

# UC Irvine

## UC Irvine Previously Published Works

### Title

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

### Permalink

<https://escholarship.org/uc/item/4w93z8sw>

### Journal

Molecular Psychiatry, 8(5)

### ISSN

1359-4184

### Authors

Grady, DL

Chi, H-C

Ding, Y-C

et al.

### Publication Date

2003-05-01

### DOI

10.1038/sj.mp.4001350

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

## ORIGINAL RESEARCH ARTICLE

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

DL Grady<sup>1</sup>, H-C Chi<sup>1</sup>, Y-C Ding<sup>1</sup>, M Smith<sup>2</sup>, E Wang<sup>1</sup>, S Schuck<sup>3</sup>, P Flodman<sup>2</sup>, MA Spence<sup>2</sup>, JM Swanson<sup>3</sup> and RK Moyzis<sup>1,3</sup>

<sup>1</sup>Department of Biological Chemistry, College of Medicine, University of California, Irvine, CA, USA; <sup>2</sup>Department of Pediatrics, University of California, Irvine, Medical Center, Orange, CA, USA; <sup>3</sup>Child 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.<sup>1</sup> Current estimates indicate that 3–6% of school age children are diagnosed with ADHD, making it the most prevalent disorder of childhood.<sup>2</sup> While the broad DSM-IV<sup>1</sup> phenotype of ADHD almost certainly has multiple biological etiologies,<sup>3</sup> numerous family, twin and adoption studies have documented a strong genetic basis.<sup>4,5</sup> However, given high cross-national variation in the recognition and treatment of ADHD,<sup>2,3</sup> we proposed that the ADHD-combined type (DSM-IV) without serious comorbidity should be used as a 'refined' phenotype in biological and genetic research.<sup>6</sup>

Despite the high heritability of ADHD, initial genome scan studies have failed to identify genes of major effect,<sup>7</sup> although a region on chromosome 16p13 has been implicated in subsequent studies by the same group.<sup>8</sup> 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.<sup>9–13</sup> 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.<sup>2,14</sup> 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 sequence<sup>15–17</sup> has provided information sufficient to examine multiple candidate genes in parallel, often representing most of the proteins in a relevant biochemical pathway.

Correspondence: Dr RK Moyzis, Department of Biological Chemistry, College of Medicine, Sprague Hall, University of California, Irvine, CA 92697, USA. E-mail: rmoyzis@uci.edu

Received 11 September 2002; revised 20 November 2002; accepted 24 January 2003

One of these candidate genes, *DRD4*,<sup>18</sup> located near the telomere of chromosome 11p, is one of the most variable human genes known.<sup>19–22</sup> Most of this diversity is the result of length and single-nucleotide polymorphism (cSNP) variation in a 48bp 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.<sup>22</sup>

A number of investigations have found associations between particular alleles of this highly variable gene and behavioral phenotypes.<sup>2,3,5</sup> While some studies have suggested that the 7R allele of the *DRD4* gene might be associated with the personality trait of novelty seeking,<sup>23</sup> the most reproduced association is between the 7R allele and ADHD.<sup>2,3,5</sup> 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.<sup>22</sup> 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.<sup>22</sup> 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.<sup>22</sup>

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.<sup>22</sup> 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 ( $\lambda = 1.9$ ), calculated from a recent meta-analysis<sup>5</sup> is consistent with the common variant–common disorder (CVCD) hypothesis (also called the common disease–common variant hypothesis).<sup>9,12,24</sup> 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.<sup>22</sup> 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,<sup>25,26</sup> contributing to the observation of unexpectedly large blocks of LD in the human genome.<sup>27–30</sup>

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.<sup>10,11,24</sup> Alleles at low prevalence, most of which have not been identified by current SNP searches targeting a small sample size,<sup>31</sup> could also contribute to complex disease.<sup>10,32</sup> 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,<sup>22</sup> 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.<sup>3,6,33</sup> 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.<sup>22</sup> 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)<sup>34</sup> at the

University of California, Irvine. The refined phenotype of ADHD was diagnosed by a research assessment battery described in detail elsewhere,<sup>34,35</sup> 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.<sup>20,22</sup>

#### *PCR amplification and DNA sequencing*

The *DRD4* exon 3 VNTR was amplified with primer sets described previously (5'-CGTACTGTGCGGCCT-CAACGA-3' and 5'-GACACAGCGCCTGCCGTGATGT-3'; 705 nucleotide product for the 4R-allele;<sup>22</sup>). PCR reactions were conducted in 25  $\mu$ l volumes, containing 100 ng genomic DNA, 200  $\mu$ M dXTPs, 0.5  $\mu$ mol of each primer, 1  $\times$  PCR buffer (Qiagen), 1  $\times$  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  $\times$  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.<sup>22</sup> DNA cycle sequencing was conducted by standard techniques, using ABI 3100 and 3700 automated sequencers.<sup>16</sup> 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.<sup>22</sup> 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.<sup>36–39</sup> 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.<sup>39</sup>

#### *Statistical analysis*

Allele distributions were compared using Fischer's exact test for a 2  $\times$  k table, as implemented in SAS (v.6.12, running on a SUN Ultra2 Enterprise workstation). In our prior worldwide sample,<sup>22</sup> 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.<sup>2,3,33,34</sup> On the SNAP, average rating per item

summary scores of inattention and hyperactivity/impulsivity above 2.0 are considered severe.<sup>34,35</sup> 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.<sup>40</sup> 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.<sup>2,3,5,33</sup> Interestingly, 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*<sup>20</sup>; Ding *et al*<sup>22</sup> 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.<sup>22,33</sup> Adjusting these values for the increased frequency of 7R alleles in some non-European populations<sup>20,22</sup> 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 (Table 3) 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

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

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

**Table 2** Genotypes of 132 ADHD Probands

Genotype	2R/4R	3R/3R	3R/4R	4R/4R	4R/6R	2R/7R	3R/7R	4R/7R	6R/7R	7R/7R	4R/8R	4R/9R
Observed	20(19)	1	9(8)	43(41)	2(1)	5	3	42	1	4	1	1
Expected	14	<1	6	70	4	2	1	23	<1	2	1	<1

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).

**Table 3** Haplotypes of 250 *DRD4* exon3 alleles from 132 ADHD probands

Allele	N	Haplotype	Allele	N	Haplotype
2R	23	1-4	7R	55	
3R	14			45	1-2-6-5-2-5-4
	8	1-7-4		2	1-2-6-5-2-5-19
	3	1-9-4		2	<b>1-2-6-1-2-3-4</b>
	2	<b>1-2-20</b>		1	<b>1-2-6-5-2-23-4</b>
	1	<b>1-6-4</b>		1	<b>1-2-6-5-8-5-4</b>
4R	156			1	<b>1-2-14-5-2-5-4</b>
	150	1-2-3-4		1	1-2-3-17-2-5-4
	2	1-2-14-4		1	1-8-25-5-2-5-4
	2	<b>1-2-5-4</b>		1	1-2-6-5-2-3-4
	1	<b>1-2-6-4</b>	8R	1	<b>1-2-6-26-5-2-3-4</b>
	1	<b>1-26-3-4</b>	9R	1	<b>1-8-25-5-2-5-2-23-4</b>
6R	3				
	1	1-2-3-2-3-4			
	1	1-2-6-5-2-20			
	1	<b>1-2-6-2-5-4</b>			

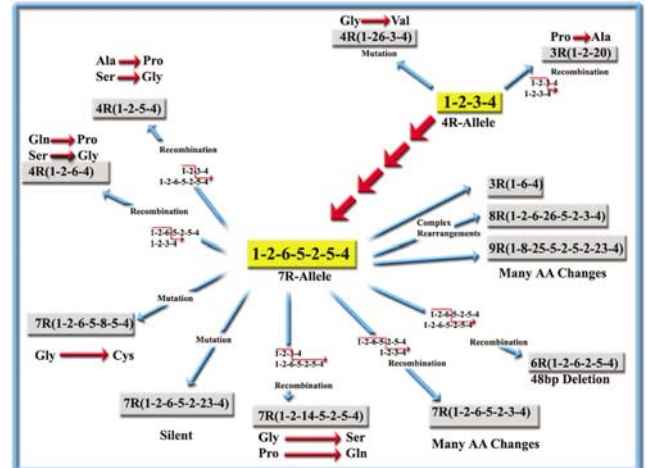
N: allele number identified by sequence analysis; haplotype nomenclature is described in Figure 1. Alleles in normal font were identified previously in a survey of 600 worldwide alleles.<sup>22</sup> Alleles in bold are unique to this study.

Repeat Variant	Nucleotide Sequence	Protein Sequence
1	A CCC GGG CCC CCC CTC CCC CAG GAC CCC TGC GGC CCC GAC TGT GCG CC	Pro Ala Pro Arg Leu Pro Gln Ser Pro Cys Gly Pro Asp Cys Ala Pro
2	C ... .. G...T ... .. G...GT	Gly ... .. Arg Gly
3	C G...A...A... ..T ... ..	Gln ... .. Ser ... ..
4	C ... .. G...C... ..T... ..T ... ..	Gly ... .. Pro ... .. Ser ... ..
5	C ... .. G... ..	Gly ... ..
6	C G...A...G... ..C... ..	Gln ... .. Ser ... ..
7	C ... .. G...T ... ..T ... ..	Gly ... ..
8	C ... .. T...T ... ..G...GT	... .. Ser ... .. Arg Gly
9	C ... .. G...T ... ..G...GT	Gly ... .. Ser ... .. Arg Gly
10	C ... .. A...T ... ..G...GT	... .. Ser ... .. Arg Gly
11	C ... .. A...Ser ... ..T ... ..	Gly ... .. Ser ... ..
12	C G...A...G... ..T ... ..	Gln ... .. Gly ... ..
13	C G...A...A... ..T ... ..	Gln ... .. Ala ... ..
14	C G...A...A... ..T ... ..	Gln ... .. Ala ... ..
15	C G...A...G... ..T ... ..	Gln ... .. Ala ... .. Arg ... ..
16	... .. G... ..	Gly ... ..
17	C ... .. G...C... ..T... ..T ... ..	Gly ... .. Pro ... .. Ser ... ..
18	C ... .. G...T ... ..	Gly ... ..
19	C ... ..T...Gly ... ..C... ..T... ..T ... ..	... .. Ser ... .. Pro ... .. Ser ... ..
20	C G...A...G... ..C... ..C... ..T... ..T ... ..	Gln ... .. Ala ... .. Ser ... .. Pro ... .. Ser ... ..
21	C G...A...A...Ser ... ..C... ..T... ..T ... ..	Gln ... .. Ala ... .. Ser ... .. Pro ... .. Ser ... ..
22	C G...A...G... ..T ... ..T... ..T ... ..	Gln ... .. Ala ... .. Gly ... .. Ser ... ..
23	C ... .. G...T ... ..T ... ..	... .. Gly ... ..
24	... .. G...T ... ..G...GT	... .. Gly ... .. Arg Gly
25	C G...A...A...Ser ... ..C... ..C... ..	Gln ... .. Ala ... .. Ser ... .. Pro ... ..
26	C ... .. G...T ... ..G...TT	... .. Gly ... .. Arg Val
27	C ... .. G...C... ..C... ..G ... ..T... ..T ... ..	Gly ... .. Pro ... .. Ser ... ..
28	C ... .. G...T ... ..C... ..G ... ..T... ..T ... ..	Gly ... .. Pro ... .. Ser ... ..
29	C G...A...A...Ser ... ..T ... ..	Gln ... .. Ala ... .. Ser ... ..
30	... ..	
31	C G...A...A... ..T ... ..	Gln ... .. Ala ... ..
32	C G...A...A... ..T ... ..	Gln ... .. Ala ... ..
33	C ... .. G...T ... ..C... ..T... ..T ... ..	Gly ... .. Pro ... .. Ser ... ..
34	T ... .. G...T ... ..G...GT	... .. Gly ... .. Arg Gly
35	C ... .. G...A... ..C... ..T... ..T ... ..	Gly ... .. Ala ... .. Ser ... ..

**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.

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 48bp 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.<sup>22</sup> The remaining nine alleles are rare 3R, 4R, 6R, and 7R variants observed previously<sup>22</sup> (Table 3).

The other half of the observed haplotypes were unique, not identified in our extensive prior analysis<sup>22</sup> (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



**Figure 2** Proposed origin of ADHD/*DRD4* allele diversity. A model for *DRD4* exon 3 repeat sequence diversity is shown. The ancestral 4R(1-2-3-4) and 7R(1-2-6-5-2-5-4) alleles are noted in yellow, with large red arrows indicating the multistep origin of the 7R-allele.<sup>22</sup> The proposed mutational or recombinational origins of the 12 novel alleles reported in this study are indicated along the blue arrows. Amino acid changes are also indicated. Haplotype nomenclature as described in Figure 1.

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,<sup>22</sup> 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). The 5 and 6 variant motifs (Figure 1) are diagnostic for the 7R allele, found only in this allele and its derivatives.<sup>22</sup> Ten of the 12 haplotype variants contain these motifs (Table 3), and hence likely arose as recombination/mutation 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 alleles<sup>22</sup> 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,<sup>22</sup> in which rare variants were found to be equally derived from 4R

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.<sup>22</sup> 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,<sup>22</sup> 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 al*<sup>22</sup>), 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 previously<sup>22</sup> 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.

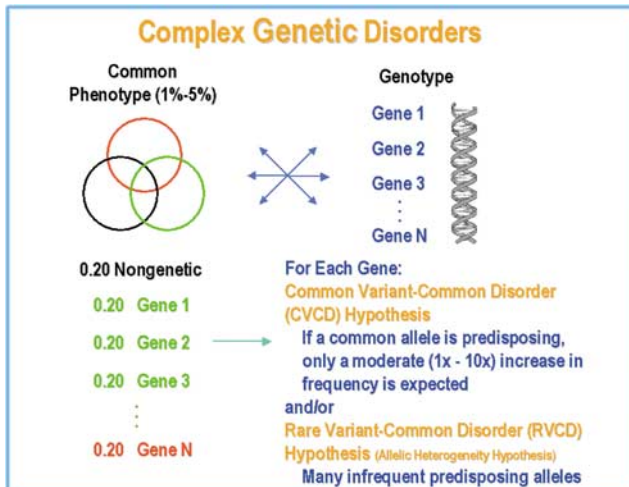
We also considered the possibility that this difference is related not to the diagnosis of ADHD, but rather to population stratification. Indeed, one of the reasons we sequenced such a large worldwide sample<sup>22</sup> was to address this issue. We constructed a series of comparison groups from our 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 ( $P \leq 0.0001$ ). 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).<sup>9,12,24</sup> 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 previously<sup>22</sup> (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 association<sup>5</sup> 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-





**Figure 3** A simplified diagram of complex genetic disorders. The left colored circles represent the potentially overlapping phenotypes classified together as a single disorder. In the current study, the refined phenotype of ADHD associated with a *DRD4* 7R allele is proposed to represent one of the circles. The Gene 1–Gene N displayed along the DNA 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.<sup>40</sup> The Genes 1–Gene N account for some fraction of the disorder (arbitrarily represented as 0.2 each in the diagram). Two widely discussed models for how genetic variants predispose to common disorders are shown, the Common Variant-Common Disorder (CVCD) hypothesis, and the Allelic Heterogeneity or Rare Variant-Common Disorder (RVCD) hypothesis.

tion/mutation was random, one would expect that the majority of derivative alleles would have 4R origins, since this is the most common allele, even in ADHD probands. The *DRD4* 4R allele is also older than the 7R allele,<sup>22</sup> and hence there has been greater time to accumulate mutations in this allele (unless they have been selected against). In our prior population study,<sup>22</sup> approximately equal numbers of 4R and 7R derivative alleles were uncovered, suggesting a mutation/recombination bias toward 7R alleles (or a stronger selection against 4R variants). In comparison with our prior population survey, however, over 90% of the rare derivative alleles in this ADHD sample have 7R origins (Figure 2).

We estimate that there are less than 85 *DRD4* alleles with population frequency greater than 0.001, and we have identified a minimum of 79% of these alleles. While there could be hundreds of extremely rare *DRD4* alleles (at a population frequency of 0.0001), such alleles could only contribute a few examples to our original population sample.<sup>22</sup> Therefore, given the sample sizes used in this and our prior population study, it is expected that, at most, two to three alleles

might be found only in one sample and not the other. It is extremely unlikely ( $P \ll 0.0001$ ), therefore, that finding 12 new alleles (on 15 chromosomes) in the ADHD population was due to chance or population stratification. We propose, then, that our ascertainment of the sample by diagnosis of ADHD was the reason for this observed increase in derivative *DRD4* 7R alleles.

Further studies, including more extensive population sampling, can refine the number and frequency distribution of rare *DRD4* alleles. In particular, it would be informative to know if rare *DRD4* alleles exhibit biased geographic/ethnic ancestry distributions. Such information would be essential for the design and interpretation of replicate studies of the current work. In addition, family-based analyses can help determine if rare alleles are preferentially transmitted to ADHD probands. However, for behavioral disorders such as ADHD, such studies should be interpreted with caution. It is common in such disorders to be unable to consent key members of a trio (mostly fathers). An inability to ascertain a truly 'random' sample of parental genotypes (for example, if there is preferential absence of a parent transmitting a putative predisposing gene) could contribute to biases in tests such as the TDT.<sup>41</sup>

The high frequency of amino-acid changing variants in these rare haplotypes (>90%) and the low probability that we uncovered these variants by chance ( $P \ll 0.0001$ ) suggest that allelic heterogeneity is also playing a role in the association of the *DRD4* gene and ADHD (RVCD Model, 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.<sup>10</sup> For example, while the common  $\Delta F508$  mutation is found in 70% of cystic fibrosis probands, hundreds of rarer mutations have also been identified.<sup>10,42</sup> There is no strong experimental or theoretical reason why genes associated with complex genetic disorders involving multiple genes should utilize a different mutational spectrum than genes for single-hit disorders.<sup>10–13,24</sup> 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,<sup>43,44</sup> 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,<sup>45</sup> seems likely. For example, many of the observed changes are quite dramatic (ie, substituting a Pro for a Gln in 4R(1-2-6-



4); Figure 2), and might be expected to alter the DRD4 protein structure/function. Clearly, further biochemical studies would be helpful. Such studies should be interpreted with caution, however. Observed biochemical differences<sup>43,44</sup> do not necessarily imply differences at the behavioral level. Many genetic/biochemical systems exhibit great buffering capacity, and biochemical variation often has little physiological effect.<sup>46</sup> Likewise, not finding biochemical differences between DRD4 variant proteins does not imply that functional differences do not exist at a behavioral level. It is often unclear which biochemical parameter is relevant to test, especially for proteins like DRD4, where most of the interacting proteins are as yet unknown.<sup>45</sup> Further, subtle biochemical changes, difficult to detect *in vitro* and *in vivo*, can have large effects at the organismal level.<sup>46</sup> The decade long search for the relevant biochemical basis of Huntington Disease, following the identification of the mutation, is but one recent example.<sup>47</sup> For these reasons, we suggest that genetic approaches will remain more powerful than biochemical approaches at detecting associations with behavioral disorders. We therefore suggest that in addition to further biochemical analysis of DRD4 variants, direct genotype/phenotype correlations continue to be pursued, including brain imaging<sup>14</sup> and model organism experiments.<sup>48</sup> It is the physiological/behavioral outcome of genetic variation that is most relevant. The finding that individuals with ADHD who possess a DRD4 7R allele perform normally on critical neuropsychological tests of attention in comparison with other ADHD probands<sup>40</sup> points to but one of many areas of future investigation.

Based on the current work and the hypothesized origin of human DRD4 diversity,<sup>22</sup> we suggest that future studies might group individuals based on DRD4 genotype differently than in the past. Only VNTR length was considered, usually split into 7R(+) and 7R(-) categories.<sup>5</sup> The DRD4 locus appears to behave like a 'two-allele' system (4R and 7R) under balanced selection.<sup>22</sup> The common 4R allele appears to be the ancestral allele, with the 7R allele being a much younger allele. All rare variants appear to be recombination/mutation products of these common 4R and 7R alleles (Figure 2 and Ding et al<sup>22</sup>). For example, the 2R allele likely has both a 4R and 7R origin.<sup>22</sup> Hence, simple 7R(+) and 7R(-) categories may not be appropriate divisions, and one should entertain other potential groupings. In particular, one might hypothesize that any amino-acid alteration from the conserved ancestral 4R(1-2-3-4) haplotype might lead to altered biochemistry/phenotype. Tests of this hypothesis would group individuals as 4R/4R vs non-4R/4R for purposes of hypothesis testing.

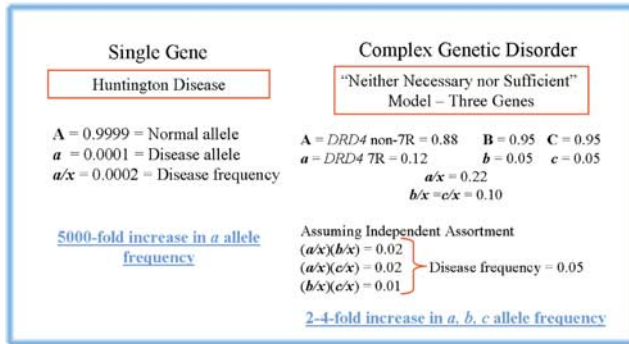
What does the DRD4/ADHD association mean? We have speculated that the very traits that may be selected for in individuals with a DRD4 7R allele may predispose behaviors that are deemed inap-

propriate in the typical classroom setting and hence diagnosed as ADHD.<sup>22</sup> This environmental mismatch hypothesis<sup>49</sup> 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.<sup>22</sup> 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.<sup>2,3,5</sup> However, the approximately two-fold risk associated with the DRD4 7R allele and ADHD has been described as 'small'.<sup>5,7</sup> The implication is that DRD4 7R is but one of many predisposing alleles (a classic QTL<sup>50</sup>), 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,<sup>40</sup> 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  $\lambda$ ).<sup>5,7</sup> Most current genome scans of complex genetic disorders,<sup>13,17</sup> including one for ADHD,<sup>7</sup> would not have detected genomic regions with  $\lambda < 2-3$ .

Are  $\lambda$  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,<sup>47</sup> where the disease allele is rare and the allelic relative risk is large (>5000 fold, Figure 4), what if alleles predisposing to ADHD are



**Figure 4** Contrast between rare single gene disorders and common complex genetic disorders. For single gene disorders, for example Huntington Disease 47 (left), predisposing alleles (indicated by  $a=0.0001$ ) and the disease frequency (indicated by  $a/x=0.0002$ ) are rare. Therefore, one observes a dramatic increase in allele frequency (and relative risk) in probands. For complex disorders related to common alleles, however, only modest increases in allele frequency (and relative risk) are expected. In the example shown (right), three predisposing alleles (*DRD4* 7R,b,c) in three different genes are hypothesized to interact. Each allele is proposed to be at polymorphic frequency in the population (0.05–0.12). Individuals with predisposing genotypes [(*DRD4* 7R/x)(b/x), (*DRD4* 7R/x)(c/x), (b/x)(c/x)] represent 0.05 of the population, the approximate frequency of ADHD.<sup>2,3,5</sup> The observed increase in alleles *DRD4* 7R,b, and c in probands ranges from 4-fold (if all cases are caused by these genes) to 2-fold (if only 50% of cases are caused by these genes). For example, a significant fraction of ADHD may have nongenetic causes<sup>40</sup>, yet these cases will be included in our proband population (Figure 3).

common in the population? Figure 4 outlines one such model, in which three different dominant alleles (designated *DRD4* 7R, b, c in three different genes) interact to predispose to the disorder. In this model, each of these alleles is at polymorphic frequency (0.05–0.12), and it is assumed that any two of them in combination predispose to ADHD. In such hypothetical interacting gene systems, any of the three ‘disease’ alleles (*DRD4* 7R, b, or c) could also be described as ‘modifier’ alleles, since their presence or absence affect the ‘penetrance’ of the other alleles.<sup>46</sup> Such interacting genetic systems should be common, since most gene products are part of multiprotein assemblies or biochemical pathways. Obviously, many other models could be proposed, involving recessive alleles, additional genes, etc. For example, each predisposing ‘allele’ could be many rare alleles (the RVCD model, Figure 3), which in total have a frequency of 0.05. However, the model proposed in Figure 4 is one of the simplest in which interacting alleles are neither necessary nor sufficient. In this example, approximately 5% of individuals would have one of the hypothesized predisposing genotypes ((*DRD4* 7R/x)(b/x), (*DRD4* 7R/x)(c/x), (b/x)(c/x)), approximately the observed incidence of ADHD (Figure 4). None of the predisposing alleles would be either necessary or sufficient to ‘cause’ ADHD.

None of the hypothetical predisposing alleles would have a high  $\lambda$  (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.<sup>40</sup> If one assumes that ADHD predisposition is related to many different genes/alleles,<sup>7,50</sup> 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

- 1 American Psychiatric Association. *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, 4th Edn. American Psychiatric Association: Washington, DC, 1994.
- 2 Swanson JM, Flodman P, Kennedy J, Spence MA, Moyzis R, Schuck S et al. Dopamine genes and ADHD. *Neurosci Behav Rev* 2000; **24**: 21–25.
- 3 Swanson J, Deutsch C, Cantwell D, Posner M, Kennedy JL, Barr CL et al. Genes and attention-deficit hyperactivity disorder. *Clin Neurosci Res* 2001; **1**: 207–216.
- 4 Faraone SV, Biederman J. Genetics of attention-deficit hyperactivity disorder. *Child Adolesc Clin North Am* 1994; **3**: 285–291.
- 5 Faraone SV, Doyle AE, Mick E, Biederman J. Meta-analysis of the association between the 7-repeat allele of the dopamine D4 receptor gene and attention deficit hyperactivity disorder. *Am J Psychiatry* 2001; **158**: 1052–1057.
- 6 Swanson JM, Sergeant JA, Taylor E, Sonuga-Barke EJS, Jensen Cantwell DP. Attention deficit disorder and hyperkinetic disorder. *Lancet* 1998; **351**: 429–433.
- 7 Fisher SE, Franks C, McCracken JT, McGough JJ, Marlov AJ, MacPhie IL et al. A genomewide scan for loci involved in attention-deficit/hyperactivity disorder. *Am J Hum Genet* 2002; **70**: 1183–1196.
- 8 Smalley SL, Kustanovich V, Minassian SL, Stone JL, Ogdie MN, McGough JJ et al. Genetic linkage of attention-deficit/hyperactivity disorder on chromosome 16p13, in a region implicated in autism. *Am J Hum Genet* 2002; **71**: 959–963.
- 9 Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.

- 10 Terwilliger JD, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* 1998; **9**: 578–594.
- 11 Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs? *Nat Genet* 2000; **26**: 151–157.
- 12 Zwick ME, Cutler DJ, Chakravarti A. Patterns of genetic variation in Mendelian and complex traits. *Ann Rev Genomics Hum Genet* 2000; **1**: 387–407.
- 13 Sklar P. Linkage analysis in psychiatric disorders: the emerging picture. *Ann Rev Genomics Hum Genet* 2002; **3**: 371–413.
- 14 Volkow ND, Wang GJ, Fowler JS, Logan J, Franceschi D, Maynard L *et al*. Relationship between blockade of dopamine transporters by oral methylphenidate and the increases in extracellular dopamine: therapeutic implications. *Synapse* 2002; **43**: 181–187.
- 15 International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001; **409**: 860–921.
- 16 Riethman HC, Xiang Z, Paul S, Morse E, Hu X-L, Flint J *et al*. Integration of telomeric sequences with the draft human genome sequence. *Nature* 2001; **409**: 948–951.
- 17 Cowan WM, Kopnisky KL, Hyman SE. The Human Genome Project and its impact on psychiatry. *Annu Rev Neurosci* 2002; **25**: 1–50.
- 18 Van Tol HHM, Bunzow JR, Guan H-C, Sunahara RK, Seeman P, Niznik HB *et al*. Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine. *Nature* 1991; **350**: 610–614.
- 19 Lichter JB, Barr CL, Kennedy JL, VanTol HHM, Kidd KK, Livak KJ. A hypervariable segment in the human dopamine receptor D4 (DRD4) gene. *Hum Mol Genet* 1993; **2**: 767–773.
- 20 Chang F-M, Kidd JR, Livak KJ, Pakstis AJ, Kidd KK. The worldwide distribution of allele frequencies at the human dopamine D4 receptor locus. *Hum Genet* 1996; **98**: 91–101.
- 21 Ding Y-C, Wooding S, Harpending HC, Chi H-C, Li H-P, Fu Y-X *et al*. Population structure and history in East Asia. *Proc Natl Acad Sci USA* 2000; **97**: 14003–14006.
- 22 Ding Y-C, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK *et al*. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc Natl Acad Sci USA* 2002; **99**: 309–314.
- 23 Klugar AN, Siegfried Z, Ebstein RP. A meta-analysis of the association between DRD4 polymorphism and novelty seeking. *Mol Psychiatry* 2002; **7**: 712–717.
- 24 Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001; **17**: 502–510.
- 25 Harpending H, Rogers A. Genetic perspectives on human origins and differentiation. *Ann Rev Genomics Hum Genet* 2000; **1**: 361–385.
- 26 Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G *et al*. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* 2001; **293**: 455–462.
- 27 Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ *et al*. Linkage disequilibrium in the human genome. *Nature* 2001; **411**: 199–204.
- 28 Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; **29**: 229–232.
- 29 Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR *et al*. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 2001; **294**: 1719–1723.
- 30 Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF *et al*. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; **419**: 832–837.
- 31 The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2002; **409**: 928–933.
- 32 Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001; **69**: 124–137.
- 33 La Hoste GJ, Swanson JM, Wigal SB, Glabe C, Wigal T, King N *et al*. Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder. *Mol Psychiatry* 1996; **1**: 21–24.
- 34 MTA Cooperative Group. A 14-month randomized clinical trial of treatment strategies for attention deficit/hyperactivity disorder. *Arch Gen Psychiatry* 1999; **56**: 1073–1086.
- 35 Hinshaw SP, March JS, Abikoff H, Arnold LE, Cantwell DP, Conners CK *et al*. Comprehensive assessment of childhood attention-deficit hyperactivity disorder in the context of a multisite, multimodel clinical trial. *J Attention Disord* 1997; **1**: 217–234.
- 36 Ewing B, Hiller L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred I. Accuracy assessment. *Genome Res* 1998; **8**: 175–185.
- 37 Ewing B, Green P. Base-calling of automated sequencer traces using phred II. Error probabilities. *Genome Res* 1998; **8**: 186–194.
- 38 Nickerson DA, Tobe VO, Taylor SL. Polyphred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 1997; **14**: 2745–2751.
- 39 Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J *et al*. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 1998; **19**: 233–240.
- 40 Swanson JM, Oosterlaan J, Murias M, Schuck S, Flodman P, Spence MA *et al*. Attention deficit/hyperactivity disorder children with a 7-repeat allele of the dopamine receptor D4 gene have extreme behavior but normal performance on critical neuropsychological tests of attention. *Proc Natl Acad Sci USA* 2000; **97**: 4754–4759.
- 41 West A, Langley K, Hamshere ML, Kent L, Craddock N and Owen MT. Evidence to suggest biased phenotypes in children with attention deficit hyperactivity disorder from completely ascertained trios. *Mol Psychiatry* 2002; **7**: 962–966.
- 42 Serre JL, Simon-Bouy B, Mornet E, Jaume-Roig B, Balassopoulou A, Schwarz M *et al*. Studies of RFLP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations in population genetics. *Hum Genet* 1990; **84**: 449–454.
- 43 Asghari V, Sanyal S, Buchwaldt S, Paterson A, Jovanovic V, VanTol HHM. Modulation of intercellular cyclic AMP levels by different human dopamine D4 receptor variants. *J Neurochem* 1995; **65**: 1157–1165.
- 44 Jovanovic V, Guan H-C, VanTol HHM. Comparative pharmacological and functional analysis of the human dopamine D4.2 and D4.10 receptor variants. *Pharmacogenetics* 1999; **9**: 561–568.
- 45 Civelli O, Bunzow JR, Grandy DK. Molecular diversity of the dopamine receptors. *Annu Rev Pharmacol Toxicol* 1993; **32**: 281–307.
- 46 Hartman JL, Garvik B, Hartwell L. Principles for the buffering of genetic variation. *Science* 2001; **291**: 1001–1004.
- 47 Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is unstable on Huntington's Disease chromosomes. *Cell* 1993; **72**: 971–983.
- 48 Dulawa SC, Grandy DK, Low MJ, Paulas MP, Geyer MA. Dopamine D4 receptor-knock-out mice exhibit reduced exploration of novel stimuli. *J Neurosci* 1999; **19**: 9550–9556.
- 49 Jensen PS, Mrazek D, Knapp PK, Steinberg L, Pfeffer C, Schowalter J *et al*. Evolution and revolution in child psychiatry: ADHD as a disorder of adaptation. *J Am Acad Child Adolesc Psychiatry* 1997; **36**: 1672–1679.
- 50 Lynch M, Walsh B. *Genetic analysis of Quantitative traits*. Sinauer Associates, Inc: Sunderland, MA, USA, 1998.