Am. J. Hum. Genet. 53:1038–1050, 1993

Linkage Disequilibrium in the Neurofibromatosis 1 (NF1) Region: Implications for Gene Mapping

L. B. Jorde,* W. S. Watkins,* D. Viskochil,† P. O'Connell,§ and K. Ward‡

Departments of *Human Genetics, †Pediatrics, and ‡Obstetrics and Gynecology, University of Utah Health Sciences Center, Salt Lake City; and §Department of Pathology, University of Texas Health Science Center, San Antonio

Summary

To test the usefulness of linkage disequilibrium for gene mapping, we compared physical distances and linkage disequilibrium among eight RFLPs in the neurofibromatosis 1 (NF1) region. Seven of the polymorphisms span most of the NF1 gene, while the remaining polymorphism lies approximately 70 kb 3′ to a stop codon in exon 49. By using Centre d'Étude du Polymorphisme Humain (CEPH) kindreds, 91–110 unrelated parents were genotyped. A high degree of disequilibrium is maintained among the seven intragenic polymorphisms ($r > .82$, $P < 10^{-7}$), even though they are separated by as much as 340 kb. The 3′ polymorphism is only 68 kb distal to the next polymorphism, but disequilibrium between the 3′ polymorphism and all others is comparatively low ($|r| < .33$, P values .27–.001). This result was replicated in three sets of unrelated kindreds: the Utah CEPH families, the non-Utah CEPH families, and an independent set of NF1 families. Trigenic, quadrigenic, three-locus, and four-locus disequilibrium measures were also estimated. There was little evidence of higher-order linkage disequilibrium. As expected for a disease with multiple mutations, no disequilibrium was observed between the disease gene and any of the RFLPs. The observed pattern of high disequilibrium within the gene and a loss of disequilibrium 3′ to the stop codon could have implications for gene mapping studies. These are discussed, and guidelines for linkage disequilibrium studies are suggested.

Introduction

Linkage studies are frequently plagued by a lack of recombinations among closely linked polymorphisms. When recombination fractions are less than .01, the number of informative meioses needed to establish gene order tends to become prohibitively large. Linkage disequilibrium, which reflects the effects of recombination over many past generations, is sometimes used as an alternative. Disequilibrium analysis has been used, for example, in studies of cystic fibrosis (Estivill et al. 1987), Huntington disease (Theilmann et al. 1989; MacDonald et al. 1991; Skraastad et al. 1992), Friedreich ataxia (Hanauer et al. 1990), myotonic dystrophy (Harley et al. 1991), adult polycystic kidney disease (Pound et al. 1992), dystonia-Parkinsonism syndrome (Kupke et al. 1992), torsion dystonia (Ozelius et al. 1992), and diastrophic dysplasia (Hästbacka et al. 1992).

The use of linkage disequilibrium in gene mapping assumes a predictable relationship between disequilibrium and genetic distance between polymorphisms. Theoretically, these two quantities have an inverse relationship of the form $r^2 = 1/(1 + 4N_e c)$, where r is a standard measure of linkage disequilibrium, N_e is effective population size, and c is the recombination fraction (Hill and Robertson 1968; Sved 1971). However, this relationship is expected to hold only when $4N_e c$ is fairly large. Thus, a monotonic relationship between disequilibrium and distance may not necessarily exist in small genomic segments, where the effects of recombination may be overcome by factors such as drift, mutation, admixture, and gene conversion (Litt and Jorde 1986; Templeton et al. 1987; Lalouel and Jorde 1988). In addition, the variance of linkage disequilibrium statistics becomes large when $N_e c$ is small (Hudson 1985; Hill

Received April 30, 1993; revision received July 1, 1993.

Address for correspondence and reprints: Dr. L. B. Jorde, Department of Human Genetics, 2100 Eccles Institute of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, UT 84112.

0002-9297/93/5305-0008$02.00

and Weir 1988), and the statistical power to detect disequilibrium can be quite low (Thompson et al. 1988). Even if recombination and disequilibrium are closely related, it is clear that recombination rates are not uniform across the genome (Steinmetz et al. 1987). Because of these considerations, some have questioned the utility of linkage disequilibrium for gene mapping.

The empirical evidence on this issue is conflicting. Some studies have found a fairly uniform inverse relationship between physical distance and disequilibrium in small genomic regions (Bech-Hansen et al. 1983; Aschbacher et al. 1985; Chakravarti et al. 1986; Chakraborty et al. 1987; Daiger et al. 1989; Leitersdorf et al. 1989; Elbein 1992), while others have not (Barker et al. 1984; Chakravarti et al. 1984*c*; Litt and Jorde 1986; Børresen et al. 1988; Thompson et al. 1988; Tzall et al. 1989; Hegele et al. 1990; Benlian et al. 1991; Haviland et al. 1991; Walter and Cox 1991; Zerba et al. 1991; Miserez et al. 1993). Nonuniform disequilibrium patterns have led some investigators to infer the presence of recombinational "hot spots" (Chakravarti et al. 1984*a*, 1986; Migone et al. 1985; Steinmetz et al. 1987) and "cold spots" (Murray et al. 1984). For the reasons cited above, the inference of hot spots from disequilibrium patterns has been criticized (Weir and Hill 1986; Hedrick 1987, 1988).

Some of the inconsistencies regarding disequilibrium and physical distance can be attributed to small sample sizes, poorly characterized populations, and inadequate statistical techniques. To clarify whether linkage disequilibrium may be useful in gene mapping studies, further empirical tests are needed. In this paper, we present the results of a study of linkage disequilibrium and physical distance among eight RFLPs in the neurofibromatosis 1 (NF1) region. The sample, generated from the Centre d'Étude du Polymorphisme Humain (CEPH) collection, is large and well characterized. We show that our results are replicated in two subsets of the CEPH collection and in an independent sample of NF1 families. Several different analytic measures of disequilibrium are used, and patterns of higher-order disequilibrium are explored.

Material and Methods

Study Subjects

All polymorphisms were typed in 60 CEPH kindreds. All generations of each kindred were typed so that phase-known haplotypes could be scored. Only the haplotypes from unrelated parents are used in the disequilibrium analysis. Since members of several CEPH kindreds are relatives (kindreds 13291, 13292, 13293, and 13294), only one of these kindreds, 13294, was analyzed. This provided a maximum of 228 phase-known haplotypes, although the number of two-locus haplotypes that could be successfully typed and inferred was 138–210 (most were in the 160–190 range).

The CEPH collection includes 43 Utah kindreds as well as French, Venezuelan, and Amish kindreds. Since linkage disequilibrium statistics can be affected by population structure, the Utah and non-Utah kindreds were analyzed both separately and together.

A second data set consisted of 17 NF1 families that were typed for five polymorphisms in the NF1 region. Sixty phase-known haplotypes were scored in unrelated members of these families. Among these haplotypes, 17 carried an NF1 disease mutation. The NF1 kindreds, which include Caucasian, Hispanic, and Chinese families, are ethnically more diverse than the CEPH collection.

To assess the age of the polymorphisms, several systems were also typed in 10 Asians and 10 Africans. These DNA samples were purchased from the Coriell Institute, Camden, NJ.

Laboratory Methods

CEPH, Asian, and African DNAs were prepared from lymphoblast cell lines according to the method of Bell et al. (1981). Restriction enzyme digests were performed according to the manufacturer's (Molecular Biology Resources) instructions, using twofold excess enzyme. Restriction enzyme–digested samples (5 μg) were fractionated in 0.8% agarose gels in 2 × Tris-acetate buffer. Gels were denatured in 0.4 N NaOH for 30 min, transferred in 0.4 N NaOH to Hybond-N+ (Amersham) for 8–12 h, and neutralized in 2 × SSC. Membranes were prehybridized in 10% polyethylene glycol, 8% SDS, and 200 μg total human DNA/ml for 4 h and hybridized overnight with a radiolabeled probe at 65°C. Final stringency washes were 0.1 × SSC, 0.1% SDS at 65°C. Radiolabeled probes were prepared by random priming using standard methods (Feinberg and Vogelstein 1984). Probes f7G4-GL.2, fHB5-E.3, fPL37-B.2, and T315.5 were isolated from whole cosmid and were labeled in low-melting-temperature agarose (SeaPlaque, FMC).

Eight probes detecting two-allele RFLPs of high heterozygosity were used to genotype individuals from 60 two- and three-generation CEPH families. Clone pHu39.3 and probe f7G4-GL.2 have been described

elsewhere (O'Connell et al. 1990; Watkins et al. 1991). Clone p11-1F10 is derived from cosmid c11-1F10 (O'Connell et al. 1989*a*) and detects two polymorphic *Pst*I alleles of 3.0 and 2.0 kb. To identify additional RFLPs at the NF1 locus, whole cosmids were used to screen a panel of six random individuals for polymorphisms, with nine common restriction enzymes.

Physical distances between polymorphisms were determined by restriction mapping. Sequence data and existing map data were both used to deduce the locations of restriction fragments.

Statistical Analysis

For each pair of diallelic loci, A and B, the linkage disequilibrium coefficient, D, was measured as $D = P_{11} - p_1q_1$, where P_{11} is the haplotype frequency of the two more common alleles at A and B, and p_1 and q_1 are the gene frequencies of these alleles (Lewontin and Kojima 1960). A positive value of D thus indicates that the two common alleles tend to co-occur on the same chromosome more often than expected, while a negative value of D indicates that the common allele at one locus and the rarer allele at the second locus tend to co-occur. Following Hill and Robertson (1968), it is common to estimate a linkage disequilibrium coefficient, r, which is standardized by gene frequencies and varies between −1.0 and 1.0, as

$$r = \frac{D}{\sqrt{p_1q_1p_2q_2}}.$$

Ninety-five–percent confidence limits for r were obtained using Fisher's z-transformation. Non-normality of gene frequencies may in principle violate the assumptions underlying the use of this method. However, z-transformations have been shown to work well even with small sample sizes where non-normality tends to be greatest (Weir 1979). The confidence limits provide a convenient test of the null hypothesis that $r = 0.0$. In addition, a χ^2 test of the null hypothesis is given by $\chi^2 = Nr^2$, where N is the number of haplotypes in the sample.

Although r is defined to be within the range −1.0–1.0, its true range as a measure of linkage disequilibrium is dictated by the allele frequencies. Thus, it can be misleading to compare r values when one value is close to its maximum while another is not. In particular, Hedrick (1987) showed that frequency-dependent measures such as D and r can lead one to falsely infer the presence of a recombinational hot spot. Accordingly, D was also compared to its maximum possible value, D_{max}, which is given by $\min(p_1q_2, p_2q_1)$ (Lewontin 1964). Then, $D' = D/D_{max}$. Lewontin (1988) has shown that even this measure is not strictly independent of gene frequencies.

In addition to estimating standard gametic disequilibrium coefficients, genotypic disequilibrium estimates were obtained (Weir 1990). These calculations were first done by using the genotypes of only those subjects who could be haplotyped and later by using all subjects.

Trigenic and quadrigenic disequilibria were evaluated using the procedures of Weir and Cockerham (1989). Trigenic disequilibrium measures association between one allele at locus A and both alleles at locus B, after adjusting for standard digenic disequilibrium and any departure from Hardy-Weinberg equilibrium. There are thus two trigenic disequilibrium coefficients, D_{ABB} and D_{AAB}. Similarly, quadrigenic disequilibrium, Δ_{AABB}, measures association between both alleles at both loci, after adjusting for Hardy-Weinberg departures and digenic and trigenic disequilibrium. Computational details, as well as methods for obtaining the variances of these measures, are given in Weir and Cockerham (1989).

Disequilibrium was also estimated for the three- and four-locus cases (Weir 1990). D_{ABC} measures disequilibrium for loci A, B, and C after adjusting for two-locus disequilibrium. Similarly, D_{ABCD} measures disequilibrium for loci A, B, C, and D after adjusting for two- and three-locus disequilibrium. Further details are given in Weir (1990).

Traditional disequilibrium statistics may not be appropriate in disease families, since they represent a selected sample. Accordingly, conditional disequilibrium (Chakravarti et al. 1984*b*) was estimated for the NF1 genotype versus other marker loci in the NF1 disease families.

All possible pairs of the eight loci were considered, giving a total of 28 pairwise disequilibrium coefficients. With this number of comparisons, it is likely that at least one r value would be "significant" at the .05 level simply by chance. Since the null hypothesis is that there is no disequilibrium for any pair of loci, the familiar Bonferonni procedure (Weir 1990) was used to adjust for multiple comparisons. An adjusted significance level, α', of .05, is given by $\alpha = 1 - (1 - \alpha')^{1/28}$. Thus, the required significance level becomes .0018. While we regard this procedure as appropriately conservative, some investigators feel that it unnecessarily increases the probability of type II error (Zerba et al. 1991).

Physical distance and pairwise disequilibrium were compared in order to test the hypothesis that there is a uniform negative relationship between distance and disequilibrium. The distance values and disequilibrium coefficients each form an 8×8 matrix whose elements are not independent of one another. The Mantel matrix comparison technique (Mantel 1967; Smouse et al. 1986) takes this lack of independence into account. An empirical distribution of association coefficients for the two matrices is generated by randomly permuting the columns of one of the matrices a given number of times (in this case, 10,000 times). A significance level is then obtained by comparing the actual correlation between the two matrices with the empirical distribution.

Results

Cosmids detecting new RFLPs of high heterozygosity include cHB5, cPL37, cEVI36, and cT315. Fragments and subclone probes were isolated from cosmids and include markers fHB5-E.3, fPL37-B.2, pEVI36.5, and fT315.5. Probe fHB5-E.3 is a 5.5-kb *Eco*RI fragment from cosmid cHB5, which reveals polymorphic *Eco*RI alleles of 5.5 and 4.9 kb. fPL37-B.2 represents a 6.5-kb *Bam*HI fragment derived from a *Bam*HI digest of cosmid cPL37 and detects polymorphic *Bam*HI alleles of 10.0 and 6.5 kb. A 1.2-kb *Eco*RI fragment from cosmid cEVI36 was subcloned into pUC18 to create probe pEVI36.5, which identifies polymorphic *Taq*I alleles of 6.9 and 3.2 kb. fT315.5 is a 1.5-kb *Eco*RI end fragment from cosmid cT315 that detects polymorphic *Eco*RI alleles of 4.2 and 3.8 kb. Subclone pDV1.9 was derived from a genomic clone containing the t(17;22) NF1 translocation and represents a 4.2-kb *Eco*RI fragment that reveals polymorphic *Hin*dIII alleles of 2.0 and 1.6 kb. pDV1.9 also detects RFLPs with other restriction enzymes, and, in each case, allele sizes differ by approximately 400 bp, which indicates a possible insertion or deletion in this region. Probes fT315.5 and pDV1.9 cross-hybridize and identify *Eco*RI alleles of 4.2 and 3.8 kb, which suggests that both markers detect the same DNA alteration. A summary of allele sizes, frequencies, and heterozygosities for each of the eight polymorphisms is presented in table 1. The table also shows how many parents were genotyped for each polymorphism.

The physical distances between the markers at the NF1 locus were determined by restriction mapping. Markers pHu39.3, pEVI36.5, fPL37-B.2, pDV1.9, and fT315.5 were positioned on an NF1 cosmid contig (Viskochil et al. 1990) by using existing restriction map data. Sequence data (GenBank accession no. L03723) were later used to refine probe positions and polymorphic fragments located between the 5′ end of cosmid cEVI36 and the 3′ end of cosmid c7D5. Good concordance was seen between initial and refined distance estimates. To position markers f7G4-GL.2 and p11-1F10, we extended the existing cosmid contig to include cosmids c7G4 and c1F10-12. A 14-kb *Not*I fragment from cosmid c7G4 shows approximately 10 kb of overlap with cosmid c7D5. An overlap of 6.5 kb between cosmids c7G4 and c1F10-12 was established by cross-hybridization and extends the contig distally to include marker p11-1F10. Cross-hybridization between marker p11-1F10 and cosmid c1F10-12 was tested by using several restriction enzymes and indicates the position of p11-1F10 to be about 3–5 kb from the 3′ end of cosmid c1F10-12. The distance between probes f7G4-GL.2 and p11-1F10 was estimated by using restriction-fragment-size data.

The location of the 3′ end of the NF1 gene has not yet been established conclusively (Viskochil et al. 1993). However, a stop codon in exon 49 (fig. 1) lies 15 kb centromeric to the distal end of cosmid c7D5. The f7G4-GL.2 polymorphism maps just 4–8 kb 3′ of the stop codon. The p11-1F10 polymorphism, which is 68 kb 3′ from f7G4-GL.2, lies well beyond the stop codon and most likely represents sequence outside of the NF1 gene.

Probe fHB5-E.3 is not represented in the cosmid contig. However, a relatively accurate estimate of the physical distance between it and other markers can be calculated by using the *Not*I site in cHB5. Cosmid cHB5 spans the proximal *Not*I site of the 360-kb *Not*I pulse-field gel fragment containing the NF1 gene; the polymorphic *Eco*RI site detected by probe fHB5-E.3 lies 10 kb distal to the 5′ *Not*I site (J. Stevens, personal communication).

Since it was not always possible to determine unambiguously the location of the polymorphic site within each restriction fragment, distances between pairs of RFLPs are given as a narrow range. The locations of probes and midpoint distance estimates between polymorphic sites are shown in figure 1.

Nine distinct haplotypes were observed in the CEPH kindreds. These haplotypes and their frequencies in the Utah and non-Utah CEPH kindreds are shown in table 2. Several additional haplotypes with unscored genotypes for the most proximal and distal polymorphisms are also shown here. One haplotype, 12112221, was

Table 1

Probe-Enzyme Combinations, Allele Sizes and Frequencies, Heterozygosity, and Sample Size (No. of Typed Individuals) for Each Polymorphism

Probe	Enzyme	Allele Sizes (kb)	Allele Frequencies	Heterozygosity	*N*
fHB5-E.3	*Eco*RI	5.5/4.9	.69/.31	.37	97
pHu39.3	*Eco*RI	9.6/7.2	.62/.38	.42	108
pEVI36.5	*Taq*I	6.9/3.2	.63/.37	.43	91
fPL37-B.2	*Bam*HI	10.0/6.5	.60/.40	.42	95
pDV1.9	*Hin*dIII	2.0/1.6	.35/.65	.41	106
fT315.5	*Eco*RI	4.2/3.8	.38/.62	.49	92
f7G4-GL.2	*Bgl*II	4.1/3.3	.36/.64	.41	101
p11-1F10	*Pst*I	3.0/2.0	.48/.52	.43	110

seen only once. It is most easily attributed to a single crossover occurring between the pHu39.3 and cEVI36 polymorphisms of haplotypes 1 and 4.

None of the eight systems showed significant departures from Hardy-Weinberg equilibrium. All of the systems tested in Africans and Asians (fHB5-E.3, fPL37-B.2, f7G4-GL.2, and p11-1F10) were polymorphic in these populations, which indicates that they probably predate the divergence of the major human races. No obligate recombinants were observed in the CEPH families.

Pairwise disequilibrium coefficients, *r*, and their 95% confidence limits, are given in table 3. The *r* values are given above the diagonal of the table, and pairwise physical distances are given below the diagonal. The table shows that the six internal polymorphisms are all in extremely tight disequilibrium with one another. In fact, five of the six systems are in complete disequilibrium with one another ($r = 1.0$), while the sixth (pHu39.3) yields *r* values exceeding .985 with the other five systems. Pairwise values for the fHB5-E.3 polymorphism, which is 260 kb 5′ to the next polymorphism, are slightly lower but still very high and statistically significant. The p11-1F10 polymorphism, which is only 68 kb

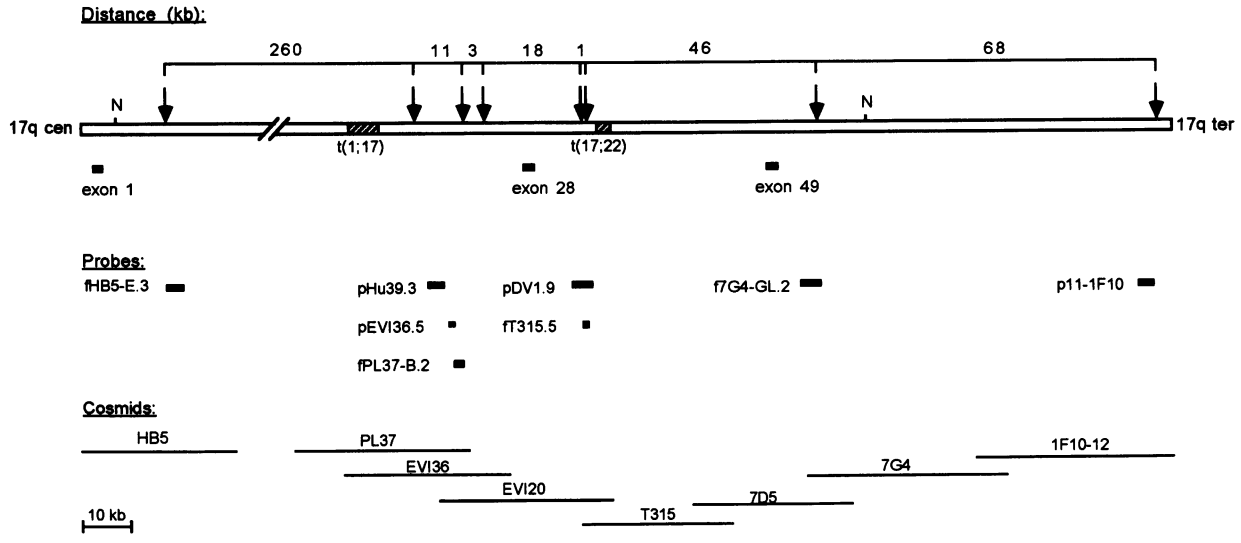

Figure 1 Map of the NF1 region, showing the locations of each of the eight RFLPs, several exons of the NF1 gene, and a cosmid contig. Only the portion of exon 49 5′ to the stop codon is shown. Also shown are the locations of previously described translocation breakpoints t(1;17) and t(17;22). N = *Not*I.

Table 2

NF1-Region Haplotype Frequencies

Haplotype No. and Composition	No. in Utah Kindreds	No. in Non-Utah Kindreds
1. 11112221	61	12
2. 11112222	38	6
3. 12112221	1	0
4. 12221111	2	0
5. 12221112	10	3
6. 22221111	12	8
7. 22221112	28	7
8. 01112221	5	5
9. 01112222	7	6
10. 02221112	2	1
11. 22221110	2	0

NOTE.—For haplotypes in which a genotype for systems 2–7 was unknown, the genotype was inferred on the basis of the observed complete disequilibrium with other polymorphisms.

3′ to the other polymorphisms, exhibits much lower disequilibrium values. As the 95% confidence limits indicate, the disequilibrium values for the p11-1F10 comparisons are significantly lower than those of the other seven systems. After the Bonferroni correction for multiple comparisons is applied, all of the pairwise comparisons except those involving p11-1F10 are highly significant, with corrected significance values less than 10^{-7}. Among the comparisons involving p11-1F10, one corrected value (p11-1F10 vs. fHB5-E.3) was not significant, while the other corrected significance values had a range of .03–.001.

The D' values for these comparisons show a very similar pattern. All comparisons except those involving p11-1F10 yielded D' values of .90 or greater (most were 1.0). The D' values for comparisons involving p11-1F10 had a range of .20–.40.

The disequilibrium estimates based on genotypic data yielded patterns nearly identical to those based on haplotype data. This was true both for the genotypes from only the haplotyped subjects as well as for the total set of genotypes. This result is expected, since all systems were in Hardy-Weinberg equilibrium and since it was possible to haplotype the great majority of subjects.

Table 4 presents r values both for the Utah kindreds only (above the diagonal) and for the non-Utah kindreds (below the diagonal). The patterns observed in these two populations are remarkably similar to those seen in table 2 for the whole population. To test further the hypothesis that haplotype distributions (and hence disequilibrium patterns) are similar in these two populations, a multicolumn extension of Fisher's exact test was applied to the haplotype distributions shown in table 2 (only the nonzero haplotypes were used). As expected, the test indicated that the two haplotype distributions do not differ significantly ($P > .29$). These results are consistent with genetic distance analyses that show that the Utah population is genetically representative of European-derived populations (McLellan et al. 1984).

Table 5 presents pairwise disequilibrium values for the NF1 family data set. Once again, the pattern for the five polymorphisms typed here is very similar to that observed in the CEPH kindreds. Virtually no disequilibrium is observed between NF1 itself and any of the

Table 3

Linkage Disequilibrium Coefficients (r) for All CEPH Kindreds (above the Diagonal) and Physical Distances (in kb, below the Diagonal) for Each Pair of Loci

	fHB5-E.3	pHu39.3	pEVI36.5	fPL37-B.2	pDV1.9	fT315.5	f7G4-GL.2	p11-1F10
fHB5-E.3		.85 (.80, .89)	.85 (.80, .89)	.83 (.77, .87)	.83 (.78, .87)	.86 (.82, .90)	.82 (.77, .87)	.18 (.05, .32)
pHu39.3	247–273		1.00	.99 (.98, .99)	.99 (.98, .99)	.99 (.98, .99)	.99 (.98, .99)	−.24 (−.10, −.37)
pEVI36.5	258–284	9–13		1.00	1.00	1.00	1.00	−.33 (−.18, −.46)
fPL37-B.2	261–287	12–16	1–5		1.00	1.00	1.00	−.26 (−.12, −.40)
pDV1.9	279–305	30–34	19–23	16–20		1.00	1.00	−.28 (−.15, −.40)
fT315.5	280–306	31–35	20–24	17–21	0–2		1.00	−.25 (−.11, −.38)
f7G4-GL.2	325–353	76–82	65–71	62–68	44–50	43–49		.27 (.14, .40)
p11-1F10	391–423	142–152	131–141	134–144	110–120	109–119	62–74	

NOTE.—Ninety-five-percent confidence limits are shown after each r value (confidence limits cannot be estimated by Fisher's transformation when $r = 1.0$).

Table 4

Linkage Disequilibrium Coefficients (*r*) for Utah CEPH Kindreds (above the Diagonal) and Non-Utah CEPH Kindreds (below the Diagonal)

	fHB5-E.3	pHu39.3	pEVI36.5	fPL37-B.2	pDV1.9	fT315.5	f7G4-GL.2	p11-1F10
fHB5-E.3		.86 (.81, .90)	.86 (.81, .90)	.83 (.78, .88)	.83 (.77, .87)	.85 (.79, .89)	.82 (.75, .87)	.24 (.08, .38)
pHu39.3	.79 (.58, .90)		1.00	.98 (.98, .99)	.98 (.98, .99)	.98 (.98, .99)	.98 (.98, .99)	−.28 (−.12, −.42)
pEVI36.5	−.78 (−.46, −.92)	1.00		1.00	1.00	1.00	1.00	−.33 (−.17, −.47)
fPL37-B.2	−.78 (−.54, −.90)	1.00	1.00		1.00	1.00	1.00	−.30 (−.14, −.44)
pDV1.9	.85 (.72, .92)	1.00	1.00	1.00		1.00	1.00	−.31 (−.16, −.44)
fT315.5	.93 (.85, .97)	1.00	1.00	1.00	1.00		1.00	−.28 (−.12, −.43)
f7G4-GL.2	−.83 (−.67, −.91)	1.00	1.00	1.00	1.00	1.00		−.31 (−.16, −.45)
p11-1F10	.04 (−.29, .36)	.11 (−.22, .41)	.37 (−.07, .68)	.13 (−.23, .45)	.20 (−.09, .45)	.15 (−.18, .45)	−.16 (−.43, .14)	

NOTE.—Ninety-five-percent confidence limits are shown after each *r* value (confidence limits cannot be estimated by Fisher's transformation when $r = 1.0$).

RFLP systems. The conditional disequilibrium estimates (not shown) also indicated no significant disequilibrium between NF1 and any of the marker loci (all *P* values for the χ^2 estimates were greater than .05).

All of the trigenic and quadrigenic disequilibrium coefficients were very small, and none approached statistical significance. The three-locus disequilibrium analysis yielded 56 different three-locus combinations among eight loci. The χ^2 values for the three-locus comparisons were much smaller than those of the two-locus comparisons, and, after correction for multiple comparisons, only one of them was significant at the .05 level. This involved three of the six internal polymorphisms, pHu39.3, pDV1.9, and f7G4-GL.2. In general, the χ^2 values for comparisons involving the p11-1F10 polymorphism were somewhat lower than those involving other loci. The four-locus disequilibrium analysis yielded 70 comparisons. Again, the χ^2 values are much smaller than those for the two-locus comparisons, and those involving p11-1F10 are smaller still. After correction for multiple comparisons, 20 of the comparisons yielded χ^2 values that could be considered significant (with corrected *P* levels of .05–.005). It is somewhat difficult to discern a recognizable pattern in these associations, although most of them involved the fHB5-E.3 locus and adjacent polymorphisms.

None of the higher-order comparisons reached significance for the non-Utah haplotypes alone. In part, this probably reflects a smaller sample size. For the Utah haplotypes, none of the three-locus comparisons reached significance, and only one of the four-locus comparisons did (systems 1, 4, 5, and 7).

The relationship between physical distance and disequilibrium is shown in figure 2. The upper points in the graph correspond to the pairwise comparisons involving all polymorphisms except p11-1F10. These show the theoretically predicted negative relationship between distance and disequilibrium. Application of

Table 5

Linkage Disequilibrium Coefficients (*r*) for NF1 Families

	NF1	fHB5-E.3	pHu39.3	pDV1.9	f7G4-GL.2	p11-1F10
NF1		.09 (−.21, .36)	.10 (−.18, .37)	.17 (−.12, .42)	.06 (−.21, .32)	−.01 (−.28, .26)
fHB5-E.3			.62 (.39, .78)	.73 (.55, .85)	.75 (.57, .86)	.18 (−.13, .45)
pHu39.3				.91 (.84, .95)	.92 (.86, .95)	.45 (.18, .65)
pDV1.9					1.00	.40 (.11, .62)
f7G4-GL.2						.42 (.15, .63)
p11-1F10						

NOTE.—Ninety-five-percent confidence limits are shown after each *r* value (confidence limits cannot be estimated by Fisher's transformation when $r = 1.0$).

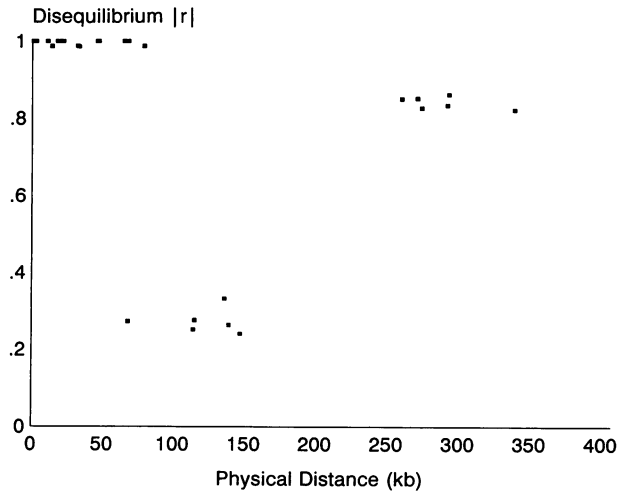

Figure 2 The relationship between physical distance (*X*-axis) and linkage disequilibrium (*Y*-axis) for all CEPH kindred data. For comparability to other studies, the absolute value of the linkage disequilibrium coefficient, *r*, is shown here.

the Mantel matrix comparison test shows that the correlation between physical distance and disequilibrium for these seven loci is large and statistically significant ($r = -.98$, $P < .04$). However, when comparisons involving p11-1F10 (represented by the lower set of points in fig. 2) are included in the matrix comparison test, the distance-disequilibrium correlation becomes nonsignificant ($r = -.27$, $P > .17$).

Discussion

These results show that disequilibrium and physical distance are quite highly correlated for the seven polymorphisms located within the NF1 gene. Virtually complete disequilibrium is maintained among six internal polymorphisms in an 80-kb region that lies within the coding sequence (with the exception of only several kb). A slight decrease in disequilibrium is observed between these polymorphisms and the relatively distant fHB5-E.3 polymorphism. Yet disequilibrium decreases substantially when one proceeds only 68 kb in the 3′ direction to the p11-1F10 polymorphism, which lies well outside the coding region of the gene. Several explanations for this pattern can be considered.

First, the age of the polymorphisms could differ. Less disequilibrium is expected between ancient polymorphisms, since disequilibrium dissipates through time as a result of repeated recombinations. The fact that p11-1F10 shows low disequilibrium with *all* other polymorphisms argues against this interpretation. In addition, the four polymorphisms assayed in Africans and Asians, which span the region under study, were polymorphic in both populations. This means that each polymorphism is relatively ancient, predating the divergence of the major human races.

Second, lower disequilibrium with p11-1F10 could be the result of a substantially higher mutation rate at this locus. This may explain the lower disequilibrium values sometimes observed for some repeat polymorphisms (Elbein 1992) and for polymorphisms recognized by the *Taq*I or *Msp*I enzymes. The p11-1F10 polymorphism, however, is a simple two-allele restriction site polymorphism recognized by *Pst*I.

Third, it could be posited that this result is a statistical artifact resulting from high sampling variance and a lack of statistical power. Several lines of evidence argue against this conclusion. The sample sizes used here are large enough to detect even relatively small amounts of disequilibrium. The formulas devised by Brown (1975) show that 138 haplotypes (the minimum number used here) will detect an *r* value as low as .31 with 95% power (this is a conservative estimate based on two-allele systems with gene frequencies of .70 and .30). The confidence limits around the disequilibrium estimates are small, indicating a high level of precision. Two different measures of disequilibrium, *r* and *D′*, yield the same pattern. While these two measures are correlated, it has been shown that *D′* is less likely to indicate a false recombinational hot spot than is *r* (Hedrick 1987).

A related possibility, one that is more difficult to exclude, is that this result is due to the effects of evolutionary factors—such as drift, admixture, gene conversion, and mutation—in small genomic regions (although mutation is unlikely to have substantial effects on disequilibrium [Carothers and Wright 1992]). These effects, which arise in the history of the population, cannot necessarily be overcome by increasing sample sizes (Kaplan and Weir 1992). However, replication of the same disequilibrium pattern in different populations, with different histories, increases confidence that the pattern is biologically meaningful. In the present study, the same pattern is seen in the Utah CEPH kindreds, the non-Utah CEPH kindreds, and an ethnically diverse sample of NF1 kindreds. In addition, a recent brief report (Messiaen et al. 1993) demonstrated strong disequilibrium among five polymorphisms (not the same ones used here) within the NF1 gene in a series

of Belgian kindreds. Replication of the result in several different populations argues strongly for its reality.

A fourth explanation is that this pattern is the result of low recombination in the pericentromeric region of chromosome 17. It is well known that recombination is relatively rare near centromeres, and previous studies have noted a lack of recombinants 5′ to the NF1 gene (O'Connell et al. 1989*b*, 1993; Ward et al. 1990). While this phenomenon may help to account for the high level of disequilibrium among seven of the polymorphisms, it is somewhat difficult to reconcile with the relatively rapid decrease in disequilibrium for the p11-1F10 polymorphism.

Finally, it seems possible that recombination rates in the region near p11-1F10 may be greater than those within the coding portion of the NF1 gene. This could be interpreted either as a hot spot for recombination near p11-1F10 or as a cold spot for the remainder of the region. The fact that disequilibrium, while lower, is still significant for the p11-1F10 polymorphism seems to argue against a recombinational hot spot. In the 5′ region of the β-globin gene, which may contain a hot spot, disequilibrium values approach zero within 9 kb (Chakravarti et al. 1984*a*). Also arguing against this interpretation is a lack of recombinants involving p11-1F10 in the CEPH kindreds. Additional polymorphisms near the p11-1F10 polymorphism would have been useful in better resolving this issue. However, no other polymorphisms were found near p11-1F10.

It is more striking that there is so little decay of disequilibrium across 340 kb of the NF1 gene. Typically, disequilibrium dissipates more quickly than this. For example, *r* values within the phenylalanine hydroxylase locus are generally less than .50 for polymorphisms separated by 50 kb, and the values approach zero for those separated by 90–100 kb (Chakraborty et al. 1987). Values that range from only .10 to .34 are seen in polymorphisms 35 kb apart or closer in the proα2(I) collagen locus (Børresen et al. 1988). Correlation values drop below .50 for polymorphisms separated by 60–70 kb in the immunoglobulin heavy-chain region on chromosome 14 (Bech-Hansen et al. 1983). Correlations within the 43-kb Apo B gene have a range of .09–.37 (Zerba et al. 1991), and many of these are statistically nonsignificant. Three polymorphisms located within 10 kb of one another in the LDL receptor locus have *r* values of .10–.47 (Hegele et al. 1990). In the insulin receptor locus, Elbein (1992) obtained *D′* values less than .20 for most pairs of polymorphisms separated by more than 60 kb, while the *D′* values in the NF1 region are 1.00 for polymorphisms separated by over 300 kb. Disequilibrium also decreases rapidly with physical distance when extragenic regions are examined (Barker et al. 1984; Litt and Jorde 1986). Significant disequilibrium is maintained between major-histocompatibility-complex class II polymorphisms that are approximately 1 map unit apart, but this is likely to be caused by natural selection (Begovich et al. 1992). It appears, then, that the degree of disequilibrium maintained across the NF1 gene is rather remarkable in comparison with most other measured genomic regions.

Reduced disequilibrium can be caused by sequences thought to be recombinogenic, such as Alu inserts (Vnencak-Jones and Phillips 1990). Sequence analysis of a portion of the NF1 gene has shown that there are at least six Alu inserts located between the pHu39.3 and f7G4-GL.2 polymorphisms, but virtually complete disequilibrium is maintained in this region. Strong disequilibrium has also been observed near Alu sequences in the insulin receptor (Elbein 1992) and LDL receptor loci (Leitersdorf et al. 1989), and Alu repeats near the human growth hormone (GH1) locus are not associated with increased recombination (Vnencak-Jones and Phillips 1990).

Evidence from yeast systems indicates that promoter regions may also be associated with increased recombination (Oliver et al. 1992). Promoters may be located within the NF1 gene, since three smaller genes are embedded within a large intron of NF1 and are transcribed from the opposite strand (Cawthon et al. 1990, 1991; Viskochil et al. 1991). These three genes, *OMGP*, *EVI2A*, and *EVI2B*, are located in the same region as the six internal polymorphisms reported in tight disequilibrium here. If the promoters are located near the embedded genes, the yeast data, if applicable to mammalian systems, would predict increased recombination in this region. The disequilibrium results clearly indicate otherwise.

It has often been suggested that linkage disequilibrium results may be useful for gene mapping and for guiding chromosome walking experiments (Murray et al. 1987; Pandolfo et al. 1990; Kupke et al. 1992). In some cases, such as those of cystic fibrosis and Huntington disease, disequilibrium patterns have been valuable in this regard (Kerem et al. 1989; Huntington's Disease Collaborative Research Group 1993). But there are good reasons for caution and skepticism (Hedrick 1987, 1988; Thompson et al. 1988; Weir 1989, 1992). The results presented here, as well as those reported in other studies, offer the following guidelines: (1) Issues

such as sample size, statistical power, and the age and type of polymorphisms must be considered. (2) The disequilibrium approach tends to be more useful for locating disease genes caused primarily by one or a few mutations. The disequilibrium gradient observed between cystic fibrosis and nearby polymorphisms reflects, in part, the fact that a single lesion, ΔF508, accounts for 50%–70% of cystic fibrosis mutations in Caucasians (Cystic Fibrosis Genetic Analysis Consortium 1990). Similarly, about one-third of Huntington disease genes are found on the same chromosome haplotype (MacDonald et al. 1992). For a disease such as NF1, in which any particular mutation accounts for only a small proportion of cases, linkage disequilibrium comparisons using the disease genotype would not be useful for isolating the disease gene. (3) Even with large sample sizes and a small number of mutations, evolutionary forces such as genetic drift and admixture may upset the predicted relationship between disequilibrium and physical distance, especially in small genomic regions. A judicious choice of populations may help to overcome this problem. Isolated, well-defined populations with well-characterized histories may be most useful (Hästbacka et al. 1992). (4) As with any association study, disequilibrium results are much more convincing when replicated in multiple populations.

It is intriguing that disequilibrium is much higher within the NF1 coding sequence than it is in the region 3′ of the stop codon. A similar finding was reported in the β-globin region, where two intragenic polymorphisms defined by *Hgi*AI and *Ava*II were in strong disequilibrium with one another but revealed no disequilibrium with a *Hin*fI polymorphism located 1 kb 5′ to the gene (Chakravarti et al. 1984*a*). Although studies of other regions tend to show less intragenic disequilibrium than in NF1, few have systematically compared disequilibrium in extragenic versus intragenic polymorphisms. Such comparisons should be made in additional genomic regions to test the generality of the pattern observed in this study. If this pattern proves to be sufficiently general, then the observation of increased disequilibrium among polymorphisms in previously uncharacterized regions could be useful in helping to locate genes.

Note added in proof.—The 3′ end of the NF1 gene was recently cloned by one of us (D.V.) and lies 3.5 kb telomeric to the stop codon in exon 49. This proves that the p11-1F10 polymorphism is located approximately 70 kb from the 3′ end of the NF1 gene.

Acknowledgments

We are grateful for assistance and comments from John Bird, Melanie Culver, Russ Eldridge, Tami Elsner, Amanda Ewart, Jean-Marc Lalouel, Mark Leppert, Ken Morgan, Brith Otterud, Alan Rogers, Jeff Stevens, and Ray White. Bruce Weir kindly provided computer programs for estimating higher-order linkage disequilibrium. This research was supported by NIH grants HG-00347, HG-00470, and HG-00367, NSF grant BNS-8720330, and the Howard Hughes Medical Institute.

References

Aschbacher A, Buetow K, Chung D, Walsh S, Murray J (1985) Linkage disequilibrium of RFLP's associated with α, β, and γ fibrinogen predict gene order on chromosome 4. Am J Hum Genet Suppl 37:A186

Barker D, Holm T, White R (1984) A locus on chromosome 11p with multiple restriction site polymorphisms. Am J Hum Genet 36:1159–1171

Bech-Hansen NT, Linsley PS, Cox DW (1983) Restriction fragment length polymorphisms associated with immunoglobulin Cγ genes reveal linkage disequilibrium and genomic organization. Proc Natl Acad Sci USA 80:6952–6956

Begovich AB, McClure GR, Suraj VC, Helmuth RC, Fildes N, Bugawan TL, Erlich HA, et al (1992) Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. J Immunol 148:249–258

Bell GI, Karem JH, Rutter JR (1981) Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. Proc Natl Acad Sci USA 78:5759–5763

Benlian P, Boileau C, Loux N, Pastier D, Masliah J, Coulon M, Nigou M, et al (1991) Extended haplotypes and linkage disequilibrium between 11 markers at the APOA1-C3-A4 gene cluster on chromosome 11. Am J Hum Genet 48:903–910

Børresen AL, Moller P, Berg K (1988) Linkage disequilibrium analyses and restriction mapping of four RFLPs at the proα2(I) collagen locus: lack of correlation between disequilibrium and physical distance. Hum Genet 78:216–221

Brown AHD (1975) Sample sizes required to detect linkage disequilibrium between two or three loci. Theor Popul Biol 8:184–201

Carothers AD, Wright AF (1992) The effect of mutation on linkage disequilibrium. Ann Hum Genet 56:155–158

Cawthon RM, Andersen LB, Buchberg AM, Xu G, O'Connell P, Viskochil D, Weiss RB, et al (1991) cDNA sequence and genomic structure of EVI2B, a gene lying within an intron of the neurofibromatosis type 1 gene. Genomics 9:446–460

Cawthon R, O'Connell P, Buchberg A, Viskochil D, Weiss R, Culver M, Stevens J, et al (1990) Identification and charac-

terization of transcripts from the neurofibromatosis 1 region: the sequence and genomic structure of EVI2 and mapping of other transcripts. Genomics 7:555–565

Chakraborty R, Lidsky AS, Daiger SP, Güttler F, Sullivan S, Dilella AG, Woo SLC (1987) Polymorphic DNA haplotypes at the human phenylalanine hydroxylase locus and their relationship with phenylketonuria. Hum Genet 76: 40–46

Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH (1984*a*) Nonuniform recombination within the human β-globin gene cluster. Am J Hum Genet 36:1239–1258

Chakravarti A, Elbein SC, Permutt MA (1986) Evidence for increased recombination near the human insulin gene: implication for disease association studies. Proc Natl Acad Sci USA 83:1045–1049

Chakravarti A, Li CC, Buetow KH (1984*b*) Estimation of the marker gene frequency and linkage disequilibrium from conditional marker data. Am J Hum Genet 36:177–186

Chakravarti A, Phillips JA, Mellits KH, Buetow KH, Seeburg PH (1984*c*) Patterns of polymorphism and linkage disequilibrium suggest independent origins of the human growth hormone gene cluster. Proc Natl Acad Sci USA 81:6085–6089

Cystic Fibrosis Genetic Analysis Consortium (1990) Worldwide survey of the ΔF508 mutation—report from the Cystic Fibrosis Genetic Analysis Consortium. Am J Hum Genet 47:354–359

Daiger SP, Chakraborty R, Reed L, Fekete G, Schuler D, Berenssi G, Nasz I, et al (1989) Polymorphic DNA haplotypes at the phenylalanine hydroxylase (PAH) locus in European families with phenylketonuria (PKU). Am J Hum Genet 45:310–318

Elbein SC (1992) Linkage disequilibrium among RFLPs at the insulin-receptor locus despite intervening Alu repeat sequences. Am J Hum Genet 51:1103–1110

Estivill X, Farrall M, Scambler PJ, Bell GM, Hawley KMF, Lench NJ, Bates GP, et al (1987) A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands. Nature 326:840–845

Feinberg PF, Vogelstein B (1984) Addendum: a technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Hanauer A, Chery M, Fujita R, Driesel AJ, Gilgenkrantz S, Mandel JL (1990) The Friedreich ataxia gene is assigned to chromosome 9q13-q21 by mapping of tightly linked markers and shows linkage disequilibrium with D9S15. Am J Hum Genet 46:133–137

Harley HG, Brook JD, Floyd J, Rundle SA, Crow S, Walsh KV, Thibault M-C, et al (1991) Detection of linkage disequilibrium between the myotonic dystrophy locus and a new polymorphic DNA marker. Am J Hum Genet 49:68–75

Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nature Genet 2:204–211

Haviland MB, Kessling AM, Davignon J, Sing CF (1991) Estimation of Hardy-Weinberg and pairwise disequilibrium in the apolipoprotein AI-CIII-AIV gene cluster. Am J Hum Genet 49:350–365

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117:331–341

——— (1988) Inference of recombinational hotspots using gametic disequilibrium values. Heredity 60:435–438

Hegele RA, Plaetke R, Lalouel J-M (1990) Linkage disequilibrium between DNA markers at the low-density lipoprotein receptor gene. Genet Epidemiol 7:69–81

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78

Hudson RR (1985) The sampling distribution of linkage disequilibrium under an infinite allele model without selection. Genetics 109:611–631

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Kaplan N, Weir BS (1992) Expected behavior of conditional linkage disequilibrium. Am J Hum Genet 51:333–343

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, et al (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1080

Kupke KG, Graeber MB, Müller U (1992) Dystonia-Parkinsonism syndrome (XDP) locus: flanking markers in Xq12-q21.1. Am J Hum Genet 50:808–815

Lalouel JM, Jorde LB (1988) Idiopathic hemochromatosis: significance and implications of linkage and association to HLA. Ann N Y Acad Sci 526:34–46

Leitersdorf E, Chakravarti A, Hobbs HH (1989) Polymorphic DNA haplotypes at the LDL receptor locus. Am J Hum Genet 44:409–421

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

Lewontin RC, Kojima KI (1960) The evolutionary dynamics of complex polymorphisms. Evolution 14:458–472

Litt M, Jorde LB (1986) Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2q. Am J Hum Genet 39:166–178

MacDonald ME, Lin C, Srinidhi L, Bates G, Altherr M, Whaley WL, Lehrach H, et al (1991) Complex patterns of link-

age disequilibrium in the Huntington disease region. Am J Hum Genet 49:723–734

MacDonald ME, Novelletto A, Lin C, Tagle D, Barnes G, Bates G, Taylor S, et al (1992) The Huntington's disease candidate region exhibits many different haplotypes. Nature Genet 1:99–103

McLellan T, Jorde LB, Skolnick MH (1984) Genetic distances between the Utah Mormons and related populations. Am J Hum Genet 36:836–857

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Messiaen L, De Bie S, Moens T, Van den Enden A, Leroy J (1993) Lack of independence between five DNA polymorphisms in the NF1 gene. Hum Mol Genet 2:485

Migone N, de Lange G, Piazza A, Cavalli-Sforza LL (1985) Genetic analysis of eight linked polymorphisms within the human immunoglobulin heavy-chain region. Am J Hum Genet 37:1146–1163

Miserez AR, Schuster H, Chiodetti N, Keller U (1993) Polymorphic haplotypes and recombination rates at the LDL receptor gene locus in subjects with and without familial hypercholesterolemia who are from different populations. Am J Hum Genet 52:808–826

Murray JC, Buetow KH, Donovan M, Hornung S, Motulsky AG, Disteche C, Dyer K, et al (1987) Linkage disequilibrium of plasminogen polymorphisms and assignment of the gene to human chromosome 6q26-6q27. Am J Hum Genet 40:338–350

Murray MC, Mills KA, Demopulos CM, Hornung S, Motulsky AG (1984) Linkage disequilibrium and evolutionary relationships of DNA variants (restriction enzyme fragment length polymorphisms) at the serum albumin locus. Proc Natl Acad Sci USA 81:3486–3490

O'Connell P, Buchberg AM, Fountain J, Cawthon RM, Culver M, Stevens J, Rich DC (1990) The human homolog of murine Evi-2 lies between two von Recklinghausen neurofibromatosis translocations. Genomics 7:547–554

O'Connell P, Leach R, Cawthon R, Culver M, Stevens J, Viskochil D, Fournier REK, et al (1989*a*) Two NF1 translocations map within a 600-kilobase segment of 17q11.2. Science 244:1087–1088

O'Connell P, Leach RJ, Ledbetter DH, Cawthon RM, Culver M, Eldridge JR, Frej A-K, et al (1989*b*) Fine structure DNA mapping studies of the chromosomal region harboring the genetic defect in neurofibromatosis type 1. Am J Hum Genet 44:51–57

O'Connell P, Plaetke R, Matsunami S, Odelberg S, Jorde L, Chance P, Leppert M, et al (1993) An extended genetic linkage map and an "index" map for human chromosome 17. Genomics 15:38–47

Oliver SG, van der Aart QJM, Agostoni-Carbone ML, Aigle M, Alberghina L, Alexandraki D, Antoine G, et al (1992) The complete DNA sequence of yeast chromosome III. Nature 357:38–46

Ozelius LJ, Kramer PL, de Leon D, Risch N, Bressman SB, Schuback DE, Brin MF, et al (1992) Strong allelic association between the torsion dystonia gene (*DYT1*) and loci on chromosome 9q34 in Ashkenazi Jews. Am J Hum Genet 50:619–628

Pandolfo M, Sirugo G, Antonelli A, Weitnauer L, Ferretti L, Leone M, Dones I, et al (1990) Friedreich ataxia in Italian families: genetic homogeneity and linkage disequilibrium with the marker loci D9S5 and D9S15. Am J Hum Genet 47:228–235

Pound SE, Carothers AD, Pignatelli PM, Macnicol AM, Watson ML, Wright AF (1992) Evidence for linkage disequilibrium between D16S94 and the adult onset polycystic kidney disease (PKD1) gene. J Med Genet 29:247–248

Skraastad MI, Van de Vosse E, Belfroid R, Höld K, Vegtervan der Vlis M, Sandkuijl LA, Bakker E, et al (1992) Significant linkage disequilibrium between the Huntington disease gene and the loci D4S10 and D4S95 in the Dutch population. Am J Hum Genet 51:730–735

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool 35:627–632

Steinmetz M, Uematsu Y, Lindahl KF (1987) Hotspots of homologous recombination in mammalian genomes. Trends Genet 3:7–10

Sved JA (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor Popul Biol 2:125–141

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes. Genetics 117:343–351

Theilmann J, Kanani S, Shiang R, Robbins C, Quarell O, Huggins M, Hedrick A, et al (1989) Non-random association between alleles detected at D4S95 and D4S98 and the Huntington's disease gene. J Med Genet 26:676–681

Thompson EA, Deeb S, Walker D, Motulsky AG (1988) The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am J Hum Genet 42:113–124

Tzall S, Ellenbogen A, Eng F, Hirschhorn R (1989) Identification and characterization of nine RFLPs at the adenosine deaminase (ADA) locus. Am J Hum Genet 44:864–875

Viskochil D, Buchberg AM, Xu G, Cawthon RM, Stevens J, Wolff RK, Culver M, et al (1990) Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. Cell 62:187–192

Viskochil D, Cawthon R, O'Connell P, Xu G, Stevens J, Culver M, Carey J, et al (1991) The gene encoding the oligodendrocyte-myelin glycoprotein is embedded within the neurofibromatosis type 1 gene. Mol Cell Biol 11:906–912

Viskochil D, White R, Cawthon R (1993) The neurofibromatosis type 1 gene. Annu Rev Neurosci 16:183–205

Vnencak-Jones CL, Phillips JA (1990) Hot spots for growth

hormone gene deletions in homologous regions outside of Alu repeats. Science 250:1745–1748

Walter MA, Cox DW (1991) Nonuniform linkage disequilibrium within a 1,500-kb region of the human immunoglobulin heavy-chain complex. Am J Hum Genet 49:917–931

Ward K, O'Connell P, Carey JC, Leppert M, Jolley S, Plaetke R, Ogden B, et al (1990) Diagnosis of neurofibromatosis I by using tightly linked, flanking DNA markers. Am J Hum Genet 46:943–949

Watkins WS, O'Connell P, Stevens J, Culver M, Jorde LB (1991) A *Bagl*II RFLP near the human neurofibromatosis type 1 (NF1) gene. Nucleic Acids Res 19:6662

Weir BS (1979) Inferences about linkage disequilibrium. Biometrics 35:235–254

——— (1989) Locating the cystic fibrosis gene on the basis of linkage disequilibrium with markers? In: Elston RC, Spence MA, Hodge SE, MacCluer JW (ed) Multipoint mapping and linkage based upon affected pedigree members: genetic analysis workshop 6. Alan R Liss, New York, pp 81–86

——— (1990) Genetic data analysis. Sinauer, Sunderland, MA

——— (1992) Disequilibrium on chromosome 21 in some Utah families. Cytogenet Cell Genet 59:128–129

Weir BS, Cockerham CC (1989) Complete characterization of disequilibrium at two loci. In: Feldman MW (ed) Mathematical evolutionary theory. Princeton University Press, Princeton, pp 86–109

Weir BS, Hill WG (1986) Nonuniform recombination within the human β-globin gene cluster. Am J Hum Genet 38: 776–778

Zerba KE, Kessling AM, Davignon J, Sing CF (1991) Genetic structure and the search for genotype-phenotype relationships: an example from disequilibrium in the Apo B gene region. Genetics 129:525–533