Am. J. Hum. Genet. 62:1180–1188, 1998

An Analysis of Linkage Disequilibrium in the Interleukin-1 Gene Cluster, Using a Novel Grouping Method for Multiallelic Markers

Angela Cox,* Nicola J. Camp,* Martin J. H. Nicklin, Francesco S. di Giovine, and Gordon W. Duff

Division of Molecular and Genetic Medicine, University of Sheffield, Sheffield, United Kingdom

Summary

In population- and family-based association studies, it is useful to have some knowledge of the patterns of linkage disequilibrium that exist between markers in candidate regions. When such studies are carried out with multiallelic markers, it is often convenient to group the alleles into a biallelic system, for analysis. In this study, we specifically examined the interleukin-1 (IL-1) gene cluster on chromosome 2, a region containing candidates for many inflammatory and autoimmune disorders. Data were collected on eight markers, four of which were multiallelic. Using these data, we investigated the effect of three allele-grouping strategies, including a novel method, on the detection of linkage disequilibrium. The novel approach, termed the "δ method," measures the deviation from the expected haplotype frequencies under linkage equilibrium, for each allelic combination. This information is then used to group the alleles, in an attempt to avoid the grouping together of alleles at one locus that are in opposite disequilibrium with the same allele at the second locus. The estimate haplotype frequencies (EH) program was used to estimate haplotype frequencies and the disequilibrium measure. In our data it was found that the δ method compared well with the other two strategies. Using this method, we found that there was a reasonable correlation between disequilibrium and physical distance in the region ($r = -.540$, $P = .001$, one-tailed). We also identified a common, eight-locus haplotype of the IL-1 gene cluster.

Received July 14, 1997; accepted for publication March 11, 1998; electronically published April 17, 1998.

Address for correspondence and reprints: Dr. Angela Cox, Division of Molecular and Genetic Medicine, Royal Hallamshire Hospital, University of Sheffield, Sheffield S10 2JF, United Kingdom. E-mail: a.cox@sheffield.ac.uk

*These authors contributed equally to this work.

0002-9297/98/6205-0024$02.00

Introduction

Genetic association studies provide a powerful tool for fine-mapping disease loci (Lander and Schork 1994; Copeman et al. 1995; Weeks and Lathrop 1995). These studies test for an association, assumed to be present because of linkage disequilibrium between the marker locus and a putative disease locus. It is therefore very useful to have some prior knowledge of the degree of disequilibrium in any candidate gene regions and of any common multimarker haplotypes present in the normal population. This information is important since it can indicate the density of markers required to cover a region for analysis. Regions of relatively weak linkage disequilibrium will require marker maps that are denser than those for regions where the disequilibrium is strong. Also, knowledge of which specific alleles at two loci are in linkage disequilibrium can be used to reduce the number of tests necessary when mapping the position of maximum association across a region. Prior hypotheses can be formed regarding which allele to test at a multiallelic locus by consideration of the linkage disequilibria with alleles at a neighboring locus that has already been analyzed.

It has been shown by Jorde et al. (1994) that over distances of 50–500 kb there is generally a good correlation between linkage disequilibrium and physical distance. However, at <50 kb this relationship breaks down, presumably because of the fact that over short distances the effects of mutation, genetic drift, and population admixture outweigh those of recombination. In addition, some genomic regions depart from these general rules (e.g., the β-globin region; Chakravarti et al. 1984), and the linkage disequilibrium may be greater within genes than in intergenic regions (Jorde et al. 1994). It might be expected that microsatellite markers, because of their higher mutation rate, would contribute more to the disequilibrium in a particular genomic region than do biallelic markers, but this has been shown not to be the case in at least one instance (Watkins et al. 1994). Several other factors affect the detection of disequilibrium; for example, the power to detect disequilibrium is low when the disequilibrium is in the neg-

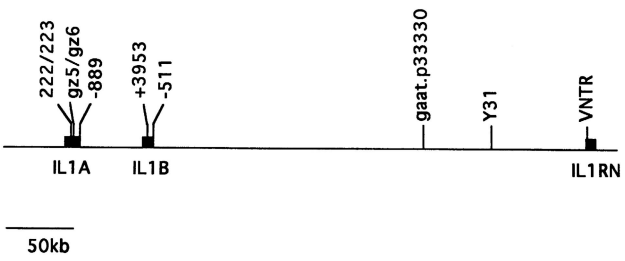

Figure 1 Polymorphic loci in the IL-1 gene cluster

ative direction (when the rare allele at one locus is associated with the common allele at the other; Thomson et al. 1988), and, when multiallelic markers are used, the method of grouping markedly affects the power to detect disequilibrium (Weir and Cockerham 1978).

In the present study, we have tried to maximize the detection of disequilibrium involving multiallelic markers by using a novel grouping strategy, the δ method, which takes into account available information about the disequilibrium between individual alleles at adjacent loci before grouping alleles together. We have compared this method with two other, simpler grouping methods, using data on eight markers in the interleukin-1 (IL-1) gene cluster on the long arm of chromosome 2 (2q13). The cluster contains the genes for IL-1α (*IL1A*), IL-1β (*IL1B*), and the IL-1 receptor antagonist (*IL1RN*), within a region of 430 kb (Nicklin et al. 1994). As part of an ongoing study of the role of the IL-1 gene family in autoimmune and inflammatory diseases, we have identified a number of biallelic and multiallelic markers in and around the IL-1 genes (di Giovine et al. 1992; Tarlow et al. 1993; McDowell et al. 1995; Clay et al. 1996; Spurr et al. 1996). We show that there is moderate linkage disequilibrium across the region and identify a common haplotype in the healthy Caucasian population studied here.

Subjects and Methods

Subjects

Subjects genotyped for disequilibrium analysis were unrelated, healthy blood donors from Sheffield (n = 112) and from Manchester (n = 100). They were all of Caucasian origin. This study was carried out with the approval of the South Sheffield Ethics Committee, and informed consent was obtained from all subjects. Family haplotype studies were carried out on samples from the first 20 families of the Arthritis and Rheumatism Council for Research National Repository (Worthington et al. 1994).

Genotyping

The positions of the marker loci are shown in figure 1. The microsatellite markers *222/223*, *gz5/gz6* (in *IL1A*), and *gaat.p33330* (*gaat*) were identified from the Genome Database (http://gdbwww.gdb.org/). Marker *Y31* is a novel microsatellite identified by screening of a YAC contig spanning the IL-1 gene cluster (Spurr et al. 1996).

Microsatellite PCRs were carried out by use of fluorescently labeled forward primers (Cruachem) in a 10-μl reaction volume containing 50 mM KCl, 10 mM Tris-HCl, pH 9.0, 1.5 mM $MgCl_2$, 200 μM dNTPs, 25 ng each primer, 50 ng DNA, 0.004% W-1 (Gibco-BRL), and 0.2 units *Taq* polymerase. The PCR conditions were 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min, for 30 cycles. One unit Perfect Match (Stratagene) was

Table 1

Estimated Frequencies of Marker Alleles

	Marker Allele Frequency (Allele Size, in Mobility Units)[a]							
Allele No.	*222/223* [*N* = 384]	*gz5/gz6* [*N* = 392]	*−889* [*N* = 398]	*+3953* [*N* = 398]	*−511* [*N* = 398]	*gaat* [*N* = 404]	*Y31* [*N* = 370]	VNTR [*N* = 398]
1	.005 (126)	.003 (79)	.714	.812	.618	.659 (189)	.091 (148)	.744
2	.018 (128)	.005 (82)	.286	.188	.382	.002 (193)	.008 (158)	.256
3	.378 (130)	.676 (88)	...	...	...	.255 (197)	.454 (160)	...
4	.299 (132)	.316 (91)	...	...	...	.084 (201)	.062 (162)	...
5	.016 (134)	...	...	...	...	...	.003 (164)	...
6	.208 (136)	...	...	...	...	...	.122 (166)	...
7	.055 (138)	...	...	...	...	...	.035 (168)	...
8	.003 (140)	...	...	...	...	...	.030 (170)	...
9	.010 (142)	...	...	...	...	...	.095 (172)	...
10	.008 (144)	...	...	...	...	...	.086 (174)	...
11	...	...	...	...	...	...	.003 (176)	...
12	...	...	...	...	...	...	.011 (178)	...

Note.—Frequency estimates are determined directly from the sample, not from haplotype frequencies.

[a] N = no. of chromosomes.

added to *gz5/gz6* PCRs. The primer sequences were as follows: for *222/223*, 5′-ATG TAT AGA ATT CCA TTC CTG and 5′-TAA AAT CAA GTG TTG ATG TAG; for *gz5/gz6*, 5′-GGG ATT ACA GGC GTG AGC CAC CGCG and 5′-TTA GTA TTG CTG GTA GTA TTC ATAT; for *gaat*, 5′-GAG GCG TGA GAA TCT CAA GA and 5′-GTG TCC TCA AGT GGA TCT GG; and, for *Y31*, 5′-GGG CAA CAG AGC AAT GTT TCT and 5′-CAG TGT GTC AGT GTA CTG TT. A sample of PCR product was examined by agarose-gel electrophoresis, and the remainder of the PCR products were pooled according to the intensity of ethidium-bromide staining. Two microliters of the pool was analyzed on an ABI 373A automated sequencer, and allele sizes were determined against the ABI Genescan 500-rox (6-carboxyrhodamine) size standard, by use of the Genescan and Genotyper software. Alleles were globally binned by use of a simple computer program and were numbered in order of size (table 1).

IL1A −889, *IL1B −511*, and *IL1RN* VNTR markers were genotyped by PCR RFLP as described elsewhere (di Giovine et al. 1992; Tarlow et al. 1993; McDowell et al. 1995). The *IL1* VNTR polymorphism is essentially biallelic in this data set, since only four-repeat and two-repeat alleles were detected. *IL1B +3953* PCRs were carried out with primers 5′-CTC AGG TGT CCT CGA AGA AAT CAAA and 5′-GCT TTT TTG CTG TGA GTC CCG, for 35 cycles, with annealing at 67.5°C. The resulting PCR products were digested with restriction enzyme *Taq*I. Allele 1 yielded fragments of 97, 85, and 12 bp, and allele 2 yielded fragments of 182 and 12 bp.

Physical Distances

Intergenic distances were determined by estimation based on restriction-fragment sizes from genomic DNA (Nicklin et al. 1994) and the insert sizes of relevant P1 artificial chromosome clones from a contig spanning the IL-1 gene cluster (Nothwang et al. 1997). Intragenic distances were determined from the relevant nucleotide sequence obtained from the GenBank database (http://www.ncbi.nlm.nih.gov/Web/Genbank/).

Estimation of Linkage Disequilibrium

The estimate haplotype frequencies (EH) program (Terwilliger and Ott 1994) was used to determine maximum-likelihood estimates of disequilibrium ($\hat{D}_{ij}$) between each pairwise combination of alleles, where $D_{ij} = h_{ij} - p_i q_j$ and p_i and q_j are the frequencies for allele i at locus 1 and for allele j at locus 2, respectively, and h_{ij} is the frequency of haplotype ij. The program calculates maximum-likelihood values for the haplotype frequencies (and, hence, the allele frequencies) under H_0 (no association) and for the haplotype frequencies under H_1 (allelic association allowed). However, for markers with >2 alleles, the EH estimates for allele frequencies correlated poorly with the allele frequencies as estimated directly from the sample population and, therefore, gave no confidence to the $\hat{D}_{ij}$ estimates given. It was therefore necessary to group alleles of the multiallelic markers into a biallelic system. Analysis of the markers in a biallelic format has the added advantage that the notations $\hat{D}_{ij}$, p_i, and q_j can be simplified to $\hat{D}$, p, and q, respectively, where we define p and q to be the frequencies of the rarer alleles at loci 1 and 2 (such that, without loss of generality, $p \leqslant q \leqslant .5$) and $\hat{D}$ to be the estimated disequilibrium between those alleles. Under a biallelic system, power is also much simpler to determine by use of the equations detailed by Hill (1974) (see Appendix). In addition, the sign of $\hat{D}$ becomes informative, such that $\hat{D}$ is positive when the rarer alleles at each of the two loci are associated and negative when the rare allele at one locus is associated with the common allele at the other locus.

Table 2

Calculation of δ_{ij} for the Loci Pair *222/223* and *gz5/gz6*

Haplotype				
222/223	*gz5/gz6*	o_{ij}	e_{ij}	δ_{ij}
3	3	132.4898	97.1348	3.587
3	4	7.0102	41.6854	−5.371
4	3	17.6340	72.1573	−6.419
4	4	86.3660	30.9663	9.956
5	3	.0000	(4.1629)	(−2.040)
5	4	6.0000	(1.7865)	(3.152)
6	3	69.8762	51.3427	2.587
6	4	3.6238	22.0337	−3.922
7	3	18.0000	12.4888	1.560
7	4	.0000	5.3596	−2.315

Note.—Parentheses indicate $1.0 \leqslant e_{ij} < 5.0$ and $\delta_{ij} > 1.0$ (see text).

Since the method of allele grouping clearly affects the power to detect disequilibrium (Weir and Cockerham 1978), a preliminary analysis was carried out to investigate three different grouping methods. The three methods used were the following: (1) The "common allele versus the rest" method compared the most common allele (see table 1) against the remaining alleles grouped together. (2) The "allele size" method grouped the alleles according to the bimodal distribution of allele frequency versus size (see table 1). The groupings were as follows: for *222/223*, alleles 1–4 and alleles 5–10; for *gz5/gz6*, alleles 1–3 and allele 4; for *gaat*, alleles 1 and 2 and alleles 3 and 4; and, for *Y31*, alleles 1–4 and alleles 5–12. (3) The δ method used the observed allelic associations between pairs of markers to determine the grouping, as follows:

i. The value of $\delta_{ij} = (o_{ij} - e_{ij})/\sqrt{e_{ij}}$ was calculated for each haplotype, where e_{ij} is the expected number of haplotypes ij, assuming equilibrium ($e_{ij} = 2np_iq_j$, where n is

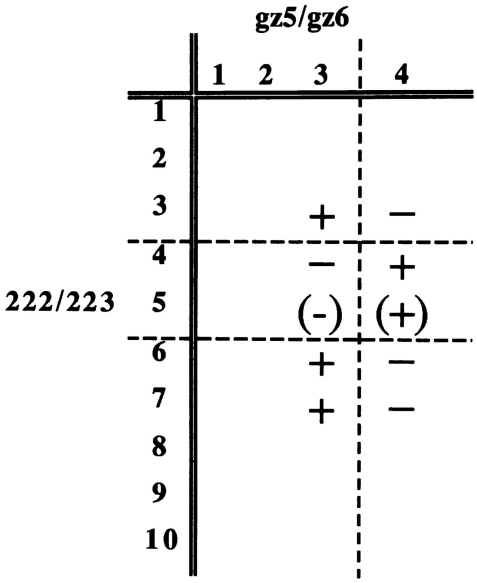

Figure 2 Grouping diagram for the marker pair *222/223* and *gz5/gz6*, by use of the δ method. Parentheses indicate $1.0 \leqslant e_{ij} < 5.0$ and $\delta_{ij} > 1$ (for details, see text).

the number of individuals in the study), and where o_{ij} is a basic estimate for the observed haplotype count, determined as follows: Genotypes that could be unambiguously resolved were haplotype counted. Those that could not be unambiguously resolved were double heterozygotes (*i*1, *i*2/*j*1, *j*2), which could be assigned to two possible haplotype sets, either (*i*1*j*1, *i*2*j*2) or (*i*1*j*2, *i*2*j*1). By use of the haplotype frequencies estimated from the unambiguous haplotype count, the probability of each set was calculated and used as a "partial" count. In this way, haplotype scores for the ambiguous genotypes were obtained, and the total scores (ambiguous plus unambiguous) constituted the o_{ij}'s used for δ_{ij}.

ii. For each two-locus allele combination for which $e_{ij} \geqslant 5.0$ and $\delta_{ij} > 1.0$, a plus sign (+) or a minus sign (−) (indicating the sign of the deviation) was entered in the relevant cell of the grouping table. Table 2 shows the o_{ij}, e_{ij}, and δ_{ij} values for the relevant allelic combinations for the *222/223* and *gz5/gz6* pairing, and figure 2 shows the resulting grouping table for this pair.

iii. Dividing lines were drawn on to the grouping table, between any opposite signs that were adjacent and in the same row or column (see fig. 2). (*a*) If the opposing signs directly flanked each other, the line position was unambiguous, and these lines were drawn. (*b*) For ambiguous lines, additional plus (+) and minus (−) signs were entered for the interim cells if $1.0 \leqslant e_{ij} < 5.0$ and $\delta_{ij} > 1.0$, and the lines were drawn accordingly. These additional plus (+) and minus (−) signs are shown in parentheses in figure 2, and e_{ij} and δ_{ij} values for the corresponding allele combinations are in parentheses in table 2. (*c*) If line assignments were still ambiguous, a line was drawn at any position between the closest opposing signs.

iv. The grouping table then was used to form the allele groupings. To group the first locus, the table was read top to bottom and then, for the second locus, left to right. Each time a dividing line was crossed, the subsequent alleles were assigned to the opposite group.

Thus, for the *222/223* and *gz5/gz6* pairing, the alleles of *222/223* were divided into two groups: one group containing alleles 1–3 and 6–10 and the other containing alleles 4 and 5. Likewise, *gz5/gz6* alleles were divided into one group containing alleles 1–3 and another containing allele 4 only (fig. 2 and table 3).

In order to compare the degree of disequilibrium between different pairwise combinations of loci, a frequency-independent measure of disequilibrium ($\tilde{D}$, the proportion of maximum possible disequilibrium in the given direction) was calculated, where $\tilde{D} = \hat{D}/|D_{\max}|$. Without loss of generality, we have assumed that $p \leqslant q \leqslant 0.5$, and therefore $-pq \leqslant D \leqslant p(1-q)$, such that if $\hat{D} < 0$ then $D_{\max} = -pq$, or if $\hat{D} > 0$ then $D_{\max} = p(1-q)$.

Table 3

Multiallelic Marker Allele Grouping, by Use of the δ Method

	Allele Grouping						
Marker	*gz5/gz6*	*−889*	*+3953*	*−511*	*gaat*	*Y31*	VNTR
222/223	(1–3)(4) (4,5)(1–3,6–10)	(1)(2) (1–3)(4–10)	(1)(2) (1–3)(4–10)	(1)(2) (4,5)(1–3,6–10)	(1,2,4)(3) (4,5)(1–3,6–10)	(1,5,6)(2–4, 7–11) (1–3,6)(4,5,7–10)	(1)(2) (1–3)(4–10)
gz5/gz6	…	(1)(2) (1–3)(4)	(1)(2) (1–3)(4)	(1)(2) (1–3)(4)	(1,2,4)(3) (1–3)(4)	(1,9–11)(2–8) (1–3)(4)	(1)(2) (1–3)(4)
gaat	…	(1)(2) (1–3)(4)	(1)(2) (1–3)(4)	(1)(2) (1,2,4)(3)	…	(1,5–11)(2–4) (1,2,4)(3)	(1)(2) (1,2,4)(3)
Y31	…	(1)(2) (1–6)(7–11)	(1)(2) (1–6)(7–11)	(1)(2) (1,5–11)(2–4)	…	…	(1)(2) (1,5–11)(2–4)

Note.—In each case, the groupings for the column-heading markers are given in the top row, and the groupings for the markers in column 1 are given in the bottom row.

Table 4

$\tilde{D}$ Values for Three Methods of Grouping Alleles at the Multiallelic Marker Loci

	$\tilde{D}$ VALUE		
MARKER	δ Method	"Common Allele vs. Rest" Method	"Allele Size" Method
gz5/gz6	.87	.79	.77
−889	.83	.81	.98
+3953	.71	.74	.77
−511	.54	.15[a]	.61
gaat	.43	.03[a]	.53
Y31	.38	.12[a]	.16[a]
VNTR	.5	.48	.04[a]

NOTE.—The values given are for the disequilibrium between *222/223* and the other markers listed.

[a] Indicates no significance at the $P = .05$ level, even before correction for multiple comparisons.

Output from the EH program includes the log likelihoods for the maximum-likelihood–parameter values under H_0 and H_1, and, since $-2\ln(L_0/L_1) \sim \chi^2_1$, where L_0 and L_1 are the likelihoods under H_0 and H_1, the P values then could be determined for each test. The asymptotic variance for $\hat{D}$, under H_0 ($D = 0$) and H_1 were computed by use of the formulas defined by Hill (1974) for genotypic data (see Appendix). By use of these formulas, the power to detect 50% D_{max} for each pairwise comparison was calculated.

Possible haplotypes were identified by inspection of the δ_{ij}'s described above, and these haplotypes were backed up by the magnitude and direction of the disequilibria observed. For multiallelic loci, the allele in the group that contributed most to the disequilibrium has been identified on the haplotype. To estimate the population haplotype frequencies, rates of carriage of at least one copy of the relevant alleles in the population were determined. These do not represent true haplotypes, since the phase is unknown. Monte Carlo–simulation techniques were used to test for significant deviation from a simulated null distribution for these combined carriages, under the assumption of no association. Haplotype reconstruction was performed, by use of Genehunter (Kruglyak et al. 1996), to determine maximum-likelihood haplotypes from family data, in order to confirm the proposed haplotypes identified in the population sample.

Results

Figure 1 shows the relative positions of the eight marker loci used in this study. DNA samples from 212 unrelated healthy volunteers were genotyped for each of these markers, and the resulting estimates of allele frequencies are shown in table 1. In order to determine the linkage disequilibria between pairwise combinations of loci, the EH program was used (Terwilliger and Ott 1994). This program was found to be most efficient when used with biallelic systems, and, since there are other advantages of using a biallelic format (see Subjects and Methods), we compared three grouping strategies for the multiallelic marker alleles. The first two methods were a "common allele versus the rest" approach and an "allele size" approach. The latter was based on the bimodal distribution of allele sizes, which was observed for all the multiallelic markers examined. The third method, a novel strategy termed the "δ method," incorporated, into the grouping strategy, information about the observed allelic associations between pairs of markers, such that disequilibrium between subsets of alleles was not masked. The grouping strategies are described fully in Subjects and Methods, and the δ method is illustrated for *222/223* and *gz5/gz6* in table 2 and figure 2. Table 3 shows, for all pairwise marker combinations, the actual grouping that resulted from the δ method. Other than those for *gz5/gz6*, the groupings differed for different pairwise comparisons, but there were consistencies: for *gaat,* alleles 3 and 4 were always in opposite groups, and, for *222/223,* alleles 4 and 5 and 1–3 were always in opposition. The grouping for *Y31* showed the most variation, because *Y31* is the most polymorphic marker (see Discussion).

The disequilibria between pairs of loci were expressed

Table 5

Disequilibrium ($\tilde{D} = \hat{D}/|D_{max}|$) and Physical Distances between Marker Loci

	222/223	*gz5/gz6*	*−889*	*+3953*	*−511*	*gaat*	*Y31*	VNTR
222/223	...	.872	.829	.710	.535	.433	−.379	−.499
gz5/gz6	2.5	...	−.889	−.695	.540	.517	−.681	.286
−889	10	4.5	...	.804	−.264	.337	−.462	−.207
+3953	55	55	50	...	−.617	.409	−.894	−.439
−511	60	60	55	4.5	...	.691	−.339	.448
gaat	260	260	255	205	200	...	−.470	.442
Y31	310	310	305	255	250	50	...	−.533
VNTR	380	380	375	325	320	120	70	...

NOTE.—Disequilibrium values are shown at the top right, and approximate physical distances (in kb) are shown at the bottom left. Intergenic distances are given to the nearest 5 kb and intragenic distances to the nearest 0.5 kb.

Table 6

Power to Detect 50% D_{max} and *P* Values of $-2\ln(L_0/L_1)$

	222/223	*gz5/gz6*	*−889*	*+3953*	*−511*	*gaat*	*Y31*	VNTR
222/223	...	~100 (+)	~100 (+)	98 (+)	~100 (+)	~100 (+)	84 (−)	93 (−)
gz5/gz6	$< 1 \times 10^{-10}$	...	87 (−)	60 (−)	~100 (+)	~100 (+)	81 (−)	~100 (+)
−889	$< 1 \times 10^{-10}$	3×10^{-8}	...	~100 (+)	96 (−)	89 (+)	70 (−)	78 (−)
+3953	1×10^{-7}	.009[a]	$< 1 \times 10^{-10}$	...	79 (−)	97 (+)	42 (−)	52 (−)
−511	1×10^{-9}	4×10^{-10}	.095[a]	.026[a]	...	~100 (+)	~100 (−)	~100 (+)
gaat	9×10^{-8}	2×10^{-9}	.017[a]	5×10^{-4}	$< 1 \times 10^{-10}$	...	98 (−)	~100 (+)
Y31	.034[a]	1×10^{-4}	.011[a]	1×10^{-5}	.002[a]	2×10^{-4}	...	97 (−)
VNTR	.001	.001	.298[a]	.118[a]	8×10^{-6}	1×10^{-9}	4×10^{-5}	...

NOTE.—Power is shown at the top right, with the sign of disequilibrium in parentheses; *P* values are shown (uncorrected) at the bottom left. For an overall significance level of $P = .05$, a pointwise significance level of .0018 should be used for 28 comparisons.

[a] Not significant at $P = .0018$ threshold.

as $\tilde{D}$; the ratio of $\hat{D}$ to its maximum value, D_{max}, and the $\tilde{D}$ values for all the comparisons involving *222/223,* under the three grouping methods, are shown in table 4. It can be seen that significant disequilibrium is not detected in several instances using the "common allele versus the rest" and the "allele size" grouping strategies. Note that disequilibrium is not detected between *222/223* and *IL1B −511* and *gaat,* in the "common allele versus the rest" approach; between *222/223* and *Y31,* in both the "common allele versus the rest" and the "allele size" approaches; or between *222/223* and the *IL1RN* VNTR, in the "allele size" approach. Therefore, the remainder of the analyses were carried out by grouping the alleles by use of the δ method.

Table 5 shows the disequilibrium ($\tilde{D}$) values for all 28 possible pairwise comparisons, together with the approximate physical distances (in kilobases) between the loci. Table 6 shows the power to detect 50% D_{max} for each locus combination and the *P* values for each corresponding $\hat{D}$. Significant linkage disequilibrium ($P <$ .0018 for 28 tests) was detected between most combinations of loci, and only three marker combinations were nonsignificant before correction for multiple testing ($P > .05$). The correlation between disequilibrium $\tilde{D}$ and physical distance was $r = -.540$ ($P = .001$, one-tailed) (fig. 3).

Examination of which alleles of the multiallelic loci were contributing the most to the disequilibrium (from the determination of δ_{ij}; see Subjects and Methods) revealed the existence of two haplotypes containing alleles of all eight loci. This observation was backed up by examination of the haplotype frequencies and disequilibrium values obtained after the grouping. The first haplotype (alleles 4 4 1 1 2 3 3 2, in chromosome order from *222/223* to the *IL1RN* VNTR) is the most common (carriage of 34/198) and is present seven times more frequently than expected (expected carriage of 4.5/198; $P < 10^{-6}$). The second haplotype (alleles 3 3 2 2 1 4 6 1; carriage of 2/206) was present four times more frequently than expected (expected carriage of 0.5/206), but this was not statistically significant ($P = .11$). Since phase cannot be determined by use of data from unrelated individuals, haplotype reconstructions using data from 14 Caucasian families were determined by use of the Genehunter program (Lander and Kruglyak 1996). This confirmed the existence of the more-common haplotype, for 20 individuals, including eight unrelated founders from six families, possessing 4 4 1 1 2 3 3 2 as one of their maximum-likelihood haplotypes.

Discussion

In this study, we determined the degree of linkage disequilibrium in the IL-l gene cluster, using a new grouping method (the δ method) for multiallelic markers, in an attempt not to mask disequilibrium by grouping together two alleles that are in disequilibrium with different alleles at an adjacent marker locus. The sign and magnitude of the disequilibrium ($\tilde{D}$) clearly has little meaning when the overall $\tilde{D}$ value for the collapsed "allele" results from a combination of positive and negative disequilibria for the individual alleles. In our data, the δ method compares favorably with the two other methods, in terms of sensitivity. The disadvantages of the δ method include that it is less straightforward than the other strategies. Also, the information used for the grouping was based on an estimate for the "observed" haplotype frequencies (see Subjects and Methods). For highly polymorphic markers, the increased heterozygosity means that the estimates for δ_{ij} are less accurate, because of the lower proportion of unambiguous haplotypes. This is especially problematic if both marker loci are highly polymorphic. It therefore is progressively more difficult to group the alleles with confidence as the markers become more polymorphic. In some cases, grouping alleles in opposite disequilibrium is unavoidable, and therefore it is inevitable that some disequilibrium will be masked when a highly polymorphic marker such as *Y31* is forced into a biallelic system. For situations in which a collapsed biallelic system is required,

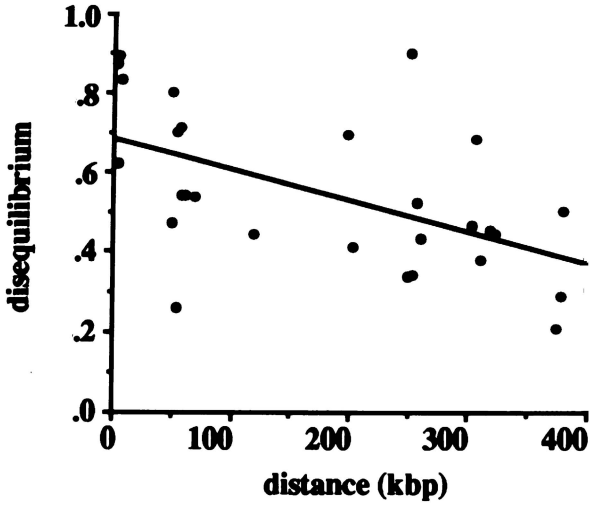

Figure 3 Correlation between disequilibrium and physical distance. Disequilibrium values ($\tilde{D}$) and physical distances were determined as described in the text. Values are for $r = -.540$ ($P = .001$, one-tailed).

however, the δ method uses the maximum amount of prior knowledge, when grouping the alleles of a multiallelic system, and is worthy of further investigation in other systems.

The data presented here indicate a moderate degree of linkage disequilibrium across an ~400-kb stretch of chromosome region 2q13. Overall, there was a reasonable correlation between physical distance and linkage disequilibrium ($r = -.540, P = .001$, one-tailed) (fig. 3), in accordance with other studies of a similar distance (Jorde et al. 1994). The disequilibrium was strong for the three markers within the *IL1A* gene (table 5), but within the *IL1B* gene only a moderate value of $\tilde{D}$ ($-.617$; *+3953/−511*) was obtained; although nominally significant, this was not significant when corrected for multiple comparisons. Other, smaller studies have shown much weaker disequilibrium within the *IL1B* gene; for example, Guasch et al. (1996) reported no significant disequilibrium, in a smaller sample size, between the *+3953* polymorphism and an *Alu*I polymorphism in the promoter, at position −31. In contrast, the degree of disequilibrium between several biallelic markers in positive disequilibrium within the *IL1RN* gene is very strong (Clay et al. 1996; Guasch et al. 1996). For this study, we genotyped only one marker in *IL1RN*, since five other biallelic markers previously had been found to be in complete association with *IL1RN* (Clay et al. 1996). Jorde et al. (1994) found that intragenic disequilibrium tended to be higher than expected; for *IL1B*, the departure from this rule probably is due to the reduction in power when disequilibrium is in the negative direction (Thompson et al. 1988) but also may reflect the age of the polymorphisms.

Table 7

Two-Locus Genotype Frequencies

Locus 2 Genotype	Frequency, at Locus 1 Genotype of		
	AA	Aa	aa
BB	$y_1 = h_1^2$	$y_2 = 2h_1h_3$	$y_3 = h_3^2$
Bb	$y_4 = 2h_1h_2$	$y_5 = 2(h_1h_4 + h_2h_3)$	$y_6 = 2h_3h_4$
bb	$y_7 = h_2^2$	$y_8 = 2h_2h_4$	$y_9 = h_4^2$

In general, an understanding of which markers are in strong linkage disequilibrium allows for the more rational design of genetic studies. In the IL-1 system in particular, where alleles of different IL-1 genes may act in concert to determine an overall inflammatory phenotype, a knowledge of the existing disequilibria is vital to our understanding of which allele combinations are important in disease.

Acknowledgments

We thank Mark Dale, Nicola Hare, and Hazel Holden, for expert technical assistance, and the Arthritis and Rheumatism Council (ARC) Epidemiological Research Unit, Manchester, for providing family material. This work was supported by ARC program grant D0528 (to M.J.H.N., G.W.D., and F.S.d.G.).

Appendix

Variance and Power Estimation

Variance

Without loss of generality, we labeled $p \leqslant q \leqslant .5$, such that p is the allele frequency for the rare allele A at locus 1 and q is the frequency for the rare allele B at locus 2. By use of this notation, the four possible haplotype frequencies can be written as follows:

$$h_1 = pq + D ,$$

$$h_2 = p(1 - q) - D ,$$

$$h_3 = (1 - p)q - D , \text{ and}$$

$$h_4 = (1 - p)(1 - q) + D ,$$

for haplotypes AB, Ab, aB, and ab, respectively. From these equations, the two-locus genotype frequencies shown in table 7 can be derived.

Let $n_i (\sum_{i=1}^{9} n_i = n)$ represent the number of two-locus genotypes y_i in a sample of n individuals, and let $\boldsymbol{\theta}$ be the vector of parameters, such that $\theta_1 = p$, $\theta_2 = q$, and $\theta_3 = D$. Then, the likelihood equation is given by

$$L(\boldsymbol{\theta}) = \frac{n!}{\prod_{i=1}^{9} n_i!} \prod_{i=1}^{9} y_i^{n_i} ,$$

such that the log likelihood is

$$l(\boldsymbol{\theta}) = \ln[L(\boldsymbol{\theta})] = \sum_{i=1}^{9} n_i \ln(y_i) + K ,$$

where K is a constant.

Maximum-likelihood–estimation values for p, q, and D are found by solving the differential equation $\partial l(\boldsymbol{\theta})/\partial\theta_i = 0$, for $i = 1, 2, 3$, and the asymptotic variance-covariance matrix, $V(\boldsymbol{\theta})$, can be calculated by use of the equation

$$V(\boldsymbol{\theta}) = I^{-1}(\boldsymbol{\theta}) , \tag{A1}$$

where $I(\boldsymbol{\theta})$ is the Fisher information matrix, defined as

$$I(\boldsymbol{\theta}) = \left[E\left(-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\theta_j \partial\theta_k} \right)_{jk\text{th element}} \right] . \tag{A2}$$

Now,

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\theta_j \partial\theta_k} = \sum_{i=1}^{9} \left[\left(\frac{n_i}{y_i}\right)\left(\frac{\partial^2 y_i}{\partial\theta_j \partial\theta_k}\right) - \left(\frac{n_i}{y_i^2}\right)\left(\frac{\partial y_i}{\partial\theta_j}\right)\left(\frac{\partial y_i}{\partial\theta k}\right) \right] ,$$

for $j, k = 1, 2, 3$, and, since $E(n_i) = ny_i$ and $\sum y_i = 1$, so that $\sum_{i=1}^{9} (\partial^2 y_i/\partial\theta_j\partial\theta_k) = 0$, equation (A2) can be written as

$$I(\boldsymbol{\theta}) = n\left(\sum_{i=1}^{9} \frac{\partial y_i}{\partial\theta_j} \frac{\partial y_i}{\partial\theta_k} y_i^{-1} \right)_{jk\text{th element}} . \tag{A3}$$

The asymptotic variance-covariance matrix is then the inverse of equation (A3), for n genotypic observations, with the 3-3 element the asymptotic variance for the disequilibrium, D.

Power

Let $V(D)$ be the asymptotic variance for the disequilibrium measure, D, calculated from equation (A1). Let $V_1(D)$ be this variance for a single observation, such that $V_n(D) = V_1(D)/n$ is the equivalent for n observations. Then, under the null hypothesis, $H_0{:}D = 0$, $V_n(D)$ simplifies to

$$V_n(D = 0) = \frac{p(1-p)q(1-q)}{n} .$$

Unfortunately, for $D \neq 0$, there is no explicit formula, and the variance must be calculated numerically, as indicated in equation (A1).

The critical value for rejection of the null hypothesis, at the 5% significance level, is 1.96, such that H_0 is rejected if

$$\left| \frac{\hat{D} - 0}{\sqrt{V_n(D = 0)}} \right| > 1.96 \text{ or}$$

$$|\hat{D}| > 1.96 \sqrt{\frac{p(1-p)q(1-q)}{n}} .$$

In the main text, we quoted the power to detect 50% of the maximum possible disequilibrium attainable, $D_{\max}$. The power to detect a true disequilibrium of $D \neq 0$—specifically, in this case, $D = .5D_{\max}$—is found as follows:

$$\begin{aligned}
\text{Power} &= P(\text{reject } H_0{:}D = 0/D \neq 0) \\
&= P\left(|\hat{D}| > 1.96 \sqrt{\frac{p(1-p)q(1-q)}{n}} / D \neq 0 \right) \\
&= P\left(\hat{D} < -1.96 \sqrt{\frac{p(1-p)q(1-q)}{n}} / D \neq 0 \right) \\
&\quad + P\left(\hat{D} > 1.96 \sqrt{\frac{p(1-p)q(1-q)}{n}} / D \neq 0 \right) \\
&= \Phi\left(\frac{-[1.96\sqrt{p(1-p)q(1-q)} + \sqrt{n}D]}{\sqrt{V_1(D)}} \right) \\
&\quad + \Phi\left(\frac{-1.96\sqrt{p(1-p)q(1-q)}}{\sqrt{V_1(D)}} \right) ,
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative probability for the standard normal distribution and where $V_1(D)$ is calculated by use of equation (A1), for a single observation.

References

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

Clay FE, Tarlow JK, Cork MJ, Cox A, Nicklin MJH, Duff GW (1996) Novel interleukin-1 receptor antagonist exon polymorphisms and their use in allele-specific mRNA assessment. Hum Genet 97:723–726

Copeman JB, Cucca F, Hearne CM, Cornall RJ, Reed PW, Ronningen KS, Undlien DE, et al (1995) Linkage disequilibrium mapping of a type 1 diabetes susceptibility gene (*IDDM7*) to chromosome 2q31-q33. Nat Genet 9:80–85

di Giovine FS, Takhsh E, Blakemore AIF, Duff GW (1992) Single base polymorphism at −511 in the human interleukin-1 beta gene (*IL1B*). Hum Mol Genet 1:450

Guasch JF, Bertina RM, Reitsma PH (1996) Five novel intragenic dimorphisms in the human interleukin-1 genes combine to high informativity. Cytokine 8:598–602

Hill WG (1974) Estimation of linkage disequilibrium in randomly mating populations. Heredity 33:229–239

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

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

McDowell TL, Symons JA, Ploski R, Forre O, Duff GW (1995) A genetic association between juvenile rheumatoid arthritis and a novel interleukin-1a polymorphism. Arthritis Rheum 38:221–228

Nicklin MJH, Weith A, Duff GW (1994) A physical map of the region encompassing the human interleukin-1α, interleukin-1β, and interleukin-1 receptor antagonist genes. Genomics 19:382–384

Nothwang HG, Strahm B, Denich D, Kubler M, Schwabe J, Gingrich JC, Jauch A, et al (1997) Molecular cloning of the interleukin-1 gene cluster: construction of an integrated YAC/PAC contig and a partial transcriptional map in the region of chromosome 2q13. Genomics 41:370–378

Spurr NK, Bashir R, Bushby K, Cox A, Cox S, Hildebrandt F, Hill N, et al (1996) Report of the Fourth International Workshop on Chromosome 2 Mapping 1996. Cytogenet Cell Genet 73:255–273

Tarlow JK, Blakemore AIF, Lennard A, Solari R, Hughes HN, Steinkasserer A, Duff GW (1993) Polymorphism in the human II-1 receptor antagonist gene intron 2 is caused by variable numbers of an 86-bp tandem repeat. Hum Genet 91:403–404

Terwilliger JD, Ott J (1994) Linkage disequilibrium between alleles at marker loci. In: Handbook of human genetic linkage, 1st ed. Johns Hopkins University Press, Baltimore, pp 188–198

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

Watkins WS, Zenger R, O'Brien E, Nyman D, Eriksson AW, Renlund M, Jorde LB (1994) Linkage disequilibrium patterns vary with chromosomal location: a case study from the von Willebrand factor region. Am J Hum Genet 55: 348–355

Weeks DE, Lathrop GM (1995) Polygenic disease—methods for mapping complex disease traits. Trends Genet 11: 513–519

Weir BS, Cockerham CC (1978) Testing hypotheses about linkage disequilibrium with multiple alleles. Genetics 88: 633–642

Worthington J, Ollier WER, Leach MK, Smith I, Hay EM, Thomson W, Pepper L, et al (1994) The Arthritis and Rheumatism Council's National Repository of Family Material: pedigrees from the first 100 rheumatoid arthritis families containing affected sibling pairs. Br J Rheumatol 33: 970–976