Molecular Psychiatry (2004), 1–12

www.nature.com/mp

ORIGINAL RESEARCH ARTICLE

COMT haplotypes suggest P2 promoter region relevance for schizophrenia

MA Palmatier[1], AJ Pakstis[1], W Speed[1], P Paschou[1], D Goldman[2], A Odunsi[3], F Okonofua[4], S Kajuna[5], N Karoma[5], S Kungulilo[6], E Grigorenko[7], OV Zhukova[8], B Bonne-Tamir[9], R-B Lu[10], J Parnas[11], JR Kidd[1], MMC DeMille[1] and KK Kidd[1]

[1]*Department of Genetics, Yale University School of Medicine, New Haven, CT, USA;* [2]*Neurogenetics Laboratory, NIAAA, Rockville, MD, USA;* [3]*Department of Gynecological Oncology, Roswell Park Cancer Institute, Buffalo, NY, USA;* [4]*Department of Obstetrics and Gynecology, Faculty of Medicine, University of Benin, Benin City, Nigeria;* [5]*The Hubert Kairuki Memorial University, Dar es Salaam, Tanzania;* [6]*Muhimbili University College of Health Sciences, Dar es Salaam, Tanzania;* [7]*Child Study Center and Department of Psychology, Yale University, New Haven, CT, USA;* [8]*NI Vavilov Institute of General Genetics RAS, Moscow, Russia;* [9]*Department of Genetics, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel;* [10]*Department of Psychiatry, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, ROC;* [11]*Copenhagen University, Psychiatry Department, Hvidovre Hospital, Hvidovre, Denmark*

A recent study found, in a large sample of Ashkenazi Jews, a highly significant association between schizophrenia and a particular haplotype of three polymorphic sites in the catechol-*O*-methyl transferase, *COMT*, gene: an IVS 1 SNP (dbSNP rs737865), the exon 4 functional SNP (Val158Met, dbSNP rs165688), and a downstream SNP (dbSNP rs165599). Subsequently, this haplotype was shown to be associated with lower levels of *COMT* cDNA derived from normal cortical brain tissue, most likely due to *cis*-acting element(s). As a first step toward evaluating whether this haplotype may be relevant to schizophrenia in populations other than Ashkenazi Jews, we have studied this haplotype in 38 populations representing all major regions of the world. Adding to our previous data on four polymorphic sites in the *COMT* gene, including the Val158Met polymorphism, we have typed the IVS 1 rs737865 and 3′ rs615599 sites and also included a novel IVS 1 indel polymorphism, yielding seven-site haplotype frequencies for normal individuals in the 38 globally distributed populations, including a sample of Ashkenazi Jews. We report that the schizophrenia-associated haplotype is significantly heterogeneous in populations worldwide. The three-site, schizophrenia-associated haplotype frequencies range from 0% in South America to 37.1% in Southwest Asia, despite the fact that schizophrenia occurs at roughly equal frequency around the world. Assuming that the published associations found between the exon 4 Val158Met SNP and schizophrenia are due to linkage disequilibrium, these new haplotype data support the hypothesis of a relevant *cis* variant linked to the rs737865 site, possibly just upstream in the P2 promoter driving transcription of the predominant form of COMT in the brain. The previously described *Hind*III restriction site polymorphism, located within the P2 promoter, varies within all populations and may provide essential information in future studies of schizophrenia.
Molecular Psychiatry advance online publication, 6 April 2004; doi:10.1038/sj.mp.4001496

Keywords: COMT; haplotypes; schizophrenia; linkage disequilibrium; functional variation; P2 promoter; population variation

Several lines of evidence suggest that *COMT* is a strong candidate gene for susceptibility to schizophrenia.[1–9] Despite the genetic linkage and association studies that have found evidence for a schizophrenia susceptibility locus on chromosome 22 in the q11.1 region containing the COMT gene, the chromosome 22 results have been inconsistent,[10–14] though recently strengthened by a sib pair study by Williams *et al*.[15]

Correspondence: Dr KK Kidd, Department of Genetics, Yale University School of Medicine, New Haven, CT 06520-8005, USA. E-mail: Kidd@biomed.med.yale.edu
Received 04 August 2003; revised 31 December 2003; accepted 28 January 2004

The reasons for the difficulty in reproducing these results could include small sample size and/or low relative risk (both leading to low power to detect linkage or association), ethnically heterogeneous samples, and differences in the expression of neurocognitive deficits in individuals with schizophrenia, all in addition to the likely existence of genetic heterogeneity and/or oligogenic inheritance of susceptibility. To reduce the likelihood of low statistical power and to assure ethnic homogeneity, Shifman *et al*[1] studied an exceptionally large patient and control sample of a homogeneous population of Ashkenazi Jews and found a highly significant association of schizophrenia and a particular haplotype of the *COMT* gene. Their findings involved the G–G–G

haplotype at the IVS 1 rs737865, exon 4 Val158Met, and downstream rs165599 polymorphic sites. The associations with schizophrenia shown by the noncoding SNPs were more significant than the association shown by the exon 4 Val158Met site; the association shown by the haplotype was even more significant. These results suggest that other sites in the gene or nearby may contribute to the susceptibility. In particular, regulatory region variation may be involved.

The COMT protein exists in two forms, a soluble form (S-COMT) translated from a short mRNA transcribed by the P1 promoter in intron 2 of the *COMT* gene, and a membrane-bound form (MB-COMT) translated from a long mRNA transcribed by the P2 promoter 5′ of exon 1 (Figure 1).[16–18] In the brain, the predominant COMT protein is MB-COMT.[17,18]

Bray *et al*[2] have recently shown that the two chromosomes are not equally represented in cDNA made from brain mRNA in individuals selected for heterozygosity at various SNPs. Their finding strongly suggests there are *cis*-acting factors operating on transcription and/or mRNA stability. Bray *et al*[2] conclude from the studies of rs165599 heterozygotes that 'it is either SNP rs165599 itself responsible for the observed differences [in expression] or one that is in very strong LD with it.' However, in their studies of samples selected for heterozygosity at multiple markers, the individuals probably heterozygous for the Shifman schizophrenia-associated haplotype had the greatest average differential expression, while those heterozygous at the exon 4 Val158Met plus rs737865 sites had a larger and more significant difference in expression than those heterozygous at the exon 4 Val158Met and rs165599 sites.[2] This emphasizes the importance of the haplotype identifying a subset of chromosomes with low expression and association with schizophrenia.

Our previous population genetic studies of the Val158Met site showed that it was polymorphic in all populations studied, but with quite varying allele frequencies depending on which population from which region of the world was considered.[19] Haplotype studies incorporating additional markers at *COMT*, including one in the P2 promoter (a *Hind*III site), further showed considerable variation in frequencies around the world.[3] However, those haplotype studies did not include two of the noncoding SNPs studied by Shifman *et al*[1] and Bray *et al*.[2] Thus, whether the G–G–G haplotype even exists in other populations was largely unknown. Moreover, the mRNA expression studies[2] showed that the strongest haplotype association involved an intron 1 SNP rather than a 3′ SNP. This places increased emphasis on the P2 promoter, the one most important for COMT expression in the brain. Other studies involving the P2 *Hind*III SNP have not found an association with schizophrenia or not found differences in total COMT expression by genotype but different ethnic groups, small numbers of subjects, and only measures of total, not allele-specific, expression confound interpretation.[4,20,21] Thus, the P2 promoter polymorphism previously described[3] is a candidate for the relevant regulatory variation or at least a surrogate marker for an etiological variant within the P2 promoter.

As a first step toward evaluating whether this haplotype and the P2 promoter polymorphism may be relevant to schizophrenia in populations other than Ashkenazi Jews, we have undertaken to determine haplotype frequencies in diverse populations.

Materials and methods

As an extension of our earlier studies of *COMT*, we have now added three sites to our four other polymorphic sites previously studied (see Figure 1) in order to include all three SNPs of the schizophrenia-associated haplotype.[3,19] Data on all seven sites are now reported on a total of 2058 individuals from 38 populations. Links to descriptions of the population samples can be found in ALFRED (the web-accessible ALlele FREquency Database, see Electronic-Database Information) associated with the allele and haplotype frequencies. Population samples analyzed for the first time for any of the *COMT* sites in this study are the Ashkenazi Jews, Chagga, Chuvash, Khanty, Komi Zyriane, and Samaritans. All studies were approved by the Yale University Human Investigations Committee and all samples were collected with informed consent. Previously studied polymorphisms were typed as described.[1,3] The new IVS 1 InDel was typed by PCR followed by gel separation, and the rs737865 and rs165599 SNPs were typed by RFLP protocols. Typing protocols, as well as further details for all sites, and the allele and

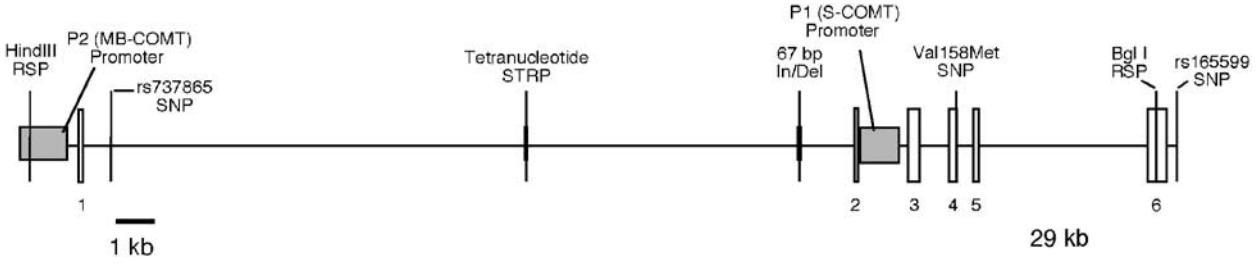


Figure 1 Map of the COMT Gene with polymorphisms.

haplotype frequencies at each site for all populations can be found in ALFRED (see Table 1 for ALFRED site UIDs; the seven-site haplotype UID is SI000939V).

As originally described in Iyengar *et al*,[22] we used the primers for the human polymorphisms to amplify the homologous regions of DNA from at least one individual from at least three of the following species; chimp, bonobo, gorilla, orangutan, and gibbon. The products were then digested and/or sequenced to reveal the nucleotides corresponding to the presence or absence of the sites polymorphic in humans and thus, by inference, which human allele is the ancestral allele.

Haplotype frequency estimates were calculated using HAPLO, which implements the expectation-maximization (EM) algorithm.[23] These estimates were used to calculate three LD measures—the standardized pairwise disequilibrium statistic D', the pairwise ξ, as well as Δ^2, which is sometimes called r^2.[24–26] The significance of the non-randomness among alleles (ie, of linkage disequilibrium) was evaluated with a permutation test.[25]

Results

All sites were polymorphic in all populations, except the IVS 1 rs737865 site in the South American populations. Genotypes existed in the expected Hardy–Weinberg ratios. Regional average heterozygosities are given in Table 1. The allele frequencies for each biallelic site, the four SNPs, two InDels, and the recoded IVS 1 STRP are graphically depicted in Figure 2 and showed large differences among populations. The numerical allele frequencies at each site for all populations can be found in ALFRED under the site UIDs in Table 1. The F_{st} values for the seven biallelic sites were within the empirical distribution of F_{st} values that we have determined for 172 sites.[27]

Ancestral alleles for the four SNP and two InDel sites are given in Table 2, along with the allele symbol synonyms and allele descriptions. At four of these sites only a single allelic variant was seen in our non-human primate samples. However, at the IVS 1 InDel and the rs16559 sites more than one allele was seen. All of the alleles seen in non-human primates at the IVS 1 InDel contained one or more copies of the 67 bp inserted sequence; therefore, the 1 (inserted) allele is the ancestral allele. For the rs165599 site, most of the non-human primates sequenced were homozygous for the G allele, including one gibbon. However, one orangutan was homozygous for the A allele, and two bonobos were heterozygous. Nonetheless, the ancestral allele appears to be the G allele.

The population frequencies of the three-site schizophrenia-associated haplotype vary widely around the world (Figure 3a). We note that the allele and haplotype frequencies in our sample of Ashkenazi Jews are similar, for the three sites in common, to those found in the Ashkenazi control sample in Shifman *et al*.[1] The frequencies of the other seven haplotypes defined by these three sites also vary significantly around the world, as do the frequencies of haplotypes defined by all seven 'biallelic' sites (data in ALFRED, site UID SI000939V). There are only 37 haplotypes, out of 128 possible, that occur at $>5\%$ in more than one population (106 have non-zero frequency estimates). Of these 37 haplotypes, only four contain the – G – – G – G haplotype studied by Shifman *et al*.[1] (The – is used to indicate the unspecified allele of a polymorphic site. Sites are ordered along the gene, beginning with the promoter *Hind*III site.) This haplotype is greatly subdivided by the other sites in all the populations in which the – G – – G – G haplotype was estimated to occur (Figure 3b).

Table 1 Heterozygosities of the seven 'biallelic' polymorphic sites

	Minimum 2N	*Promoter* Hind*III RSP*[a]	*IVS1 rs737865 SNP*	*IVS1 recoded STRP*[b]	*IVS1 InDel*	*Exon 4 Val158Met SNP*	*3′ UTR* Bgl*I RSP*	*Downstream rs165599*
ALFRED site UID		SI000268Q	SI000865T	SI000269R	SI000726P	SI000155L	SI000193N	SI000866U
Africa	662	0.44	0.22	0.41	0.37	0.37	0.45	0.35
Southwest Asia	298	0.45	0.47	0.47	0.31	0.48	0.44	0.49
Europe	814	0.47	0.35	0.47	0.35	0.49	0.40	0.45
Northwest Asia	190	0.42	0.46	0.22	0.38	0.48	0.43	0.44
East Asia	560	0.41	0.31	0.42	0.46	0.34	0.46	0.44
Pacific and Siberia	216	0.35	0.38	0.35	0.48	0.35	0.49	0.50
North and Central America	400	0.39	0.10	0.42	0.44	0.42	0.48	0.43
South America	320	0.27	0.00	0.28	0.27	0.27	0.28	0.17

[a]This SNP was also described by Li *et al* (2000) in a paper and erratum.[4,5]

[b]In this paper, we have recoded and analyzed the *COMT* tetranucleotide repeat in intron 1 as a biallelic system. With the exception of three alleles found in four individuals in the Chagga, the alleles of this STRP can be divided into primary ladder alleles that evenly follow the 4 base pair (bp) TAAA repeat unit and secondary ladder alleles that consist of 4 bp repeat units and a final 3 bp repeat unit that lacks the last A of the repeat. The three tertiary ladder alleles seen in the Chagga are of unknown sequence. Primary on-ladder alleles are combined as Allele O and all secondary and tertiary non-ladder alleles are combined as Allele N.

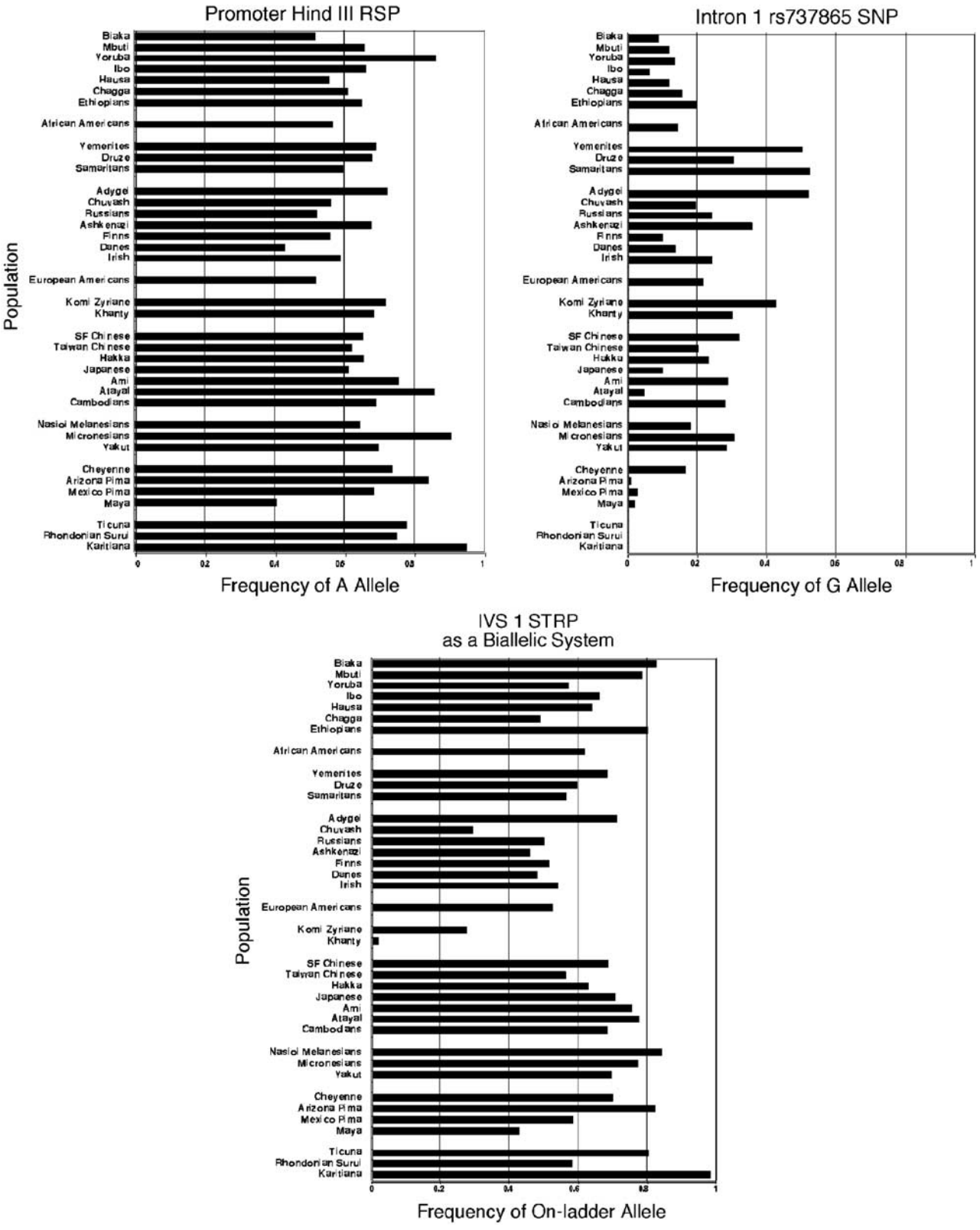


Figure 2 COMT allele frequencies in 38 human populations. The frequencies given are for the ancestral alleles at five of the seven polymorphic sites. The ancestral allele of the rs737865 site is the A allele. We chose to display the frequencies of the G allele because it is the allele present in the schizophrenia-associated haplotype of Shifman *et al*.[1] The ancestral allele of the IVS1 STRP was not determined and the frequencies of the on-ladder allele are given for this site. Populations are grouped by region as in Table 1.

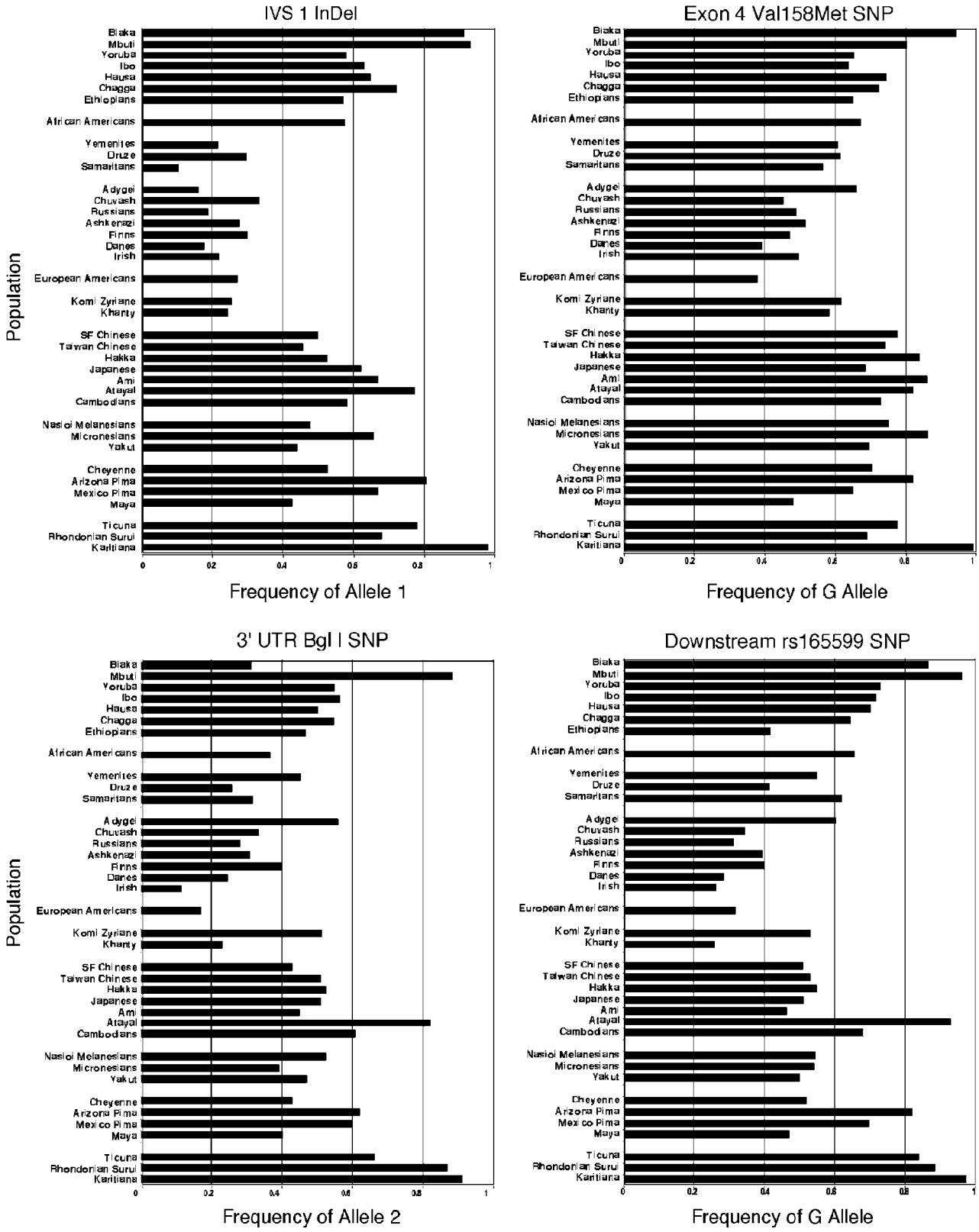


Figure 2 (*continued*)

Linkage disequilibrium between pairs of biallelic sites was significant between many of the sites in many populations (Table 3, Figure 4). Shifman *et al*[1] found significant LD ($D' = 0.85$) between the IVS 1 rs737865 SNP and the exon 4 Val158Met SNP in Ashkenazi Jews. We found essentially the same LD

Table 2 Allele symbol synonyms, allele descriptions and ancestral alleles

Polymorphic site	*Allele symbol synonyms*	*Allele description*	*Ancestral allele*
Promoter *Hind*III, rs2097603	A	Site absent	A
	G	Site present	
IVS1 rs737865	A	Site absent	A
	G	Site present	
IVS1 recoded STRP	O	On-ladder alleles	nd
	N	Non-ladder alleles	
IVS1 InDel	+ = 67 bp, present = 1	'Inserted'	1
	− = 67 bp, absent = 2	'Deleted'	
Ex4 *Nla*III (Ex 4 Val158Met), rs4680	G	Valine, high-activity allele, site absent	G
	A	Methionine, low-activity allele, site present	
3′ UTR *Bgl*I, rs362204	1	Inserted C absent, site absent	
	2	Inserted C present, site present	2
IVS1 rs165599	A	Site absent	
	G	Site present	G

($D' = 0.87$, $\xi = 0.49$, $\Delta^2 = 0.42$, $P \leq 0.01$) in our sample of Ashkenazi Jews.

Discussion

Since Shifman *et al*[1] found significant associations individually between both the IVS 1 rs737865 and 3′ rs165599 sites and schizophrenia, but very weak association between the exon 4 Val158Met SNP and schizophrenia, they suggest that positive associations between the Val158Met SNP and schizophrenia are due to LD. The expression studies of Bray *et al*[2] shift the focus strongly toward the P2 promoter governing transcription of the mRNA found primarily in the brain. Bray *et al*[2] do not specify the ethnicity of the cortical brain samples from which they derived the COMT cDNA for allelic expression studies, although they are probably northern/western European. Their haplotype frequencies do not match exactly any single population. The frequencies are similar to those of the European or Southwest Asian populations we have studied, and quite different from those of African, East Asian, and Native American populations (data in ALFRED). We find that the IVS 1 rs737865 site is fixed or nearly fixed in three populations (Figure 2). In 14 out of the remaining 35 population samples, including the Ashkenazi Jews, the exon 4 Val158Met SNP showed some LD with the IVS 1 rs737865 SNP, suggesting that, in these populations, any positive association between schizophrenia and the functional polymorphism could be due to its association with the IVS 1 rs737865 SNP. However, in most of our population samples, including all East Asian and Native American populations, there is no significant LD between the exon 4 Val158Met site and the IVS 1 rs737865 site; so LD cannot explain positive associations.

The IVS 1 rs737865 SNP does not have any evident functional relevance, but the *Hind*III SNP, only 2.2 kb upstream, is within the P2 promoter and could have functional relevance itself, and at least is a good surrogate marker for any variation within the P2 promoter, since the P2 promoter drives transcription of the membrane-bound form of COMT, the form that is more highly expressed in brain. The obvious question arises as to whether the association with schizophrenia seen by Shifman *et al*[1] and the expression level association seen by Bray *et al*[2] with the IVS 1 rs737865 SNP reflect LD with the promoter *Hind*III SNP. Both the ξ and Δ^2 pairwise measures of LD show weak to moderate, though often highly significant, LD between the promoter *Hind*III and IVS 1 rs737865 sites in 22 populations at $P < 0.05$, including 16 populations at $P < 0.01$, out of the 35 populations in which it could be calculated (Figure 4). All the significant associations have the G allele of the IVS 1 rs737865 site associated with the A allele of the promoter *Hind*III site. Though not significant, in the Nasioi and Mexican Pima the opposite pattern exists, with the two A alleles preferentially together. Such differences could occasionally occur by chance, especially as the LD is not strong between these two sites.[28] As the *Hind*III and IVS 1 rs737865 sites are not in complete LD, they will not give the same association results with another marker or trait. Thus, any extrapolation as to an association of the promoter *Hind*III SNP with schizophrenia is not warranted by the LD values themselves.

Linkage disequilibrium statistics, however, do not fully explain the results of the expression study; the associations for specific alleles at two sites are not symmetric, as shown by conditional probabilities.[2,3] As the conditional probabilities in Table 4 show, if one selects chromosomes with the G allele at the IVS 1 rs737865 site, the probability of having the A allele at the *Hind*III site is 1.0 or nearly so for almost all populations. In contrast, of the chromosomes with the A allele at IVS 1 rs737865 SNP, there will be many that also have the A allele at the *Hind*III site, as well as many with the G allele. Thus, if the *Hind*III

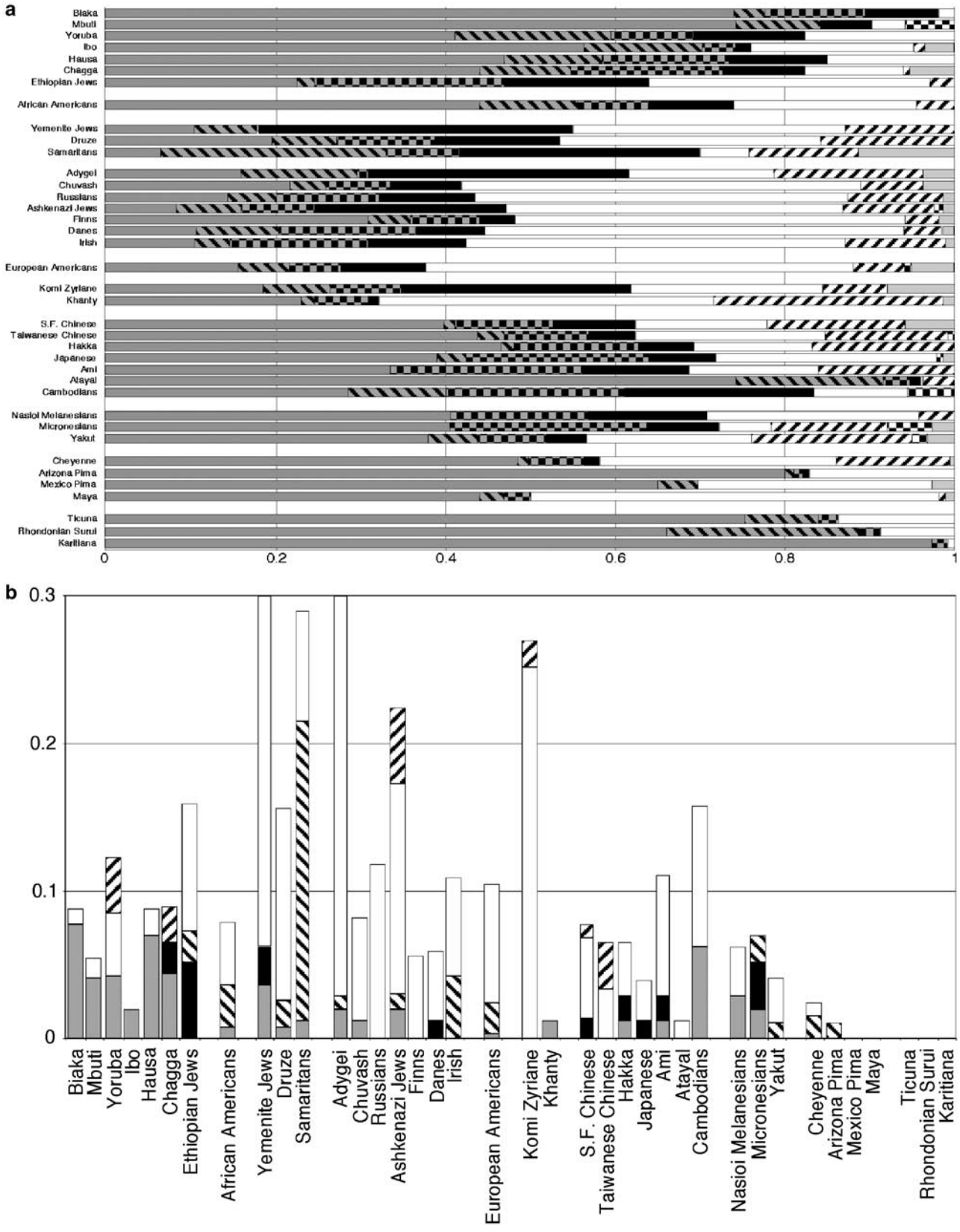
promoter site, or one in complete LD, is the site controlling expression, these data predict that virtually all of the chromosomes with the G allele at the IVS 1 rs737865 site will have low expression of COMT, while chromosomes with the A allele at the IVS 1 rs737865 site will be a mixture, varying among populations, of high and low expression of COMT. The three-site schizophrenia-associated haplotype of Shifman *et al*[1] is only associated with a single P2 promoter *Hind*III allele. Interestingly, choosing haplotypes for the G allele at the 3′ rs165599 SNP does not yield such homogeneity (data not shown). The conditional probabilities of the A allele at the *Hind*III site given the G allele at the 3′ rs165599 site range from 0.976 (in Yemenite Jews) to 0.490 (in Samaritans). The Ashkenazi Jews have one of the higher values, 0.860. Therefore, it would not be too surprising if this site showed a stronger association with schizophrenia in a particular study of this population, since a slight change in allele and haplotype frequencies could alter the relative association levels of the IVS 1 and 3′ SNPs with the promoter *Hind*III SNP.

The G allele at IVS 1 rs737865 is the derived allele and its strong association with only one of the P2 promoter alleles is consistent with its being more recent than the promoter polymorphism. The ubiquity of the IVS 1 rs737865 polymorphism argues that, although relatively more recent, its origin still predates expansion of modern humans out of Africa. This G allele is similarly more strongly associated with the exon 4 Val158Met G allele, also consistent with the more recent origin of the rs737865 polymorphism. Thus, the rs737865 G allele flags a more homogeneous set of haplotypes.

Several studies provide weak evidence against the *Hind*III SNP being the functionally relevant one. The Li *et al* study[4] of Chinese families with schizophrenia showed highly significant overtransmission of a haplotype with a specific *Hind*III allele ('G'), but also significant undertransmission of a different haplotype with the same *Hind*III allele. Considering only the P2 promoter *Hind*III site, they found nonsignificant preferential transmission of the 'A' allele. Tunbridge *et al*[21] found no difference in *COMT* mRNA abundance between schizophrenic and control dorsolateral prefrontal cortex tissue and no relationship to genotype at either the *Hind*III P2 SNP or the exon 4 Val158Met site in their small sample of unspecified ethnicity (15 controls, 15 schizophrenics). They did not analyze the data as haplotypes. Matsumoto *et al*[20] studied an ethnically mixed sample of mostly African Americans and found no association with genotype at the Val158Met SNP, but a differential laminar pattern of *COMT* mRNA expression in pyramidal neurons: schizophrenic patients had higher *COMT* mRNA abundance in intermediate/deep layers of the dorsolateral prefrontal cortex, and relatively lower levels of *COMT* mRNA in the superficial layers. Matsumoto *et al*[20] did not genotype the other two markers of the schizophrenia-associated haplotype, which is generally uncommon in African Americans (Figure 3), and so it is not clear that their data are relevant to the findings of Shifman *et al*.[1] These various studies have involved different ethnic groups, sometimes heterogeneous or unspecified, with different allele frequencies and different levels of linkage disequilibrium; some of the samples have been small and had low power. Nonetheless, the Li *et al*[4] and Tunbridge *et al*[21] data argue that the P2 promoter *HindIII* SNP is not functionally relevant; however, given the Matsumoto *et al* data,[20] it is possible that the schizophrenia-associated haplotype is in LD with another genetic polymorphism which affects COMT expression in a tissue-specific manner.

DeMille *et al*[3] found very little LD between the promoter *Hind*III site and the exon 4 Val158Met site (cf. Table 3 in DeMille *et al*[3]). The addition of six new populations in this report does not change that finding (Table 3, a supplement to the relevant part of Table 3 in DeMille *et al*[3]). Moderate, but significant ($P \leq 0.01$), LD between the promoter *Hind*III and the exon 4 Val158Met sites exists in only seven out of 38 populations, the Yemenite Jews, Khanty, Taiwanese Chinese, Cheyenne, Arizona Pima, Maya, and Ticuna. Thus, functional variation in the promoter region may be distributed in populations largely independently of the Val158Met site. It is possible that the risk of schizophrenia is truly 'epistatic' in that both specific regulatory variation and the Val158Met activity difference interact to produce a gene whose overall function confers susceptibility. Such a hypothesis would explain the Shifman *et al*[1] results; a single haplotype incorporating the promoter *Hind*III site might show an even stronger association with schizophrenia.

Interestingly, the haplotype identified by Shifman *et al*[1] and Bray *et al*[2] has the presumptive 'low expression' allele at the P2 promoter and the G allele coding for Val, the 'high activity' allele, at the Val158Met site. However, it is not known directly whether the Val158Met polymorphism affects the activity of the membrane-bound form of COMT (MB-COMT) protein, the predominant form of COMT expressed in brain dorsolateral prefrontal cortical neurons.[29] We could find no studies of the effect of the Val158Met polymorphism on the thermostability or enzyme activity of MB-COMT. Lotta *et al*[30] studied the soluble COMT protein (S-COMT, the exon 4 Val/Met polymorphism is located at amino acid 108 in S-COMT) and found no difference in the enzyme kinetics between the Val or Met variants. They did find that, in the absence of AdoMet/SAM, the cofactor in the methylation reaction, the Met (low activity, A, site present) allelic variant was thermolabile and the Val (high activity, G, site present) allelic variant was not. In the presence of AdoMet/SAM, there was no difference in the thermostability of the allelic variants. The exon 4 Val108Met amino acid, which is located on the opposite side of the S-COMT protein from the enzyme site, does not contribute to the active site, suggesting that conformational changes are not



the cause of the thermolability and enzyme activity differences seen with the different amino acids at this site. However, Lotta *et al*[30] do note that there is a direct connection of an alpha-helix and beta-strand within the structure of S-COMT between amino-acid residue 108 and the binding groove for AdoMet.

Table 3 Pairwise LD between the promoter *Hind*III and Exon 4 Val158Met sites in the six newly studied populations

Population	D′	Δ^2	ξ	*Permute test* P
Chagga	0.17	0.02	−0.01	0.382
Samaritans	0.46	0.18	0.00	0.298
Chuvash	0.64	0.27	0.28	0.001
Ashkenazi Jews	0.62	0.19	0.20	0.000
Komi Zyriane	0.61	0.25	0.10	0.016
Khanty	0.82	0.43	0.42	0.000
Karitiana	1.00	0.18	0.06	0.108[a]

[a]The statistics for the Karitiana are repeated here to correct a typographical error in the *P*-value in DeMille *et al*.[3]

Perhaps the Met variant of the S-COMT protein does have a different conformation that affects the binding of AdoMet/SAM. Lotta *et al*[30] also suggest that it is an interaction between a charged side chain on the COMT enzyme substrate and either the attached cellular membrane or the longer amino end of MB-COMT, rather than a conformational change in the protein structure, that causes the higher affinity of MB-COMT for dopamine. The higher affinity for dopamine of MB-COMT makes it more important than S-COMT at the physiological levels of dopamine found in the brain. It seems important to know whether the presence of the longer amino end of MB-COMT, or the adjacent cellular membrane, could stabilize the protein, as it seems to stabilize the enzyme/substrate interaction, and affect the thermostability of the Met allelic variant of MB-COMT, similar to the effect of AdoMet/Sam on the Met allelic variant of S-COMT.

The population variation in haplotype frequencies and associations of alleles at the different polymorphic sites of *COMT* highlight the difficulties of interpreting many of the published association studies that have examined only the Val158Met polymorphism. For example, in two recent studies ethnically heterogeneous samples without precise matching of the contrasted groups make the authors' conclusions impossible to accept based solely on the evidence presented: the question is raised whether population stratification could have resulted in an artifactual association.[31,32] In a study of Japanese, an association was found with the maintenance dose of narcoleptics.[33] The difficulty in this case comes in trying to extrapolate to other populations with very different haplotype frequencies.

The Ashkenazi Jews are a well-defined homogenous population with a large founder effect, which decreases the likelihood of false positive association due to sample stratification and increases the likelihood that a true association between genotype at *COMT* and schizophrenia, a complex behavioral disorder, was found. In our study of 38 populations at seven polymorphic sites, we find (1) the same allele and haplotype frequencies as Shifman *et al*[1] at the three polymorphic sites, IVS 1 rs737865, exon 4 Val158Met, and downstream rs165599 SNPs, in our sample of Ashkenazi Jews; (2) the frequencies of alleles and haplotypes for these three sites vary significantly around the world; (3) the haplotype consisting of these three sites can be subdivided by any of the four other sites in ways that would likely provide more information for future association studies with schizophrenia; and (4) the IVS1 rs737865 G allele is almost completely associated with one allele of the *Hind*III site in the P2 promoter in most populations studied, while the IVS1 rs737865 A allele is found in combination with both promoter alleles.

This population study emphasizes the importance of using haplotypes in association studies, not just a single marker, because *any* single non-etiological marker can have very different associations with the etiological variable in different populations and even in different samples from the same population. Haplotypes decrease that source of variation among studies and can increase the power. By using haplotypes in a large sample population, the Shifman *et al*[1] study found a highly significant association between schizophrenia and *COMT*, with a focus on the rs737865 SNP at the 5′ end of intron 1. The location of this SNP pointed to the *COMT* P2 promoter located only a couple of kb upstream. The results of our *COMT* seven-site haplotype study support the relevance of the *COMT* P2 promoter to schizophrenia. More importantly, the population data show that the schizophrenia-associated haplotype varies significantly in frequency around the world and has significant heterogeneity when other markers in *COMT* are also considered. These empiric observations must be considered in research study design and

Figure 3 Population frequencies of the three-site (IVS 1 rs737865—exon 4 Val158Met—downstream rs165599) G–G–G schizophrenia susceptibility haplotype at COMT. (a) Three-site haplotypes in 38 populations. The ancestral haplotype is AGG. Populations are grouped by region as in Table 1. ▒ A–G–G, ▨ A–A–G, ▚ A–G–A, ■ G–G–G, □ A–A–A, ▨ G–G–A, ▚ G–A–G, ▒ G–A–A. The numerical frequencies of the three-site haplotype can be found in ALFRED under the site UID SI000940N. (b) Heterogeneity of the G–G–G haplotype determined by four other sites across the gene. Haplotypes are given in the form – G – – G – G for those occurring at >5%; all other grouped in 'Residual'. The – is used to indicate the unspecified allele of a polymorphic site. Sites are ordered along the gene beginning with the promoter *Hind*III site. Note, all – G – – G – G haplotypes have the A allele at the *Hind*III site. ▨ AGN2G2G, □ AGO2G2G, ▨ AGO2G1G, ■ AGO1G2G, ▒ residual – G – – G – G. The total – G – – G – G haplotype frequencies for the two populations that are off-scale are: Yemenite Jew, 0.408, and Adygei, 0.309. The portions of the columns that are off-scale are the remainders of the AGO2G2G haplotype. The rest of the numerical frequencies of the seven-site haplotype can be found in ALFRED under the site UID SI000939V.

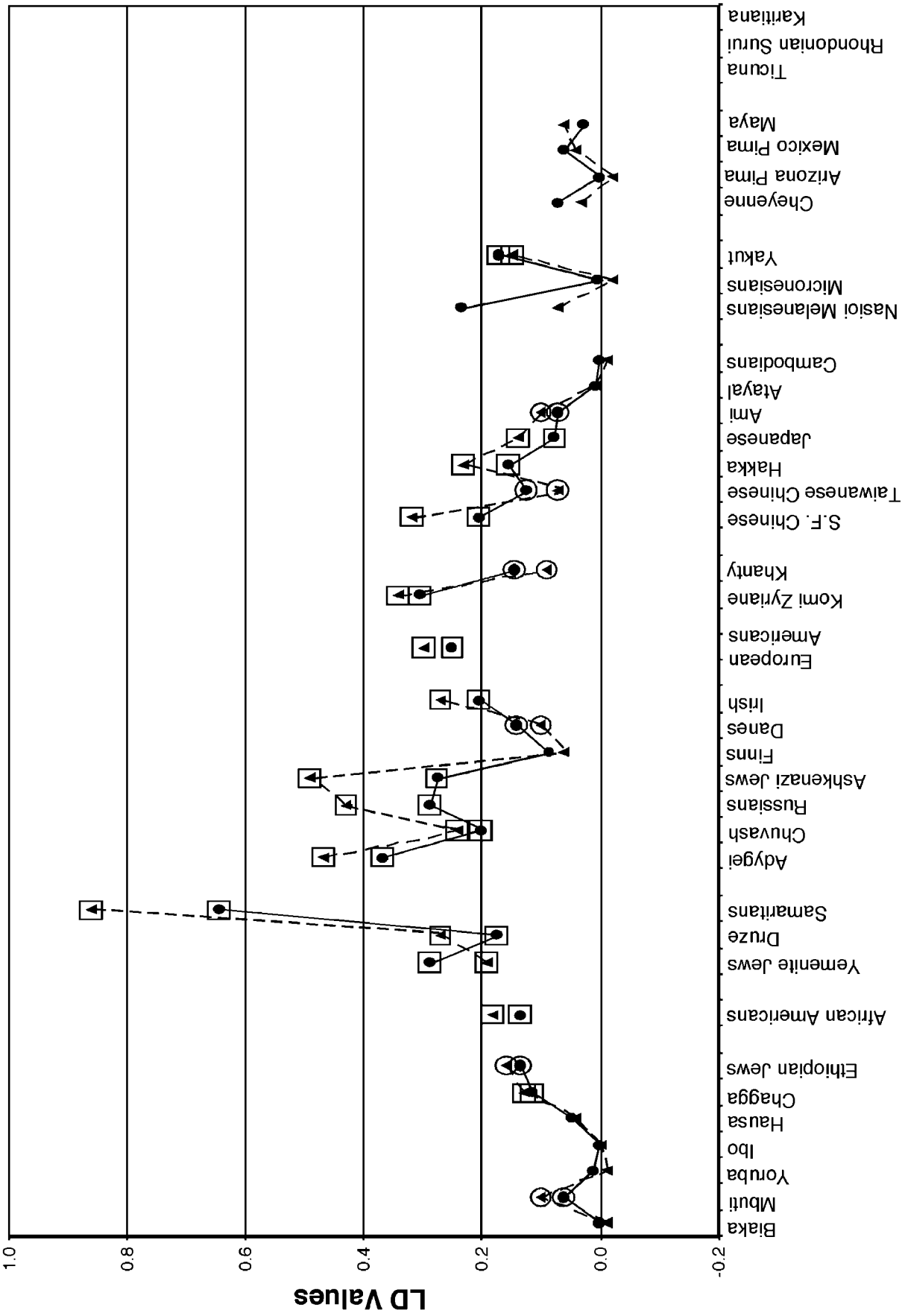


Figure 4 Pairwise LD between the IVS 1 rs737865 and 5′ *Hind*III sites. Populations are grouped by region as in Table 1. –▲– ξ values, –●– Δ^2 values, □ $P \leq 0.01$, ○ $0.01 < P \leq 0.05$. This figure also illustrates the general finding that the pairwise ξ is highly correlated with Δ^2. Our samples of three South American populations, Ticuna, R. Surui, and Karitiana, are fixed for the A allele of the IVS 1 rs737865 site (Figure 2) and, therefore, the LD could not be calculated.

Table 4 Conditional probabilities of the *Hind*III site A allele given either allele at IVS 1 rs737865

Population	*Probability of HindIII*A allele given: rs737865 G allele*	*rs737865 A allele*
Biaka	0.72	0.50
Mbuti	1.00	0.61
Yoruba	0.85	0.86
Ibo	1.00	0.64
Hausa	1.00	0.51
Chagga	1.00	0.54
Ethiopian Jews	1.00	0.55
African Americans	1.00	0.49
Yemenite Jews	0.92	0.45
Druze	1.00	0.54
Samaritans	0.97	0.19
Adygei	0.98	0.45
Chuvash	1.00	0.45
Russians	1.00	0.37
Ashkenazi Jews	1.00	0.49
Finns	1.00	0.51
Danes	0.88	0.36
Irish	0.97	0.46
European Americans	1.00	0.39
Komi Zyriane	1.00	0.50
Khanty	0.94	0.57
SF Chinese	1.00	0.49
Taiwanese Chinese	1.00	0.53
Hakka	1.00	0.56
Japanese	1.00	0.57
Ami	1.00	0.65
Atayal	1.00	0.85
Cambodians	0.71	0.67
Nasioi Melanesians	0.20	0.77
Micronesians	0.83	0.94
Yakut	1.00	0.58
Cheyenne	1.00	0.67
Arizona Pima	1.00	0.85
Mexico Pima	0	0.70
Maya	1.00	0.40
Ticuna	0	0.78
Rhondonian Surui	0	0.75
Karitiana	0	0.95

in attempts to reconcile data from studies of diverse populations.

Summary

Palmatier *et al* present a population genetic study of seven polymorphic sites at the COMT gene for samples of normal individuals from 38 populations around the world. They discuss the implications of their findings for future work aimed at explaining some intriguing recent reports implicating variation near the COMT gene in susceptibility to schizophrenia. We argue that behind the findings relating COMT with schizophrenia may be an as yet unidentified *cis*-acting variant in a nearby regulatory region, which drives transcription of the main form of COMT found in the brain.

Acknowledgements

We would like to thank Valeria Ruggeri, Roy Capper, and Lena Golovyan for their technical help. This work was supported by PHS grants, MH62495, AA09379 and GM57672, to KKK and NS01795 to MAP. We want to acknowledge and thank the following additional individuals for their help over the years in assembling the samples from the diverse populations: FL Black, William Byerley, LL Cavalli-Sforza, J Friedlaender, Kenneth Kendler, William Knowler, Frank Oronsaye, Leena Peltonen, Leslie O Schulz and Kenneth Weiss. Some cell lines were made available by the Coriell Institute for Medical Research and by the National Laboratory for the Genetics of Israeli Populations. Special thanks are due to the many hundreds of individuals who volunteered to give blood samples for studies such as this. Without such participation of individuals from diverse parts of the world, we would be unable to obtain a true picture of the genetic variation in our species.

Electronic-Database Information

ALFRED is a database of allele frequency information accessible over the worldwide web at the URL; http://alfred.med.yale.edu/

References

1 Shifman S, Bronstein M, Sternfeld M, Pisante-Shalom A, Lev-Lehman E, Weizman A *et al.* A highly significant association between a COMT haplotype and schizophrenia. *Am J Hum Genet* 2002; **71**: 1296–1302.
2 Bray NJ, Buckland PR, Williams NM, Williams HJ, Norton N, Owen MJ *et al.* A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. *Am J Hum Genet* 2003; **73**: 152–161.
3 DeMille MM, Kidd JR, Ruggeri V, Palmatier MA, Goldman D, Odunsi A *et al.* Population variation in linkage disequilibrium across the COMT gene considering promoter region and coding region variation. *Hum Genet* 2002; **111**: 521–537.
4 Li T, Ball D, Zhao J, Murray RM, Liu X, Sham PC *et al.* Family-based linkage disequilibrium mapping using SNP marker haplotypes: application to a potential locus for schizophrenia at chromosome 22q11. *Mol Psychiatry* 2000; **5**: 77–84.
5 Li T, Ball D, Zhao J, Murray RM, Liu X, Sham PC *et al.* Errata: Family-based linkage disequilibrium mapping using SNP marker haplotypes: application to a potential locus for schizophrenia at chromosome 22q11. *Mol Psychiatry* 2000; **5**: 452.
6 Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE *et al.* Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc Natl Acad Sci USA* 2001; **98**: 6917–6922.
7 Bilder RM, Volavka J, Czobor P, Malhotra AK, Kennedy JL, Ni X *et al.* Neurocognitive correlates of the COMT Val(158)Met polymorphism in chronic schizophrenia. *Biol Psychiatry* 2002; **52**: 701–707.
8 Murphy KC, Jones LA, Owen MJ. High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry* 1999; **56**: 940–945.

9 Karoum F, Chrapusta SJ, Egan MF. 3-Methoxytyramine is the major metabolite of released dopamine in the rat frontal cortex: reassessment of the effects of antipsychotics on the dynamics of dopamine release and metabolism in the frontal cortex, nucleus accumbens, and striatum by a simple two pool model. *J Neurochem* 1994; **63**: 972–979.
10 Sklar P. Linkage analysis in psychiatric disorders: the emerging picture. *Annu Rev Genomics Hum Genet* 2002; **3**: 371–413.
11 O'Donovan MC, Williams NM, Owen MJ. Recent advances in the genetics of schizophrenia. *Hum Mol Genet* 2003; **12**: R125–R133.
12 Maier W, Zobel A, Rietschel M. Genetics of schizophrenia and affective disorders. *Pharmacopsychiatry* 2003; **36**(Suppl 3)**:** 195–202.
13 Pulver AE. Search for schizophrenia susceptibility genes. *Biol Psychiatry* 2000; **47**: 221–230.
14 Riley BP, McGuffin P. Linkage and associated studies of schizophrenia. *Am J Med Genet* 2000; **97**: 23–44.
15 Williams NM, Norton N, Williams H, Ekholm B, Hamshere ML, Lindblom Y *et al.* A systematic genomewide linkage study in 353 sib pairs with schizophrenia. *Am J Hum Genet* 2003; **73**: 1355–1367.
16 Bertocci B, Miggiano V, Da Prada M, Dembic Z, Lahm HW, Malherbe P. Human catechol-*O*-methyltransferase: cloning and expression of the membrane-associated form. *Proc Natl Acad Sci USA* 1991; **88**: 1416–1420.
17 Tenhunen J, Salminen M, Lundstrom K, Kiviluoto T, Savolainen R, Ulmanen I. Genomic organization of the human catechol *O*-methyltransferase gene and its expression from two distinct promoters. *Eur J Biochem* 1994; **223**: 1049–1059.
18 Lundstrom K, Tenhunen J, Tilgmann C, Karhunen T, Panula P, Ulmanen I. Cloning, expression and structure of catechol-*O*-methyltransferase. *Biochim Biophys Acta* 1995; **1251**: 1–10.
19 Palmatier MA, Kang AM, Kidd KK. Global variation in the frequencies of functionally different catechol-*O*-methyl transferase alleles. *Biol Psychiatry* 1999; **46**: 557–567.
20 Matsumoto M, Weickert CS, Beltaifa S, Kolachana B, Chen J, Hyde TM *et al.* Catechol *O*-methyltransferase (COMT) mRNA expression in the dorsolateral prefrontal cortex of patients with schizophrenia. *Neuropsychopharmacology* 2003; **28**: 1521–1530.
21 Tunbridge E, Burnet PWJ, Sodhi S, Harrision PJ. Catechol-*O*-methyltransferase (COMT) and proline dehydrogenase (PRODH) mRNAs in the dorsolateral prefrontal cortex in schizophrenia, bipolar disorder, and major depression. *Synapse* 2004; **51**: 112–118.
22 Iyengar S, Seaman M, Deinard AS, Rosenbaum HC, Sirugo G, Castiglione CM *et al.* Analyses of cross species polymerase chain reaction products to infer the ancestral state of human polymorphisms. *DNA Seq* 1998; **8**: 317–327.
23 Hawley ME, Kidd KK. HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. *J Hered* 1995; **86**: 409–411.
24 Lewontin R. The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* 1964; **49**: 49–67.
25 Zhao H, Pakstis AJ, Kidd JR, Kidd KK. Assessing linkage disequilibrium in a complex genetic system. I. Overall deviation from random association. *Ann Hum Genet* 1999; **63**: 167–179.
26 Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 1995; **29**: 311–322.
27 Pakstis AJ, Kidd JR, Kidd KK. A reference distribution of Fst values for biallelic DNA markers. *Am J Hum Genet* 2002; **71**(Suppl)**:** 371.
28 Calafell F, Grigorenko EL, Chikanian AA, Kidd KK. Haplotype evolution and linkage disequilibrium: a simulation study. *Hum Hered* 2001; **51**: 85–96.
29 Matsumoto M, Weickert CS, Akil M, Lipska BK, Hyde TM, Herman MM *et al.* Catechol *O*-methyltransferase mRNA expression in human and rat brain: evidence for a role in cortical neuronal function. *Neuroscience* 2003; **116**: 127–137.
30 Lotta T, Vidgren J, Tilgmann C, Ulmanen I, Melen K, Julkunen I *et al.* Kinetics of human soluble and membrane-bound catechol *O*-methyltransferase: a revised mechanism and description of the thermolabile variant of the enzyme. *Biochemistry* 1995; **34**: 4202–4210.
31 Wonodi I, Stine OC, Mitchell BD, Buchanan RW, Thaker GK. Association between Val108/158 Met polymorphism of the COMT gene and schizophrenia. *Am J Med Genet* 2003; **120B**: 47–50.
32 Strous RD, Nolan KA, Lapidus R, Diaz L, Saito T, Lachman HM. Aggressive behavior in schizophrenia is associated with the low enzyme activity COMT polymorphism: a replication study. *Am J Med Genet* 2003; **120B**: 29–34.
33 Inada T, Nakamura A, Iijima Y. Relationship between catecho*l*-*O*-methyltransferase polymorphism and treatment-resistant schizophrenia. *Am J Med Genet* 2003; **120B**: 35–39.