD hypothesis, the high prevalence of a given disorder (and its associated alleles) is attributed to either (1) the interaction with a new environment (such that genotypes associated with the disorder were not eliminated in the past) or (2) the disorder has small effect on fitness (because it is late onset). We suggested a third possibility. Perhaps predisposing alleles in fact are under positive selection, and only result in deleterious effects when combined with other environmental/genetic factors. This would explain the high prevalence of common disorders in the population, since the selected allele would only be deleterious in a small fraction of those individuals carrying it. Positive selection for particular human alleles may, in fact, be common, contributing to the observation of unexpectedly large blocks of LD in the human genome.

It is a reasonable hypothesis that high-prevalence human genetic disorders will be related to some common variants in the population. However, it is unclear that single common variants will be the only relevant variants. Alleles at low prevalence, most of which have not been identified by current SNP searches targeting a small sample size, could also contribute to complex disease. Of the hundreds of ‘single-hit’ disease genes identified to date, the vast majority contain hundreds of ‘private’ mutations that alter protein function. In order to test this rare variant–common disorder (RVCD) model for complex disease, much greater depth of DNA resequencing must be conducted, ideally in individuals enriched for the putative mutant alleles (ie, probands).

All previous studies of the DRD4/ADHD association have defined alleles based only on PCR length differences. Hence, it is possible that specific sequence variants are actually associated with the disorder. For example, one could imagine that the selected DRD4 7R allele might have a higher mutation rate than the common 4R allele, and it is in fact these variant 7R alleles that predispose to ADHD. Given the large sequence diversity of this gene, in which 56 different exon 3 haplotypes were uncovered in 600 chromosomes obtained from a worldwide sample, we decided that direct DNA resequencing of DNA obtained from ADHD probands was the only method that could answer this question.

Here, we confirm the increased prevalence of DRD4 7R alleles in individuals diagnosed with the refined phenotype of ADHD. By DNA resequencing of 250 DRD4 alleles obtained from 132 ADHD probands, we show that most ADHD-associated 7R alleles are of the conserved haplotype found in our previous 600 allele worldwide DNA sample. Interestingly, however, over 10% of the ADHD probands had novel DRD4 haplotypes, not previously found in our worldwide allele sample. The probability that this high prevalence of novel alleles occurred by chance in our ADHD sample is much less than 0.0001. Most of these novel haplotypes were 7R allele derived. These results suggest that allelic heterogeneity (the RVCD model) may also be contributing to the association of the DRD4 locus with ADHD, as is routinely found for ‘single-gene’ genetic disorders.

Materials and methods
Clinical
ADHD probands were recruited to participate in either clinical trials or the Multimodality Treatment Study of Children with ADHD (MTA) at the...
University of California, Irvine. The refined phenotype of ADHD was diagnosed by a research assessment battery described in detail elsewhere, which includes psychiatric interviews and questionnaires about the symptoms of the disorder and other psychopathological behavior related to comorbid disorders. Instruments used included the Diagnostic Interview Schedule for Children, Fourth Version (DISC-IV), the SNAP-IV Rating Scale and a locally developed family and developmental history questionnaire. In addition, measures of ability and achievement were obtained using the Wechsler Intelligence Scale for Children, Third Revision (WISC-III) and the Wechsler Individual Achievement Test (WIAT). The inclusion criteria included a DSM-IV diagnosis of ADHD-combined type, which requires the endorsement of at least six of the nine symptoms of inattention and six of the nine symptoms of hyperactivity/impulsivity. High cutoffs on parent and teacher ratings of ADHD items on the SNAP rating were required. Subjects with an IQ score on the WISC-III <80 were excluded. Information was also obtained for oppositional defiant disorder (ODD), but a comorbid diagnosis of ODD did not exclude the subject. A diagnosis of other comorbid disorders (such as Tourette syndrome), or treatment of symptoms of other disorders with nonstimulant psychotropic drugs, were exclusion criteria for this study.

Establishing cell lines and DNA purification

Lymphoblastoid cell lines were established for all ADHD probands. Methods for transformation, cell culture, and DNA purification have been described. PCR amplification and DNA sequencing

The DRD4 exon 3 VNTR was amplified with primer sets described previously (5′-CGTACTGTGCGCGCTTCAACGA-3′ and 5′-GACACAGCGCGCTGCTGATGT-3′; 705 nucleotide product for the 4R-allele;22). PCR reactions were conducted in 25 μl volumes, containing 100 ng genomic DNA, 200 μM dNTPs, 0.5 μmol of each primer, 1 × PCR buffer (Qiagen), 1 × Q-solution (Qiagen) and 0.625 U Taq DNA polymerase (Qiagen). Amplification was performed using Perkin–Elmer 9700 thermal cyclers. A 20 s, 96 °C hot start was used, followed by 40 cycles of 95 °C for 20 s and 68 °C for 1 min. Following a 4-min chase at 72 °C, excess primers were eliminated with 0.5 U of shrimp alkaline phosphatase (SAP, Amersham Life Science), 0.1 U of exonuclease I (ExoI, Amersham Life Science) and 1 × SAP buffer (Amersham Life Science). The SAP/ExoI reaction was carried out at 37 °C for 1 h, followed by a 15-min heat inactivation at 72 °C. DNA from the SAP/ExoI reaction was used directly for DNA sequencing. For individuals heterozygous for DRD4 alleles, the two allelic PCR products were first separated on 1.2% agarose gels,25 DNA cycle sequencing was conducted by standard techniques, using ABI 3100 and 3700 automated sequencers.16 Overall PCR/resequencing success was greater than 95%. One allele from an ADHD proband, 9R(1-8-25-5-2-5-2-23-4), was included in our prior worldwide sample.22 DNA sequences of the novel DRD4 haplotypes reported in this paper have been submitted to GenBank (Accession numbers AY151027–AY151038).

Analysis of sequence data

Analysis of sequence data was accomplished using PHRED, PHRAP, POLYPHRED, and CONSED. These programs are used to clean and assemble the sequence files, and aid in the detection of DNA polymorphism. For every position in the DRD4 consensus sequence, POLYPHRED examines each sample sequence for evidence of polymorphism/heterozygosity. The rank limit for identifying a position as polymorphic is under user control. Based on our experience, we have configured POLYPHRED to identify all potential polymorphisms of rank 1–4, which are then independently evaluated by two skilled investigators.

Capture of individual genotypes/haplotypes into a database (SNPMAN)

The collection of SNPs into a relational database is done via an in-house software package we have designated, SNPMAN. SNPMAN is a package of three main programs written originally in PERL and SQL and now available in both binaries and open source format. The first program (SNPMAN) is designed to collect the SNP information from POLYPHRED output files and transform it into acceptable SQL command files, later to be executed by a database operator (DBO). The second program (MANIP) is the CONSED-add on extension that allows an experienced chromatogram reader to adjust or delete database information in case of false-positive or false-negative polymorphisms. The last program (GIMMEPRETTY-Base) in the SNPMAN package converts existing polymorphism tables into acceptable input files for visual genotyping via VG.

Statistical analysis

Allele distributions were compared using Fischer’s exact test for a 2 × k table, as implemented in SAS (v.6.12, running on a SUN Ultra2 Enterprise workstation). In our prior worldwide sample,22 all DRD4 repeat lengths except for 4R were oversampled by a factor of two. This was corrected for before comparisons were conducted with the present sample.

Results

DNA was isolated from 132 probands diagnosed with the refined phenotype of ADHD, sequentially identified as part of ongoing research and clinical trials programs at the UCI Child Development Center (see Materials and methods). Table 1 gives the demographics, ADHD symptoms, and psychometric test scores of these probands. As expected, the majority (80%) of individuals were of European ancestry and male. On the SNAP, average rating per item
summary scores of inattention and hyperactivity/impulsivity above 2.0 are considered severe. The average SNAP for this group of probands was 2.22 (Table 1). Ratings were also obtained for ODD, often found to be comorbid with ADHD. The observed average for ODD (1.62) was significantly higher than for population norms. Other psychometric measures of IQ (WISC) and achievement (WIAT) were in the normal range for the group (Table 1).

The exon 3 VNTR region of the DRD4 gene was amplified from these DNAs, and the distribution of DRD4 genotypes obtained in this sample is shown in Table 2. As reported in numerous other studies, including our own, the frequency of ADHD individuals with at least one DRD4 7R allele is approximately two-fold greater (43.2%) than that found in ethnically matched control individuals. Interessingly, the observed frequency of 2R and 3R alleles was also increased in this ADHD sample (Table 2). In European populations, the observed allele frequency for DRD4 is 2R = 0.07, 3R = 0.03, 4R = 0.73, 5R = 0.01, 6R = 0.02, 7R = 0.12, 8R < 0.01, 9R < 0.001 (2N = 1652; Chang et al.; Ding et al. and unpublished data). Calculating an expected genotype distribution (assuming Hardy–Weinberg equilibrium, Table 2) indicates that only 22% of Europeans should have a 7R/x genotype, consistent with prior experimental control data from our research. Adjusting these values for the increased frequency of 7R alleles in some non-European populations cannot account for the increased frequency in our predominantly European ancestry ADHD sample (Table 1).

DNA sequence analysis of 250 DRD4 alleles obtained from these ADHD probands found 24 different haplotypes (Table 3). No data were obtained on 14 alleles (5.3%; two 2R, seven 4R, and five 7R alleles) because of PCR and/or sequencing failures. Altogether, we screened over 200 000 bp of genomic DNA and 1132 48-bp repeats. Interestingly, only half (12/24) of the observed haplotypes were identified previously in our analysis of 600 DRD4 alleles obtained from a worldwide population sample (see GenBank Accession Nos. AF395210–AF395264). For example, using our proposed nomenclature for DRD4 haplotypes (Figure 1), the majority of 7R alleles were identified, whereas the majority of the DRD4 alleles were unique to this study.

### Table 1  Demographics, ADHD symptoms, and psychometric test scores of the ADHD probands

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Females (26)</th>
<th>Males (106)</th>
<th>Total (132)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>8.62</td>
<td>8.83</td>
<td>8.78</td>
<td>1.88</td>
</tr>
<tr>
<td>% European</td>
<td></td>
<td></td>
<td>79.7</td>
<td></td>
</tr>
<tr>
<td>% Hispanic</td>
<td></td>
<td></td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td>% African American</td>
<td></td>
<td></td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>% Asian</td>
<td></td>
<td></td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>% Native American</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>% Pacific Island</td>
<td></td>
<td></td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>ADHD symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattention</td>
<td>2.53</td>
<td>2.3</td>
<td>2.41</td>
<td>0.52</td>
</tr>
<tr>
<td>Hyperactivity/impulsivity</td>
<td>2.07</td>
<td>1.96</td>
<td>2.01</td>
<td>0.74</td>
</tr>
<tr>
<td>ADHD average</td>
<td>2.29</td>
<td>2.15</td>
<td>2.22</td>
<td>0.51</td>
</tr>
<tr>
<td>ODD average</td>
<td>1.68</td>
<td>1.57</td>
<td>1.62</td>
<td>0.81</td>
</tr>
<tr>
<td>Psychometric tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WISC block design</td>
<td>101.9</td>
<td>113.3</td>
<td>107.7</td>
<td>33.9</td>
</tr>
<tr>
<td>WISC vocabulary</td>
<td>96.9</td>
<td>105.4</td>
<td>101.2</td>
<td>30.4</td>
</tr>
<tr>
<td>IQ average</td>
<td>99.4</td>
<td>109.4</td>
<td>104.4</td>
<td>25.7</td>
</tr>
<tr>
<td>WIAT reading</td>
<td>96.5</td>
<td>100.5</td>
<td>98.5</td>
<td>14.6</td>
</tr>
<tr>
<td>WIAT math</td>
<td>100.6</td>
<td>101.4</td>
<td>101.9</td>
<td>13.8</td>
</tr>
<tr>
<td>WIAT spelling</td>
<td>96.6</td>
<td>97.8</td>
<td>97.2</td>
<td>13.8</td>
</tr>
<tr>
<td>Achievement</td>
<td>97.8</td>
<td>99.9</td>
<td>98.9</td>
<td>12.5</td>
</tr>
</tbody>
</table>

### Table 2  Genotypes of 132 ADHD Probands

<table>
<thead>
<tr>
<th>Genotype</th>
<th>2R/4R</th>
<th>3R/3R</th>
<th>3R/4R</th>
<th>4R/4R</th>
<th>4R/6R</th>
<th>2R/7R</th>
<th>3R/7R</th>
<th>4R/7R</th>
<th>6R/7R</th>
<th>7R/7R</th>
<th>4R/8R</th>
<th>4R/9R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>20(19)</td>
<td>1</td>
<td>9(8)</td>
<td>43(41)</td>
<td>2(1)</td>
<td>5</td>
<td>3</td>
<td>42</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Expected</td>
<td>14</td>
<td>&lt;1</td>
<td>6</td>
<td>70</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>23</td>
<td>&lt;1</td>
<td>2</td>
<td>1</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>

Numbers of observed and expected genotypes are indicated. Expected genotypes are calculated from European DRD4 allele frequencies (2N = 1652; see text) and assume Hardy–Weinberg equilibrium. Numbers in brackets refer to reduced number of genotypes in the non-7R category, following correction for 7R derivative alleles uncovered by sequence analysis (see text).
High prevalence of rare dopamine receptor D4 alleles

DL Grady et al

Molecular Psychiatry

540

found in our ADHD probands (45/55 = 81.8%) are the common 7R(1-2-6-5-2-5-4) haplotype (Table 3). In this nomenclature, the numbers in brackets refer to different 48 bp repetitive sequence motifs (Figure 1). Likewise, the majority of 2R and 4R alleles were the common 2R(1-4) and 4R(1-2-3-4) haplotypes, respectively. These three common alleles (2R, 4R, and 7R) account for 87.2% of the observed alleles (Table 3), similar to the proportion obtained in our 600 allele population sample.22 The remaining nine alleles are rare 3R, 4R, 6R, and 7R variants observed previously22 (Table 3).

The other half of the observed haplotypes were unique, not identified in our extensive prior analysis22 (Table 3). Excluding the common variants, expected to be present in all samples, 60% (12/20) of rare (<0.01 frequency) variants found in this ADHD sample were unique. In all, 15 ADHD probands had one of these 12 unique DRD4 haplotypes (15/132 = 11.4%). For seven of these probands, parental DNA was available. PCR resequencing indicated that the variant allele was present in one of the parents, and not a new mutation. All but one of these 12 novel alleles produce an altered amino-acid sequence in the resulting DRD4 protein compared to the common allele (Figure 2). For example, the observed 4R(1-2-6-4) variant would substitute a Gly for a Ser and a Pro for a Gln in comparison with the common 4R(1-2-3-4) variant (Table 3). This result is similar to our prior population studies on the DRD4 gene,22 where 87% of the observed rare variants altered the amino-acid sequence of the resulting protein.

The origin of most of these newly observed variants can be inferred to be 7R allele derivatives, based on their nucleotide sequence (Figures 1 and 2). Ten of the 12 haplotype variants contain these motifs (Table 3), and hence likely arose as recombinational or mutational events involving a 7R allele (Figure 2). For example, the 4R(1-2-5-4) allele likely arose as a recombination event between a 4R(1-2-3-4) allele and a 7R(1-2-6-5-2-5-4) allele. Genotyping six of these variant alleles for flanking SNPs diagnostic for the 4R and 7R alleles22 confirmed their hypothesized origin (data not shown; Figure 2). The finding that the majority of these rare variants are derived from 7R alleles should be contrasted with our prior population studies of the DRD4 gene,22 in which rare variants were found to be equally derived from 4R alleles.

Figure 1 Nucleotide and amino acid sequences of DRD4 VNTR motifs. The nucleotide and corresponding amino acid (red) sequences of 35 DRD4 exon 3 48-bp repeat motifs observed to date are shown. Modified from Ref 22.
and 7R alleles. There is an approximate two-fold increase in rare 7R alleles in this ADHD sample in comparison to our prior population sample (18.2% vs 11.0%; Table 3).

Including these 7R related sequence variants in the 7R allele category removes five individuals from the non-7R category, originally classified based on their PCR fragment length (numbers in brackets in Table 2). Altogether, individuals with 7R and derivative 7R alleles account for 47% of the ADHD proband population (Table 2).

In all, 20% of the ADHD DRD4 alleles sequenced in this study are of non-European origin (Table 1). However, 33% (5/15) of the individuals with novel rare alleles were non-European in genetic origin. While this difference is not statistically significant, it is possible that population stratification could account for a portion of the observed difference. Our prior worldwide sequence sample included 220 European DRD4 alleles, as well as 164 Asian, 122 African, 76 North and South American, and 18 Pacific Island ancestry alleles. Non-4R alleles were over-sampled approximately two-fold in this prior study. Given the ethnic breakdown of our ADHD probands (Table 1), then, our prior worldwide resequencing sample can serve as an extensively ‘oversampled’ control, in which we have comparable numbers of European origin alleles, and 10–20-fold larger numbers of non-European alleles.

A total of 67 different haplotype variants of DRD4 were seen in either our prior population sample,22 our ADHD sample (Table 3), or both: 60 of these haplotypes are at low (<0.01) frequency. We can therefore ask a simple question: how likely is it, assuming a pool of uncommon DRD4 alleles, that these two samples (population control and ADHD) would give the observed results? Most of the rare alleles were found only once, hence we can only estimate their frequency in the population. Our initial sample size of 600 chromosomes, however, is expected to detect 80% of variants at a frequency of 0.002 or greater. Based on DRD4 allele frequency distributions (Table 3 and Ding et al22), where the six common 2R–7R alleles account for >90% of the observed alleles, we can estimate that there can be, at most, 85 different DRD4 alleles at frequencies greater than 0.001. At a minimum, therefore, we have identified 79% (67/85) of DRD4 alleles with a frequency greater than 0.001. Alleles less frequent than 0.001 would be found rarely in population samples of the current size, and hence cannot contribute significantly to the observed distributions.

One can consider the possibility, then, that among a pool of uncommon alleles, there were 12 undetected alleles (on 15 chromosomes) that happened by chance to occur among the 250 chromosomes obtained from ADHD individuals. Likewise, one can consider the possibility that these 12 alleles were not found among 600 random chromosomes. We considered a range of allele frequencies for these 12 alleles, from 1/400 each to 1/1000 each. For each set of allele frequencies, the probability of seeing none of these 12 new alleles among the 600 chromosomes examined previously22 can be easily calculated as a multinomial probability (Probability ‘A’). Likewise, the probability of seeing nine of these new alleles once, and three twice, among 250 chromosomes can be calculated (Probability ‘B’). For all sets of allele frequencies, either probability ‘A’ or probability ‘B’ is much less than 0.0001. It is extremely unlikely that the distribution of alleles in these two samples has occurred by chance.

We also considered the possibility that this difference is related not to the diagnosis of ADHD, but rather to population stratification. Indeed, one of the reasons we sequenced such a large worldwide sample22 was to address this issue. We constructed a series of comparison groups from our worldwide population sample. Each comparison group contained the 220 alleles from samples of European origin. Added to this was a random selection from the remaining non-European samples to approximate the ethnic distribution of the ADHD sample (Table 1). In all cases, the allele distribution differed significantly between the ADHD sample and the comparison group (P<0.0001). It is extremely unlikely, therefore, that population stratification and undetected ethnic bias can account for the distribution differences in our population and ADHD samples. We conclude, then, that the most likely reason for the observed differences was our ascertainment of this sample by diagnosis of ADHD, and that variants present at low frequency in the general population were ‘enriched’ in the ADHD sample.

Discussion

The increased frequency of the DRD4 7R allele in ADHD probands is consistent with the predictions of the CVCD hypothesis (Figure 3).9,12,24 By DNA resequencing from probands diagnosed with the refined phenotype of ADHD, we determined that the majority (82%) of 7R alleles in these individuals were of the common 7R(1-2-6-5-2-5-4) haplotype found previously22 (Table 2). However, we uncovered an unusually high prevalence (50%) of novel haplotypes in the 24 haplotypes observed in our sample, most 7R allele derivatives (Table 3). Greater than 10% of ADHD probands had one of these rare alleles. Including these rare derivatives (determined by sequence analysis) in the ‘7R’ class increased the number of ADHD individuals with 7R alleles from 43.2 to 47% (Tables 2 and 3). It is impossible to know without further biochemical/physiological/behavioral experimentation if these derivatives are functionally equivalent/related to 7R alleles (see below). It is likely, however, that all previous studies of the DRD4/ADHD association6 modestly underestimated the relative risk by only examining repeat length rather than DNA sequence.

What can account for the high frequency of novel alleles uncovered in the present study? If recombina-
molecule indicates our inability to estimate the number of genes associated with the disorder. Likewise, the double-headed arrows represent our inability to predict how these genes interact to produce the phenotype(s) depicted at left. Some fraction of the disorder may have a nongenetic cause (arbitrarily represented as 0.2 nongenetic in the diagram), for example brain damage in the case of ADHD.

We suggest, then, that both CVCD association and allelic heterogeneity (RVCD) contribute to the association of the DRD4 gene and ADHD (Figure 3). The finding of allelic heterogeneity for the DRD4/ADHD association should not be surprising, since ‘private’ mutations are found frequently for the majority of ‘single-hit’ genetic diseases, even ones where a particular variant is common. For example, while the common AF508 mutation is found in 70% of cystic fibrosis probands, hundreds of rarer mutations have also been identified. We suggest, then, that both CVCD association and allelic heterogeneity (RVCD) contribute to the association of the DRD4 gene and ADHD (Figure 3). The observation of increased allelic heterogeneity adds further support to the hypothesis that the DRD4 gene itself, rather than an adjacent variant in strong LD with DRD4, is responsible for the association.

While data exist indicating that DRD4 protein variants containing different VNTR lengths exhibit different biochemical properties, little is known of the effect of sequence (amino acid) differences in this region of the protein. The functional importance of changes at this position in the DRD4 protein, however, in a region that couples to G proteins and mediates postsynaptic effects, seems likely. For example, many of the observed changes are quite dramatic (ie, substituting a Pro for a Gln in 4R[1-2-6-8]).
Hypothesis testing. vs duals as 4R/4R type. Tests of this hypothesis would group individual haplotypes might lead to altered biochemistry/phenotype alteration from the conserved ancestral 4R(1-2-3-4) categories may not be appropriate divisions, and one selected for in individuals with a may predispose behaviors that are deemed inappropriate in the typical classroom setting and hence diagnosed as ADHD. This environmental mismatch hypothesis has testable predictions, including the potential benefit of altered educational approaches. In this hypothesis, the DRD4 7R subset of individuals diagnosed with ADHD are assumed to have a different, evolutionarily successful behavioral strategy rather than a disorder. Alternatively, we also speculated that DRD4 7R, while selected for in human populations, could have deleterious effects only when combined with other genetic variants. This complex genetic model for ADHD also has testable predictions. One of the many important questions stemming from this hypothesis is the number and nature of these interacting genes. Is DRD4 7R one of only a few (or a few hundred) predisposing alleles?

The DRD4 7R/ADHD association is one of the most reproduced in complex behavioral disorders. However, the approximately two-fold risk associated with the DRD4 7R allele and ADHD has been described as ‘small’. The implication is that DRD4 7R is but one of many predisposing alleles (a classic QTL), and indeed may be only a ‘modifier’ of yet undiscovered predisposing genes. Certainly, this is a possibility. However, while a two-fold risk may be considered small in some contexts, this risk needs to be put in the perspective of observed DRD4 allele frequencies and the predictions of the CVCD hypothesis (Figure 3).

In the populations of predominantly European ancestry used in most investigations of the DRD4/ADHD association, the allele frequency of DRD4 7R is approximately 12–15%. Therefore, even if the presence of a DRD4 7R allele was a necessary predisposing condition for ADHD (ie, 100% of ADHD probands had at least one copy of this allele), and assuming Hardy–Weinberg equilibrium, the increase in observed frequency (and relative risk) would be only 3.6-fold (Figure 3). If only half of ADHD is ‘caused’ by DRD4 7R, then the increase in observed frequency would be 1.8-fold. Common alleles associated with a particular disorder, then, can only exhibit modest increases in allele frequency in affected individuals, and hence have modest relative risks (ie, small λ). Most current genome scans of complex genetic disorders, including one for ADHD, would not have detected genomic regions with λ < 2–3.

Are λ values less than 2–3 of little significance? Do they imply that the associated allele has little impact on the disorder? On the contrary, they are exactly of the magnitude one expects if the CVCD hypothesis is correct. Likewise, the RVCD model also predicts modest relative risks, if one sums the contributions of all variants in a single gene (Figure 3). It is informative to propose a simple model for ADHD based on the CVCD hypothesis and the DRD4 7R association (Figure 4). Unlike rare disorders like Huntington Disease, where the disease allele is rare and the allelic relative risk is large (>5000 fold, Figure 4), what if alleles predisposing to ADHD are
absence affect the ‘penetrance’ of the other alleles. Described as ‘modifier’ alleles, since their presence or inclusion in our proband population (Figure 3).

types (example, approximately 5% of individuals would have a 2-4 fold relative risk, Figure 4), and none would likely be detected with genome scans of typical size. Yet according to this model, these are the predisposing alleles that are the object of our search. Similar conclusions could be reached for a variety of other likely models.

What can be concluded from such models? The observed two-fold increase in DRD4 7R allele frequency in ADHD probands is approximately 54% of the maximum possible (if all ADHD is genetic and related to DRD4 7R). As discussed above, this estimate modestly underestimates the relative risk, since rare 7R derivatives, as uncovered in this study, would not have been identified in prior work. The observed risk is approximately 87% of the maximum possible if 50% of ADHD has a nongenetic cause. If one assumes that ADHD predisposition is related to many different genes/alleles, such values for a single allele are, in fact, unusually high. We conclude, therefore, that the observed DRD4 7R allele/ADHD association is not ‘small’, but is of a magnitude quite surprisingly high. It suggests that this allele is associated with a minimum of 25–50% of the observed cases of ADHD. It further suggests that as few as one or two other common alleles in other genes, in combination with DRD4 7R (Figure 4), could account for most of the disorder.

Acknowledgements

This work was supported by NIMH Grant MH60660 to RKM.

References


