The Structure of Linkage Disequilibrium at the *DBH* Locus Strongly Influences the Magnitude of Association between Diallelic Markers and Plasma Dopamine β-Hydroxylase Activity

Cyrus P. Zabetian,[1,3,*] Sarah G. Buxbaum,[4,†] Robert C. Elston,[4] Michael D. Köhnke,[5] George M. Anderson,[2] Joel Gelernter,[1,3] and Joseph F. Cubells[1,3]

[1]Department of Psychiatry and [2]Child Study Center, Yale University School of Medicine, New Haven, CT; [3]Department of Psychiatry, VA Connecticut Healthcare System, West Haven, CT; [4]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; and [5]University Hospital of Psychiatry and Psychotherapy, Tübingen University Hospital, Tübingen, Germany

There is currently a great deal of interest in using linkage disequilibrium (LD) mapping to locate both disease and quantitative-trait loci on a genomewide scale. Recent findings suggest that much of the human genome is organized in discrete "blocks" of low haplotype diversity, but the utility of such blocks in identifying genes influencing complex traits is not yet known and must ultimately be tested empirically through use of real data. We recently identified a putative functional polymorphism (−1021C→T) in the 5′ upstream region of the *DBH* gene that accounted for 35%–52% of the total phenotypic variance in plasma dopamine β-hydroxylase (DBH) activity in samples from three distinct populations. In the present study, we genotyped 11 diallelic markers at the *DBH* locus surrounding −1021C→T in 386 unrelated individuals of European origin. We identified a single 10-kb block containing −1021C→T, in which four haplotypes comprised 93% of the observed chromosomes. Only markers within the block were highly associated with phenotype ($P \leqslant 2.2 \times 10^{-10}$), with one exception. In general, association with phenotype was strongly correlated with the degree of LD between each marker and −1021C→T. Of four LD measures assessed, d^2 was the best predictor of this relationship. Had one attempted to map quantitative-trait loci for plasma DBH activity on a genomewide basis without prior knowledge of candidate regions and not included (by chance) markers within this haplotype block, the *DBH* locus might have been missed entirely. These results provide a direct example of the potential value of constructing a haplotype map of the human genome prior to embarking on large-scale association studies.

Introduction

Linkage disequilibrium (LD) mapping might prove to be more powerful than linkage analysis in identifying genes underlying complex traits (Risch 2000). Recent advances in molecular and computational technology promise to make genomewide LD mapping feasible in the near future. The growing public database (dbSNP Home Page) of ~2.8 million SNPs will be critical in this endeavor. The optimum choice of SNP marker densities, locations, and allele frequencies must first be decided, however, and some authors have suggested that an LD map of the human genome be constructed to assist with the process (Daly et al. 2001). Several recent studies have emphasized the necessity of such a tool, by demonstrating that the extent of LD throughout the genome is extremely variable even at intragenic scales (Stephens et al. 2001; Tiret et al. 2002) and that the genome might be organized in discrete "blocks" of LD (Daly et al. 2001; Gabriel et al. 2002). By understanding the underlying block structure of LD in each region of interest, one might be able to assign marker locations more efficiently and thus to minimize inadequate or redundant coverage.

Whether such blocks of LD, defined statistically on the basis of common SNPs, will be useful for detecting associations with genes underlying complex traits is presently unknown. For both quantitative and dichotomous traits, this will depend on many factors, including the frequency and effect size for each trait allele, the measure used to define LD, and the sampling strategy employed (Risch 2000; Ardlie et al. 2002). Although several authors have attempted to address these issues through use of computer simulations (Kruglyak 1999; Schork et al. 2000), the key factors need to be determined empirically. Studies examining the effect of LD

Received November 13, 2002; accepted for publication March 14, 2003; electronically published April 30, 2003.

Address for correspondence and reprints: Dr. Joseph F. Cubells, Psychiatry 116A2, VA Connecticut Healthcare System, 950 Campbell Avenue, West Haven, CT 06516. E-mail: joseph.cubells@yale.edu

[*] Present affiliation: Department of Neurology, University of Washington School of Medicine and Department of Neurology, VA Puget Sound Health Care System, Seattle.

[†] Present affiliation: Department of Human Genetics, University of Pittsburgh, Pittsburgh.

0002-9297/2003/7206-0004$15.00

structure on the association of specific markers to traits using real data are scarce, with a few notable exceptions (Martin et al. 2000).

The *DBH* gene, which encodes dopamine β-hydroxylase (DBH [MIM 223360]), represents a simple model for directly assessing the extent of useful LD surrounding a common trait allele of large effect. DBH catalyzes the conversion of dopamine to norepinephrine and is released into the circulation from sympathetic neurons, and its enzymatic activity is readily assayed in plasma or serum (Weinshilboum and Axelrod 1971). A single major QTL accounting for about half the heritability of DBH activity was initially mapped to chromosome 9q34 by linkage analysis (Goldin et al. 1982; Wilson et al. 1988), and later studies identified the *DBH* gene, located within this region, as this QTL (Wei et al. 1997; Cubells et al. 1998). We recently discovered a putative functional SNP (−1021C→T), located within the promoter of the *DBH* gene, that accounted for 35%–52% of the total variation in plasma DBH activity levels in three distinct populations (Zabetian et al. 2001). Using sequencing-based mutational analysis of the entire *DBH* coding region, intron-exon junctions, and 5′ flanking region (up to −1.5 kb from the translational start site, later extended to −2.6 kb) in 16–24 chromosomes from individuals with extreme phenotypes, we were unable to identify any other functional candidate polymorphisms of large effect.

In the present study, we describe the LD structure of the *DBH* gene and provide evidence for the presence of at least one discrete block of LD. Assuming that −1021C→T is a true functional polymorphism, we examine how the extent of LD between −1021C→T and a group of 11 surrounding diallelic markers influences the strength of association of each marker to plasma DBH activity. We illustrate the sensitivity of this relationship to the method used to calculate LD for four measures: the absolute value of Lewontin's D' ($|D'|$), the arctan transformation of the absolute value of the log odds ratio (t|LOR|), the squared difference in proportions (d^2), and the square of the correlation coefficient between two loci (Δ^2). These results are relevant to the design of genomewide LD mapping studies for quantitative traits and, possibly, for categorical phenotypes as well.

Material and Methods

Subjects

Plasma and DNA specimens were collected from a total of 386 unrelated adults in the course of several ongoing

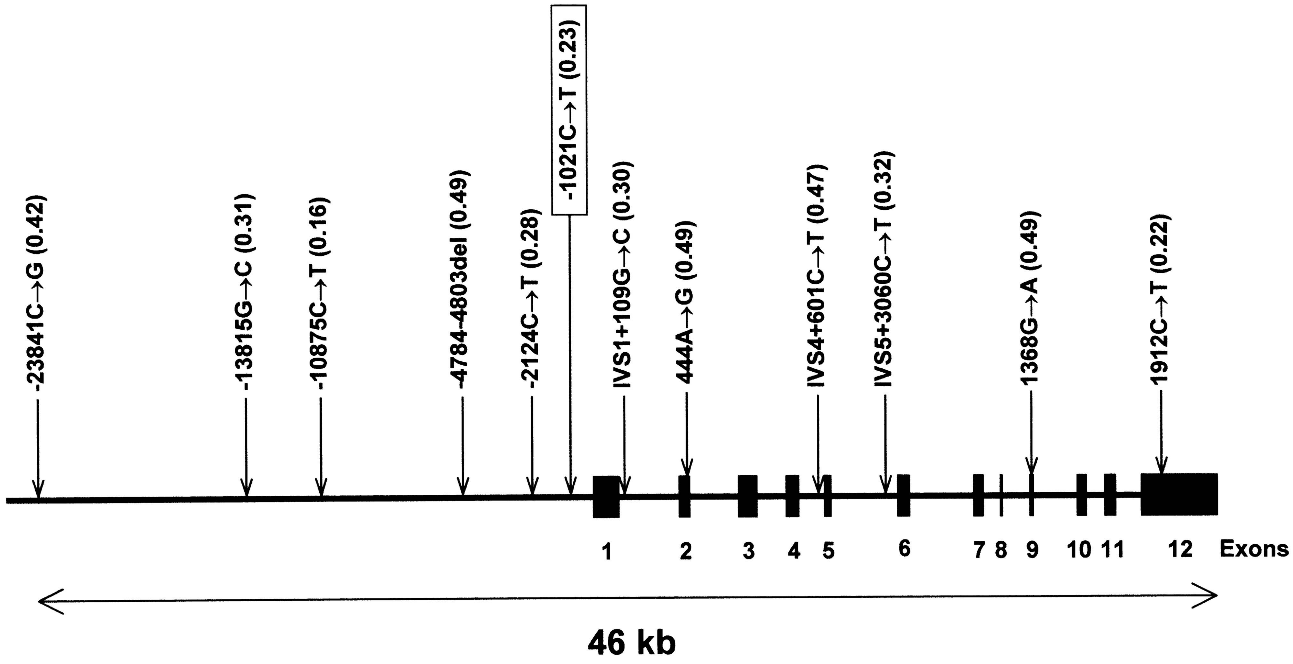

Figure 1 Location of the 12 diallelic polymorphisms at the *DBH* locus used in LD analysis. The putative functional SNP −1021C→T (*boxed*) is located at approximately the midpoint of the region spanned by these markers. The minor allele frequency of each marker is listed in parentheses. Nucleotide positions are numbered according to the cDNA sequence for exons, or genomic sequence for the 5′ flanking region, beginning at the A of the ATG initiator Met codon. Positions for introns are numbered according to the genomic sequence starting from the G of the donor site invariant GT.

genetic studies, as described elsewhere (Zabetian et al. 2001; Köhnke et al. 2002). In brief, 169 European Americans and 217 individuals of German origin were recruited from the northeastern United States and southwestern Germany, respectively. Ethnic groups were self-defined, and those of known mixed or other heritage (other than mixed European) were excluded. The groups included healthy individuals and those with psychiatric and substance-use disorders. As discussed elsewhere (Cubells et al. 1998), sampling from a variety of diagnostic groups is unlikely to obscure fundamental genetic influences on plasma DBH activity.

Laboratory Methods

We selected a total of 12 diallelic markers for study at the *DBH* locus, spanning 46 kb (fig. 1). These consisted of a 19–bp insertion/deletion (Nahmias et al. 1992) and 11 SNPs derived from our previous study (Zabetian et al. 2001) and dbSNP. The positions of the markers were roughly symmetrically distributed around −1021C→T and included the entire *DBH* coding region and ~23 kb of 5′ upstream sequence. We chose only common markers with minor allele frequencies >0.15 and did not exclude any SNPs at CpG sites.

Genotypes were determined using RFLP techniques, by digesting PCR products with the appropriate restriction enzymes. PCR primers were designed with mismatches in some cases, to generate artificial restriction sites as needed. Digested PCR products were electrophoresed on ethidium bromide–stained agarose gels, were photographed under UV transillumination, and were independently double scored. The average rate of PCR nonamplification for all genotype groups was 1.7%, and the maximum rate was 3.1%. Results from representative samples of each genotype were confirmed by direct sequencing of PCR products through use of an ABI Prism 377 DNA Sequencer. Further methodological details, PCR primer sequences, and source data for each marker are available upon request from the authors.

DBH activity was assayed in plasma samples from all subjects through use of a sensitive high-performance liquid chromatography–fluorometric method, as described elsewhere (Cubells et al. 1998). Separation and measurement of the enzyme product, octopamine, in the presence of a large excess of the substrate, tyramine, was accomplished with a detection limit of <1 pmol. All measurements were performed in duplicate, and average enzyme activities are expressed in nanomoles/minute/milliliter plasma.

Statistical Analysis

We used the EH+ program, version 1.11 (Zhao et al. 2000) to estimate relative two-locus haplotype frequencies as given in the following table:

		Locus 2	
		Allele 1	Allele 2
Locus 1	Allele 1	*a*	*b*
	Allele 2	*c*	*d*

The 2LD program (Zapata et al. 2001) was then used to calculate D', D, and $D_{\max}$, where D' is defined as $D/D_{\max}$, $D = (ad) - (bc)$, and $D_{\max} = \min(a+b)(b+d),(c+d)(a+c)$ for $D>0$ and $\min(a+b)(a+c),(c+d)(b+d)$ for $D<0$ (Lewontin 1964). The squared standardized difference in proportions was calculated using the formula $d^2 = D^2/[(a+b)(c+d)]^2$, where locus 1 is the trait locus (Nei and Li 1980). In figures 3 and 4, −1021C→T is considered the trait locus, whereas, in figure 5, each marker was sequentially designated the trait locus. The square of the correlation coefficient between two loci is given by $\Delta^2 = D^2/[(a+b)(c+d)(a+c)(b+d)]$. The odds ratio was determined by first multiplying each of the relative haplotype frequencies in a fourfold table by $2n$ and then adding 0.5 haplotype counts to each cell (Haldane 1955). This correction ensures nonzero values in all cells. The odds ratio is then given by ad/bc, where a, b, c, and d are the corrected relative haplotype frequencies, and natural logarithms were taken to obtain the LOR. The arctan transformation of the |LOR| was then divided by $\pi/2$ to obtain a measure whose maximum value is 1, which is abbreviated t|LOR|. Multilocus haplotype frequencies were estimated by the expectation-maximization (EM) method using Arlequin, version 2.000 (Schneider et al. 2000). Deviations from Hardy-Weinberg equilibrium (HWE) were assessed using the HWSIM program (Cubells et al. 1997; Kidd Lab Web site). Since some of the analyses contained small numbers of observations in some cells, P values for all analyses were estimated empirically through use of Monte Carlo simulations (10,000 iterations in each case) based on observed allele frequencies. Significance levels were estimated as the proportion of times the simulated distribution reached or exceeded the observed deviation from HWE.

A square-root transformation of plasma DBH activity was employed to stabilize the variance, as discussed elsewhere (Zabetian et al. 2001). Marker-phenotype association was measured using multiple and simple linear regression with the GLM procedure in SAS. P values were calculated under the assumption of normality and homoscedasticity of square-root DBH activity. Spearman correlation coefficients (ρ) and the corresponding P values under the null hypothesis $\rho = 0$ were calculated using the FREQ procedure in SAS.

Results

None of the marker genotypes deviated significantly from HWE after application of the appropriate Bonferroni correction for multiple tests. A plot of intermarker distance versus LD as measured by |*D*′|, for all possible pairwise combinations of markers, is displayed in figure 2*A*. Over the 46-kb interval spanned by the marker set, a clear decay of LD with increasing physical distance was evident, although the relationship for any single pairwise comparison was quite variable. In figure 2*B*, the average value of |*D*′| for all markers within sequential 5-kb intervals is plotted against distance. The "half-length" of LD, as defined by the distance over which the average value of |*D*′| falls below .5, was ~10 kb. No values of |*D*′| >0.5 were observed for any marker pairs farther apart than 22 kb (fig. 2*A*).

Visual inspection of the square matrix of |*D*′| values revealed a contiguous block of high values (|*D*′| > .79) for five sequential markers, suggesting the presence of a block of LD with limited haplotype diversity (table 1). These five markers, −2124C→T, −1021C→T, IVS1+109G→C, 444A→G, and IVS4+601C→T, spanned a 9.9-kb segment extending from the 5′ upstream area to roughly the first one-third of the genic region (fig. 1). Within this block, only 4 of the 32 possible haplotypes were estimated to occur at a frequency >2%, and together they accounted for an estimated 93% of the observed chromosomes (table 2). The ends of the block were well delineated, since adding the adjacent marker in the 5′ (−4784-4803del) or 3′ (IVS5+3060C→T) direction in a six-locus analysis increased the number of observed haplotypes with a frequency >2% to seven and nine, respectively. Although −4784-4803del was in relatively strong LD with −1021C→T (|*D*′| = .81), located 3.8 kb away, it was in weak LD with the two markers located at the 3′ end of the block (444A→G, |*D*′| = .32; and IVS4+601C→T, |*D*′| = .28) and was thus excluded

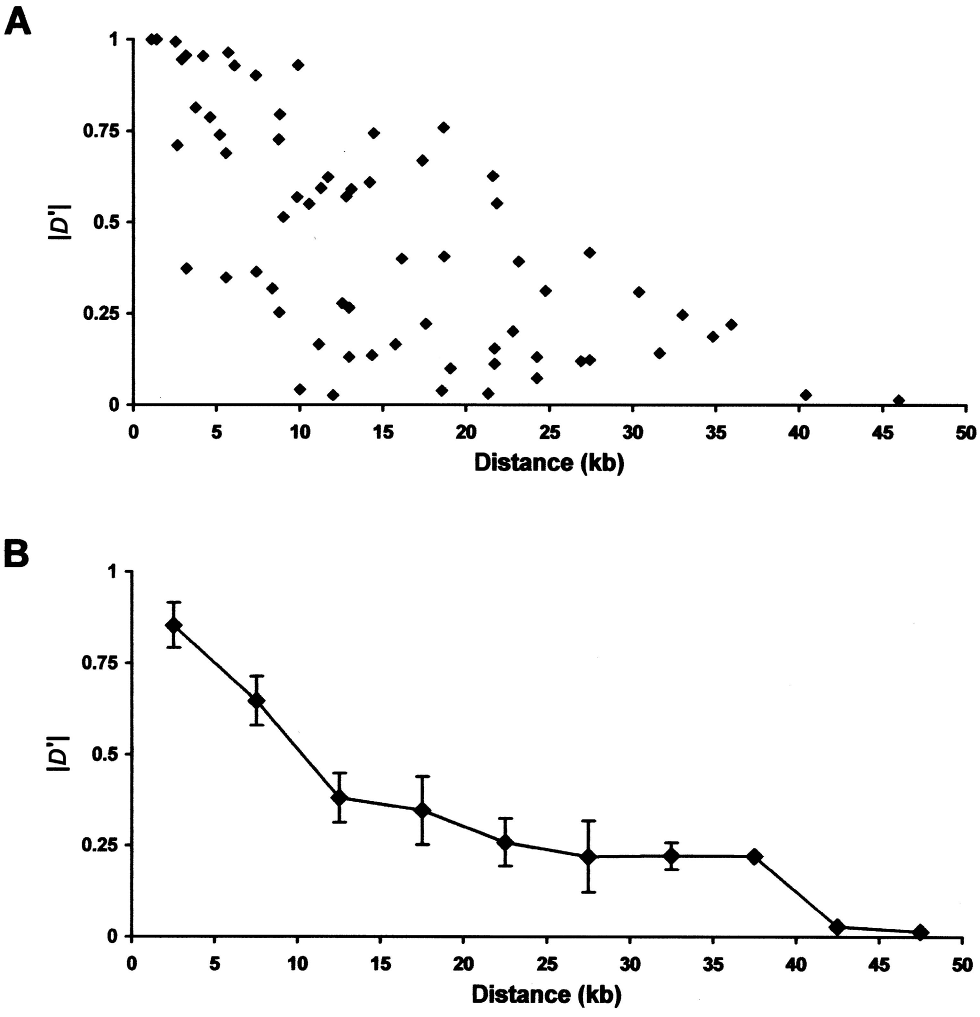

Figure 2 LD as measured by |*D*′| as a function of physical distance at the *DBH* locus for all pairs of markers (*A*) and the average of all markers within sequential 5-kb intervals (*B*) (±SEM).

Table 1

Pairwise LD as Measured by |*D*′| for All Markers at the DBH Locus

	|*D*′|											
	−23841C→G	−13815G→C	−10875C→T	−4784-4803del	−2124C→T	−1021C→T	IVS1+109G→C	444A→G	IVS4+601C→T	IVS5+3060C→T	1368G→A	1912C→T
−23841C→G	…	.04	.13	.10	.15	.20	.13	.12	.14	.19	.03	.01
−13815G→C	.04	…	.95	.51	.62	.57	.61	.67	.63	.31	.31	.22
−10875C→T	.13	.95	…	.93	.73	.57	.59	.74	.76	.55	.42	.25
−4784-4803del	.10	.51	.93	…	.71	.81	.74	.32	.28	.17	.03	.12
−2124C→T	.15	.62	.73	.71	…	1.00	.99	.96	.93	.59	.41	.07
−1021C→T	.20	.57	.57	.81	1.00	…	1.00	.79	.80	.03	.22	.39
IVS1+109G→C	.13	.61	.59	.74	.99	1.00	…	.96	.90	.55	.40	.11
444A→G	.12	.67	.74	.32	.96	.79	.96	…	.95	.36	.27	.04
IVS4+601C→T	.14	.63	.76	.28	.93	.80	.90	.95	…	.37	.25	.14
IVS5+3060C→T	.19	.31	.55	.17	.59	.03	.55	.36	.37	…	.69	.17
1368G→A	.03	.31	.42	.03	.41	.22	.40	.27	.25	.69	…	.35
1912C→T	.01	.22	.25	.12	.07	.39	.11	.04	.14	.17	.35	…

NOTE.—The boxed area indicates the contiguous block of high values for five sequential markers.

Table 2

Estimated Frequencies of the Four Common Haplotypes Observed within the Block of LD at the *DBH* Locus

Haplotype					
−2124	−1021	IVS1+109	444	IVS4+601	Frequency
C	C	G	A	C	.296
T	C	C	G	T	.267
C	T	G	A	C	.205
C	C	G	G	T	.162

from the block (table 1). Further haplotype analysis did not reveal any additional blocks of LD of three or more markers at the *DBH* locus.

The association of individual markers with phenotype was assessed by simple linear regression; the R^2 and corresponding P values are displayed in table 3. The putative functional SNP −1021C→T accounted for nearly half of the total variance in square-root plasma DBH activity levels in the sample. All of the markers within the LD block strongly associated with phenotype ($P \leqslant 2.2 \times 10^{-10}$). Multiple linear regression analysis was then performed to assess the combined effects of the markers on square-root plasma DBH activity, under the assumption of an additive model for each SNP, as described elsewhere (Zabetian et al. 2001). For the subset of the sample without missing genotypes ($n = 340$), $R^2 = .480$ in the full model with all 12 markers. In a reduced model including only −1021C→T, $R^2 = .457$; the added contribution of the remaining 11 markers was not significant ($F = 1.35$; $P = .20$).

In figure 3, LD, as measured by $|D'|$, t|LOR|, d^2, and Δ^2 between each of the markers and −1021C→T, is plotted against distance, and the R^2 values from table 3 are included for comparison. In general, both LD and association with phenotype were inversely proportional to the distance from −1021C→T. There was a strong correlation between the degree of LD with −1021C→T and the magnitude of association between each marker and phenotype; plots of these data for all four LD measures fit a straight line reasonably well but made very different predictions for a marker in absolute LD with −1021C→T (fig. 4). For such a marker, $|D'|$ and t|LOR| predicted an R^2 of .10–.15, whereas d^2 and Δ^2 suggested much higher values, of ∼.5 and .7, respectively. Though no markers with near-maximal values of d^2 or Δ^2 were included in the data set, d^2 most accurately predicted the observed association of the putative functional SNP (−1021C→T) with phenotype ($R^2 = .46$). In figure 4, this is illustrated by a data point for a hypothetical marker with $x = 1.0$ (the maximum possible value for $|D'|$, t|LOR|, d^2, and Δ^2) and $y = .46$.

To further assess the utility of LD structure and the influence of the LD measure chosen in localizing trait alleles at the *DBH* locus, we re-examined our data without assuming that −1021C→T was a functional SNP. In stepwise fashion, we designated each of the markers as the putative functional polymorphism and calculated the rank correlation between the R^2 values from table 3 and the degree of LD with each of the 11 surrounding markers (fig. 5). When d^2 and Δ^2 were used as the measures of LD, the Spearman correlation coefficient (r_s) was highest for −1021C→T as expected (fig. 5*C* and 5*D*). However, for $|D'|$ and t|LOR|, r_s was nearly equal at −1021C→T and at an adjacent marker, IVS1+109G→C (fig. 5*A* and 5*B*).

Discussion

Recently, several groups have measured the extent of LD between genetic markers both within genes and in intergenic regions in population samples of European descent (Abecasis et al. 2001; Frisse et al. 2001; Reich et al. 2001). In comparison with these studies, our results indicated that, in general, LD at the *DBH* locus extends over relatively short distances for simple pairwise marker analyses. For example, the half-length of LD for *DBH* appears shorter (10 kb; fig. 2*B*) than for all but 1 of the 19 regions examined by Reich et al. (2001). Although this might indicate a high local rate of recombination at the *DBH* locus, other factors, such as admixture, genetic drift, natural selection, and the age of the polymorphisms examined, likely influence estimates of LD over short physical distances (Hill and Weir 1994; Jorde et al. 1994). Furthermore, since $|D'|$ is biased upward with decreasing sample size (Weiss and Clark 2002), caution must be used in comparing studies that analyze widely disparate numbers of chromosomes.

Gabriel et al. (2002) surveyed 51 autosomal regions, evenly spaced throughout the genome, in four population groups, and found that most of the sequence was contained in LD blocks in which three to five common haplotypes accounted for ~90% of all chromosomes. In individuals of European ancestry, the average block length was estimated at 22 kb but ranged from <1 to

Table 3

Association of Markers to Plasma DBH Activity

Marker	R^2	P
−23841C→G	.006	.30
−13815G→C	.041	4.0×10^{-4}
−10875C→T	.008	.23
−4784-4803del	.110	3.0×10^{-10}
−2124C→T	.110	2.2×10^{-10}
−1021C→T	.463	3.0×10^{-52}
IVS1+109 G→C	.110	4.4×10^{-11}
444A→G	.130	6.1×10^{-13}
IVS4+601C→T	.128	7.2×10^{-12}
IVS5+3060C→T	.004	.49
1368G→A	.011	.14
1912C→T	.018	.03

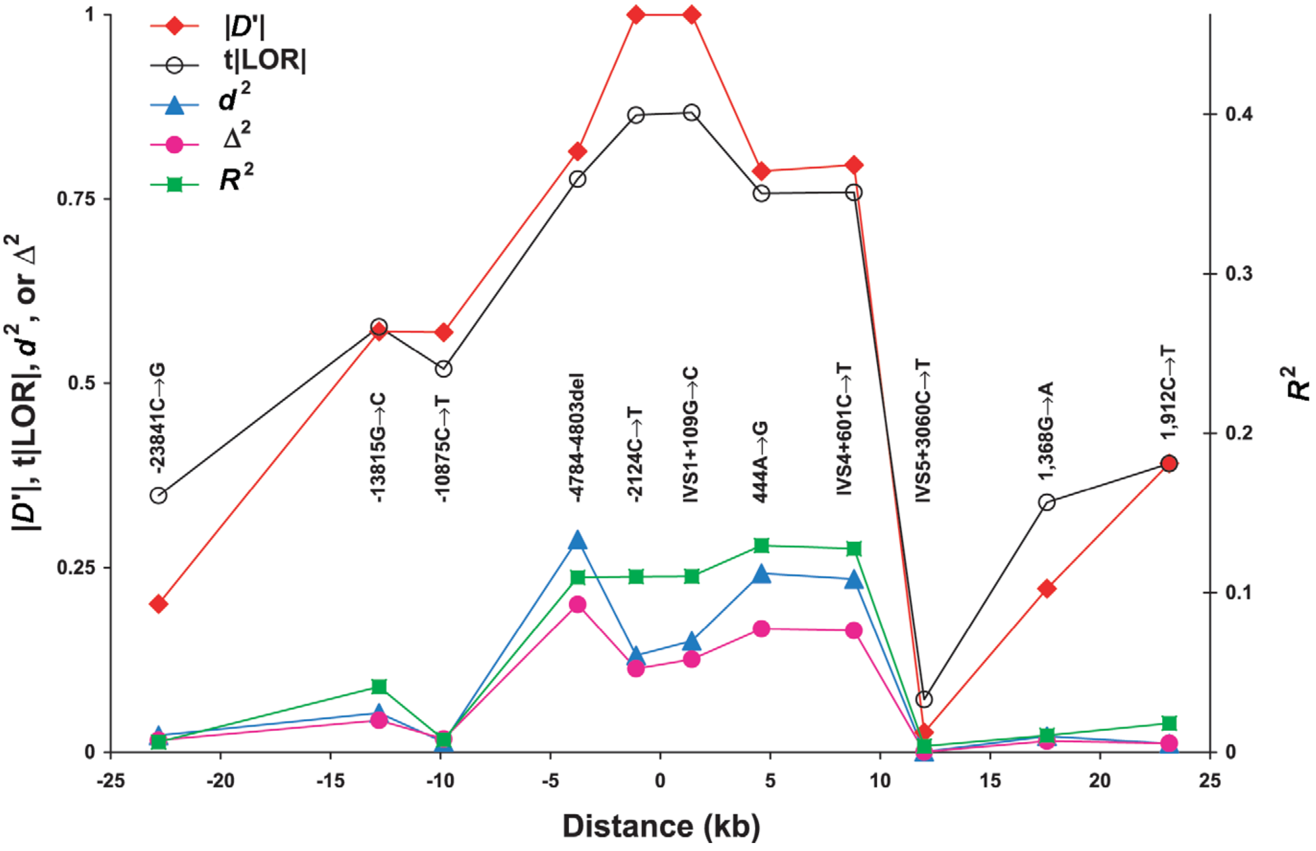

Figure 3 LD and association with phenotype for all surrounding markers as a function of physical distance from −1021C→T. LD between each marker and −1021C→T is displayed for four measures and the position of −1021C→T defines the zero-point on the *X*-axis. The right *Y*-axis is scaled such that its maximum value is .46, the R^2 obtained for −1021C→T.

173 kb, and short blocks <5 kb were very common. Similarly, Daly et al. (2001) found that a 500-kb segment on chromosome 5q31 could be divided into 11 blocks of 3–92 kb in length that encompassed ~75% of the total sequence and were characterized by 2–4 common haplotypes. The intervening regions were relatively short and might represent recombinational hotspots (Jeffreys et al. 2001). We have identified a block of LD at the *DBH* locus with similar properties, spanning nearly 10 kb, that includes the putative functional SNP −1021C→T (table 2). We were unable to find any additional blocks, which, if we assume that the block-structure model of LD is generally applicable, suggests one of two possibilities. First, the markers selected in our study were designed to cluster around −1021C→T (fig. 1). If the *DBH* locus is composed primarily of short blocks of a few kilobases in length, then the marker densities in the 5′ and 3′ regions of the 46-kb segment examined might not have been adequate to detect them. Alternatively, it is possible that the sequence flanking the block containing −1021C→T constitutes two extended “interblock” regions of high haplotype diversity.

In the present study, marker-phenotype association was well correlated with the extent of LD between each marker and −1021C→T, regardless of the LD measure used (figs. 3 and 4). The highest associations with phenotype were found for markers within the block and for a nearby marker, −4784-4803del, whereas the three markers downstream of the block were weakly correlated at best (table 3). These results provide potentially useful insights for genomewide LD mapping studies. To illustrate, let us assume that one had characterized the plasma DBH activity phenotype, had not previously identified any positional or functional candidate regions, and wished to map potential QTLs using population-based samples and regression methods. A variety of strategies using single-marker associations have been proposed for initial LD mapping studies of disease traits, and several authors agree that SNPs spaced ~10–30 kb apart might suffice for this purpose (Roses 2000; Ardlie et al. 2002). Using a similar design in our search for QTLs, we would likely place at least one marker within the *DBH* gene itself, particularly if we biased marker selection in favor of coding SNPs in known genes. Our sample would consist of several hundred unrelated European Americans. To correct for mul-

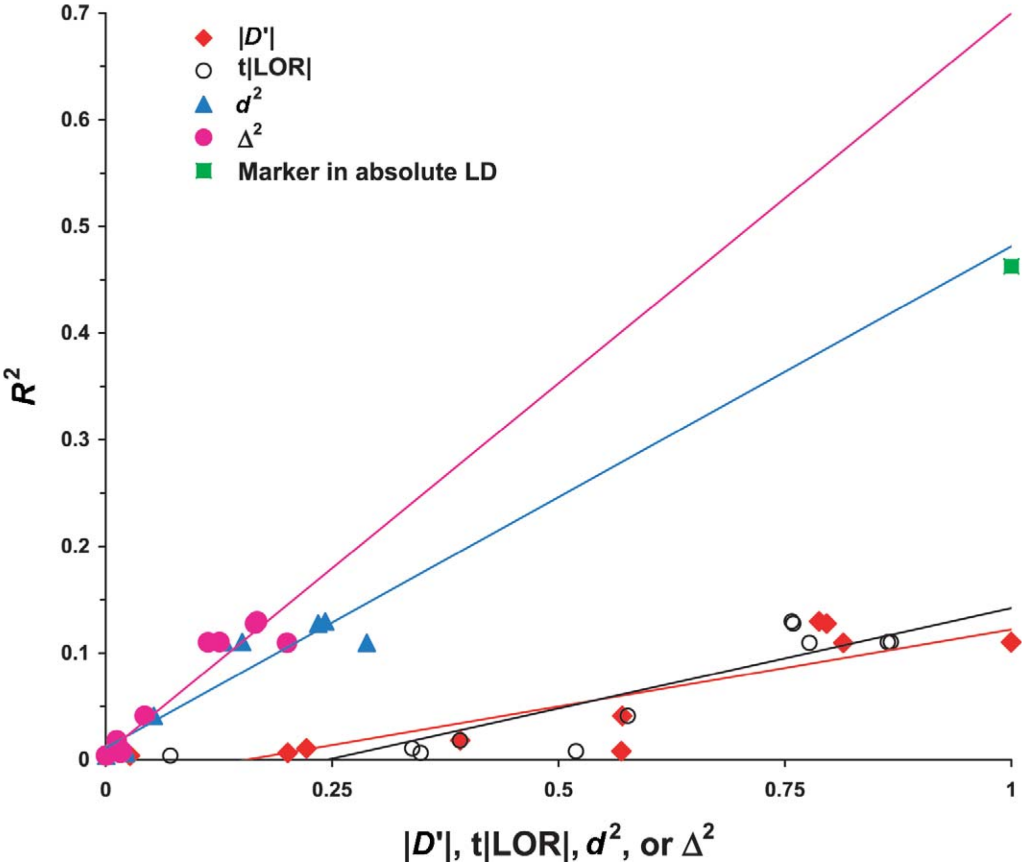

Figure 4 Comparison of the relationship between association with phenotype and LD to −1021C→T as measured by |*D*′|, t|LOR|, d^2, and Δ^2. A data point for a hypothetical marker in absolute LD with the putative functional SNP (−1021C→T) is included for reference.

tiple testing using 100,000–300,000 SNPs, we would need to apply a highly stringent significance level, on the order of $\alpha = 2.5 \times 10^{-7}$ to 8.3×10^{-8} (Schork 2002). Would such an approach have identified the *DBH* locus as a potential QTL for plasma DBH activity, given these constraints? Probably so, if one or more markers had resided within the LD block surrounding −1021C→T (table 3). However, if all markers were located outside of the block, which could easily have occurred by chance, given the marker densities employed, the *DBH* locus might well have been missed. If a genomewide mapping study based on SNP spacing alone failed to detect the single major QTL for a trait of high heritability, how much more difficult might it be to detect QTLs for complex traits, each presumably of small effect? This underscores the potential utility of constructing a genomewide haplotype map prior to undertaking such large-scale studies, as has been proposed elsewhere (Daly et al. 2001). Prior knowledge of haplotype structure at the *DBH* locus would have ensured the inclusion of markers representative of the block, and Johnson et al. (2001) have suggested formal methods to select such "haplotype tagging" SNPs (htSNPs). Inspection of table 2 reveals that information from just three of the five markers—for example, −2124C→T, −1021C→T, and 444A→G—effectively captures or tags the haplotype diversity of the entire block.

The block-structure model of LD is appealing because it suggests that a limited number of SNPs might be sufficient to adequately survey common variation within potential trait loci throughout the genome. However, consensus criteria for the definition of haplotype blocks have not yet been reached, and a variety of methods have been used to define block structure (Daly et al. 2001; Patil et al. 2001; Gabriel et al. 2002). Some authors have suggested that further SNP discovery and analysis of underlying haplotypes be prioritized to areas in or around known genes (Johnson et al. 2001). For genes spanned by a single block of LD, as few as two to five SNPs might be adequate to represent all of the common haplotypes, when existing resources such as dbSNP are used. At the other extreme, regions containing multiple short-length blocks and/or long intervening sequences (recombination hotspots) might require a

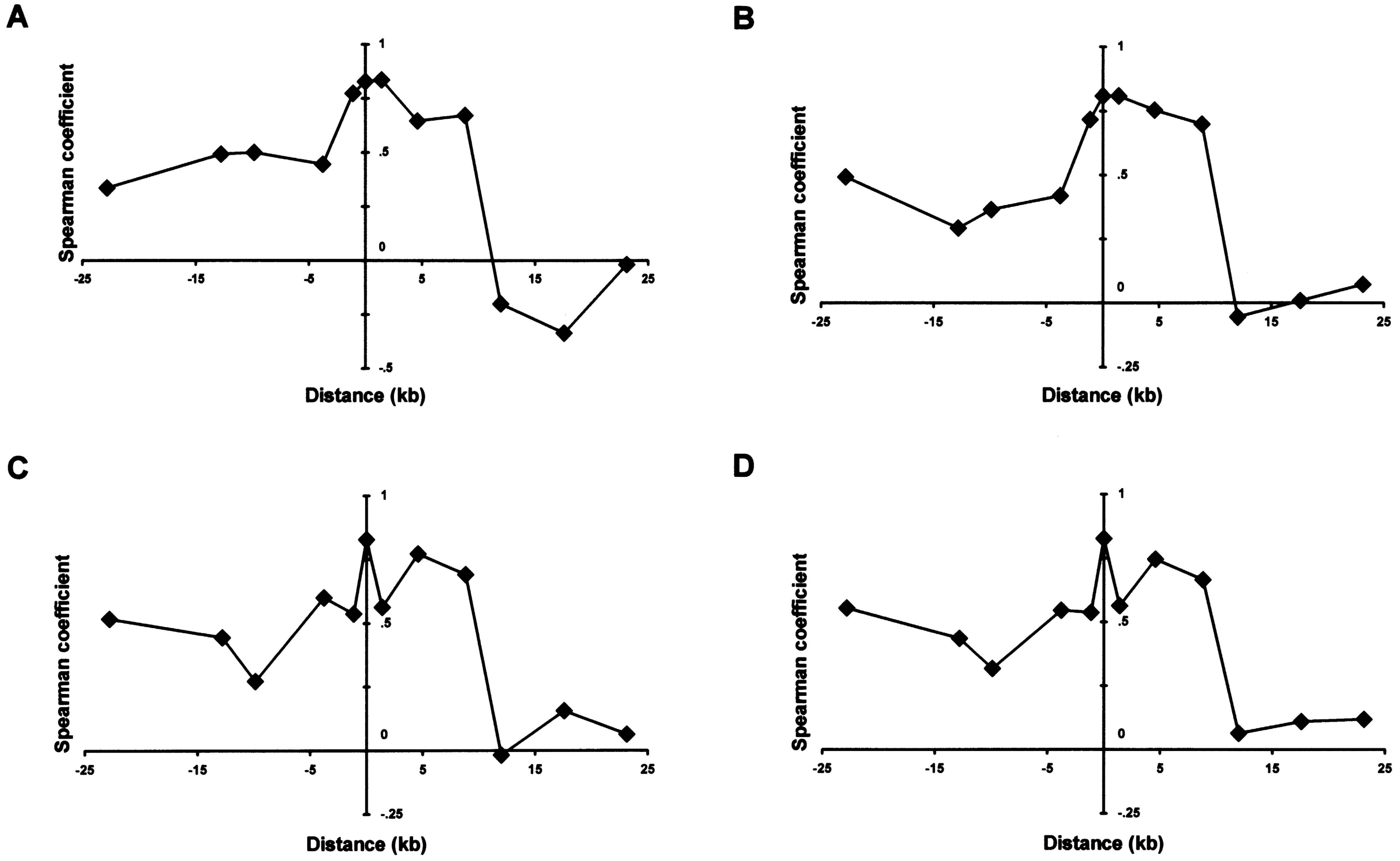

Figure 5 Rank correlation between LD and association with phenotype by varying the location of the functional polymorphism. Each marker was sequentially designated as the single functional polymorphism, and LD to each of the surrounding markers was calculated using $|D'|$ (A), t|LOR| (B), d^2 (C), and Δ^2 (D). The zero point of the *X*-axis is the position of −1021C→T.

great deal more effort to characterize, including extensive resequencing in multiple individuals. The *DBH* locus appears to fall within the latter category, since its full-length haplotype structure is complex, and will require further examination with additional markers to be fully delineated. Finally, it is important to note that the block-structure model does not preclude the occurrence of relatively strong residual LD between multiple blocks or between blocks and individual markers within intervening regions. As evidence of this, Daly et al. (2001) observed clear long-range LD among blocks over a 500-kb region. This suggests that some markers occurring outside of blocks containing functional polymorphism(s) might still be associated with the trait of interest. For example, in the present study, though −4784-4803del occurred outside the block containing −1021C→T, it was still in relatively strong LD with this putative functional SNP and was highly associated with plasma DBH activity ($P = 3.0 \times 10^{-10}$).

A variety of LD measures with differing properties based on Lewontin's *D* are in use today (Devlin and Risch 1995). $|D'|$ is perhaps the most commonly encountered measure, and its scale (0–1) is independent of allele frequency (Lewontin 1988), which has often been considered a desirable property. Thus, $|D'| = 1$ for two markers with divergent allele frequencies, provided that the two have not been separated by recombination. We chose to display our results by using $|D'|$ in figure 2, to facilitate comparison with other studies. However, $|D'|$ has several undesirable characteristics, including a sensitivity to sample size and a tendency to overestimate the magnitude of LD (Frisse et al. 2001; Ardlie et al. 2002), which is especially apparent when at least one of the variants studied is rare and the opportunities for observing recombination are consequently limited. Several authors have proposed the use of alternate LD measures in the context of association studies, including d^2 and Δ^2 (also denoted as r^2) (Kruglyak 1999; Ardlie et al. 2002; Weiss and Clark 2002). Both d^2 and Δ^2 range from 0 to 1 and attain maximum value only when two markers have not been separated by recombination and have identical allele frequencies, such that only two of the four possible haplotypes are observed. In this case, genotype at one marker perfectly predicts genotype at the other, rendering one redundant. Thus, one would predict that a marker and trait allele in "absolute LD" should be similarly associated with a disease or quantitative trait when d^2 or $\Delta^2 = 1$ but not necessarily when $|D'| = 1$. Our findings in the current study illustrate this concept. For example, $|D'| = 1$ between −1021C→T and the two closest adjacent markers, IVS1+109G→C and −2124C→T, but, because of divergent allele frequencies, the corresponding values of d^2 were much lower (.13 and .15, respectively; fig. 3). As expected, these two markers explained a much smaller proportion of phenotypic variance ($R^2 = .11$ for both) than did the putative functional SNP itself ($R^2 = .46$; table 3). Thus, the concept of "absolute LD" as defined by $|D'| = 1$ can be quite misleading in the context of association mapping. In our analysis, d^2 more accurately predicted the phenotypic association of each marker than $|D'|$ did. This was evident from the following data: (1) R^2, graphed on a separate axis with a maximum value equal to the R^2 for −1021C→T (0.46), paralleled a plot of d^2 much more closely than $|D'|$ (fig. 3); (2) the best-fit line for d^2 versus R^2 passed much closer to the data point predicted for a hypothetical marker in absolute LD with −1021C→T than did the corresponding line for $|D'|$ (fig. 4); and (3) in figure 5, the highest Spearman correlation coefficient occurred at −1021C→T for d^2 but not for $|D'|$. Thus, scaling *D* to remove allele frequency effects loses potentially valuable information. In general, Δ^2 performed similarly to d^2 (figs. 3 and 5), but the best-fit line for Δ^2 in figure 4 did not pass as close to the "absolute LD" data point. Absolute LD where $d^2 = 1$ is a useful concept to consider in several other situations. For example, as d^2 approaches 1.0 for a set of SNP markers, any one of the markers can be used as an htSNP. In the present study, $d^2 = 0.94$ and 0.88 (without specifying a trait allele, two values are possible) between −2124C→T and IVS1+109G→C. Inspection of table 2 reveals that either of these two SNPs can be used as one of the htSNPs for the haplotype block. Similarly, statistical methods cannot distinguish between a marker and trait allele when $d^2 = 1$; molecular techniques to directly assay function are required.

Edwards (1963) found that, of several measures of association, only those derived from the odds ratio, including LOR and Yule's Q, were independent of the marginal allele frequencies in a two-way table. Thus, we were interested in comparing the performance of one of these measures against that of $|D'|$. We chose LOR, but, since this has an infinite range, we have considered the arctan transformation of the absolute value of the LOR, which, when divided by $\pi/2$, can be restricted to the interval (0,1). The t|LOR| closely resembled $|D'|$ in its pattern of decay with physical distance (fig. 3), prediction of marker association with phenotype (fig. 4), and localization of the putative functional polymorphism (fig. 5). The maximum value of t|LOR| observed for any single marker pair was .92 (data not shown).

The present work lends further indirect support to the putative functional role of −1021C→T in *DBH* gene expression but, as with all statistical approaches, requires confirmation by molecular genetic studies. Unfortunately, direct molecular evidence is still lacking, since transient-transfection assays of reporter gene constructs in human neuroblastoma cell lines designed to assess whether −1021C→T directly alters transcriptional activation of the *DBH* gene have been negative to date (K. S. Kim, personal communication). However, although studies using in vitro reporter gene techniques have identified key regulatory motifs only within the proximal 400 bp of the *DBH* promoter (Ishiguro et al. 1993; Kim et al. 1998), in vivo reporter gene experiments in transgenic mice have demonstrated that a critical positive regulatory element(s) exists somewhere between −600 bp and −1,100 bp upstream of the transcriptional start site (Hoyle et al. 1994). This suggests that such in vitro assays are likely biased in favor of detecting the effects of only the most proximal promoter elements and that in vivo experiments will be required to definitively assess the functionality of −1021C→T.

The results of our multiple regression analysis argue against additional functional effects for any of the 11 markers flanking −1021C→T. It is possible, however, that the strong association between −1021C→T and plasma DBH activity is a result of tight LD with an undiscovered functional polymorphism(s) located outside of the DBH coding region, proximal 2.6 kb of 5′ flanking sequence, or intron-exon boundaries we previously examined (Zabetian et al. 2001). Our findings here suggest that, if such a cryptic variant exists, it is likely located within or very near the LD block containing −1021C→T, thus limiting future searches for alternate/additional functional candidate polymorphisms to specific regions of the *DBH* gene. Regardless of whether −1021C→T is actually functional, our results indicate that the underlying LD structure of the *DBH* gene strongly influences marker-phenotype association and demonstrate the potential utility of a genomewide haplotype resource in future mapping studies of quantitative traits.

Acknowledgments

We are grateful to Laura Heizman, Regis Hoppenot, and David M. Ocame, for technical assistance, and to Adam Frost, for preliminary analyses and useful discussions. This research was supported by grants from the National Alliance for Research on Schizophrenia and Depression (to J.F.C.); National Heart, Lung, and Blood Institute grant T32-HL07567 (to Case Western Reserve University); the National Institute on Drug Abuse (Substance Abuse Fellowship 5T32-DA-07238 [to C.P.Z.] and Yale Physician Scientist Training Program grants K-12-DA-00167 and R01-DA12422 [both to J.F.C.]); National Institute on Alcohol Abuse and Alcoholism grant R01-AA11330 (to J.G.); National Institute of Mental Health grants MH 30929 (to Yale University), K02-MH01387 (to J.G.), and T32-MH2005 (to the University of Pittsburgh); National Institute of Neurological Disorders and Stroke grant K08-NS44138 (to C.P.Z.); Tübingen University fortuene-fund f-490 (to M.D.K); the U.S. Department of Veterans Affairs (Psychiatric Research/Neurosciences Fellowship [to C.P.Z.], Presidential Early Career Award for Scientists and Engineers [to J.F.C.], Merit Review [to J.G.], and the Mental Illness Research Education and Clinical Center [to J.F.C., J.G., and C.P.Z.]); U.S. Public Health Service National Institute of General Medical Sciences grant GM 28356 (to R.C.E.); and National Center for Research Resources grant RR 03655 (to R.C.E.).

Electronic-Database Information

The URLs for data presented herein are as follows:

dbSNP Home Page, http://www.ncbi.nlm.nih.gov/SNP/
Kidd Lab, Computer Programs, http://info.med.yale.edu/genetics/kkidd/programs.html (for HWSIM)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for DBH)

References

Abecasis GR, Noguchi E, Heinzmann A, Traherne JA, Bhattacharyya S, Leaves NI, Anderson GG, Zhang Y, Lench NJ, Carey A, Cardon LR, Moffatt MF, Cookson WO (2001) Extent and distribution of linkage disequilibrium in three genomic regions. Am J Hum Genet 68:191–197

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3: 299–309

Cubells JF, Kobayashi K, Nagatsu T, Kidd KK, Kidd JR, Calafell F, Kranzler HR, Ichinose H, Gelernter J (1997) Population genetics of a functional variant of the dopamine β-hydroxylase gene (DBH). Am J Med Genet 74:374–379

Cubells JF, van Kammen DP, Kelley ME, Anderson GM, O'Connor DT, Price LH, Malison R, Rao PA, Kobayashi K, Nagatsu T, Gelernter J (1998) Dopamine β-hydroxylase: two polymorphisms in linkage disequilibrium at the structural gene DBH associate with biochemical phenotypic variation. Hum Genet 102:533–540

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29:229–232

Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29:311–322

Edwards AWF (1963) The measure of association in a 2 × 2 table. J R Stat Soc Ser A 126:109–114

Frisse L, Hudson RR, Bartoszewicz A, Wall JD, Donfack J, Di Rienzo A (2001) Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am J Hum Genet 69: 831–843

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296: 2225–2229

Goldin LR, Gershon ES, Lake CR, Murphy DL, McGinniss M, Sparkes RS (1982) Segregation and linkage studies of plasma dopamine-beta-hydroxylase (DBH), erythrocyte catechol-O-methyltransferase (COMT), and platelet monoamine oxidase (MAO): possible linkage between the ABO locus and a gene controlling DBH activity. Am J Hum Genet 34:250–262

Haldane JBS (1955) The estimation and significance of the logarithm of a ratio of frequencies. Ann Hum Genet 20: 309–311

Hill WG, Weir BS (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. Am J Hum Genet 54:705–714

Hoyle GW, Mercer EH, Palmiter RD, Brinster RL (1994) Cell-specific expression from the human dopamine beta-hydroxylase promoter in transgenic mice is controlled via a combination of positive and negative regulatory elements. J Neurosci 14:2455–2463

Ishiguro H, Kim KT, Joh TH, Kim KS (1993) Neuron-specific expression of the human dopamine beta-hydroxylase gene requires both the cAMP-response element and a silencer region. J Biol Chem 268:17987–17994

Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29:217–222

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29:233–237

Jorde LB, Watkins WS, Carlson M, Groden J, Albertsen H, Thliveris A, Leppert M (1994) Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. Am J Hum Genet 54:884–898

Kim HS, Yang C, Kim KS (1998) The cell-specific silencer region of the human dopamine β-hydroxylase gene contains several negative regulatory elements. J Neurochem 71:41–50

Köhnke MD, Zabetian CP, Anderson GM, Kolb W, Gaertner I, Buchkremer G, Vonthein R, Schick S, Lutz U, Köhnke AM, Cubells JF (2002) A genotype-controlled analysis of plasma dopamine β-hydroxylase in healthy subjects and alcoholics: evidence for alcohol-related differences in noradrenergic function. Biol Psychiatry 52:1151–1158

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139–144

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49: 49–67

——— (1988) On measures of gametic disequilibrium. Genetics 120:849–852

Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, Finch KL, Stevens JF, Livak KJ, Slotterbeck BD, Slifer SH, Warren LL, Conneally PM, Schmechel DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM (2000) SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. Am J Hum Genet 67:383–394

Nahmias J, Burley MW, Povey S, Porter C, Craig I, Wolfe J (1992) A 19 bp deletion polymorphism adjacent to a dinucleotide repeat polymorphism at the human dopamine beta-hydroxylase locus. Hum Mol Genet 1:286

Nei M, Li WH (1980) Non-random association between electromorphs and inversion chromosomes in finite populations. Genet Res 35:65–83

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719–1723

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199–204

Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405:847–856

Roses AD (2000) Pharmacogenetics and the practice of medicine. Nature 405:857–865

Schneider S, Roessli D, Excoffier L (2000) Arlequin version 2.000: a software for population genetics data analysis. University of Geneva, Geneva, Switzerland

Schork NJ (2002) Power calculations for genetic association studies using estimated probability distributions. Am J Hum Genet 70:1480–1489

Schork NJ, Nath SK, Fallin D, Chakravarti A (2000) Linkage disequilibrium analysis of biallelic DNA markers, human quantitative trait loci, and threshold-defined case and control subjects. Am J Hum Genet 67:1208–1218

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, et al (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293: 489–493

Tiret L, Poirier O, Nicaud V, Barbaux S, Herrmann SM, Perret C, Raoux S, Francomme C, Lebard G, Tregouet D, Cambien F (2002) Heterogeneity of linkage disequilibrium in human genes has implications for association studies of common diseases. Hum Mol Genet 11:419–429

Wei J, Ramchand CN, Hemmings GP (1997) Possible control of dopamine β-hydroxylase via a codominant mechanism associated with the polymorphic (GT)n repeat at its gene locus in healthy individuals. Hum Genet 99:52–55

Weinshilboum R, Axelrod J (1971) Serum dopamine-beta-hydroxylase activity. Circ Res 28:307–315

Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. Trends Genet 18:19–24

Wilson AF, Elston RC, Siervogel RM, Tran LD (1988) Linkage of a gene regulating dopamine-beta-hydroxylase activity and the ABO blood group locus. Am J Hum Genet 42:160–166

Zabetian CP, Anderson GM, Buxbaum SG, Elston RC, Ichinose H, Nagatsu T, Kim KS, Kim CH, Malison RT, Gelernter J, Cubells JF (2001) A quantitative-trait analysis of human plasma-dopamine β-hydroxylase activity: evidence for a major functional polymorphism at the DBH locus. Am J Hum Genet 68:515–522

Zapata C, Carollo C, Rodriguez S (2001) Sampling variance and distribution of the D' measure of overall gametic disequilibrium between multiallelic loci. Ann Hum Genet 65: 395–406

Zhao JH, Curtis D, Sham PC (2000) Model-free analysis and permutation tests for allelic associations. Hum Hered 50: 133–139