Am J Hum Genet 35:201-213, 1983

Heterozygous Expression of Insulin-Dependent Diabetes Mellitus (IDDM) Determinants in the HLA System

N. E. MORTON,[1] A. GREEN,[2] T. DUNSWORTH,[3] A. SVEJGAARD,[4] J. BARBOSA,[3] S. S. RICH,[3] L. ISELIUS,[1] P. PLATZ,[4] AND L. P. RYDER[4]

SUMMARY

HLA phenotypes of cases with insulin-dependent diabetes mellitus (IDDM) and identity by descent of *HLA* haplotypes in affected sib-pairs support an intermediate model in which morbid risk is increased by one *HLA*-linked IDDM determinant, and greatly increased by two determinants, which may be qualitatively different in *DR3* and *DR4* haplotypes. Linkage analysis allowing for gametic disequilibrium reveals no recombination in pedigrees with a *DR3/DR4* propositus, but spurious recombination in the remaining pedigrees. This evidence favors interaction of unlinked IDDM determinants to produce affection in a small proportion of heterozygotes for an *HLA*-linked determinant. Partition of data by *HLA* type of the propositus (ideally by *DR* and the complement types jointly) is a powerful method to resolve etiological heterogeneity for *HLA*-associated diseases.

INTRODUCTION

Green et al. [1] and Dunsworth et al. [2] recently reported segregation and linkage analyses of multiplex pedigrees with insulin-dependent diabetes mellitus (IDDM). They examined recessive, dominant, and intermediate models, with and without a multifactorial background and sporadic cases. Both studies found significant heterogeneity in the estimate of recombination between IDDM and the *HLA* locus: pedigrees with a *DR3/DR4* propositus revealed little if any recombination, whereas linkage appeared loose in the remaining pedigrees. They concluded that loose linkage was simulated by residual variation segregating independently of the *HLA* system in at least a proportion of families. The question was left open whether the

Received November 13, 1981; revised June 2, 1982.
This work was supported by grant GM-17173 from the National Institutes of Health.
[1] Population Genetics Laboratory, University of Hawaii, Honolulu, Hawaii.
[2] University Institute of Clinical Genetics, Odense, Denmark.
[3] Department of Medicine, University of Minnesota, Minneapolis, Minn.
[4] Tissue Typing Laboratory, Rigshospitalet, Copenhagen, Denmark.

unlinked determinants act independently or interactively with the *HLA* system. Here we report evidence that favors interaction, predominantly with heterozygous IDDM determinants tightly linked to the *DR* locus.

MATERIALS AND METHODS

Diagnostic criteria for IDDM were: onset before age 40, idiopathic, ketosis-prone, nonobese, and insulin-dependent. Every pedigree contajned at least one proband meeting all the diagnostic criteria. The linkage material of Green et al. [1] comprised 92 Caucasoid pedigrees, predominantly from Europe, with at least two affected sibs typed for *DR*. The linkage material of Dunsworth et al. [2] comprised 118 pedigrees living in the upper Midwest area of the United States and typed for *DR*. No other selection was deliberately applied, but pedigree size depends on accessibility of relatives for *HLA* typing. The two samples were pooled for analyses not reported separately.

For clarity, we shall refer to two classes of IDDM determinants: *HLA*-linked and unlinked. *Linked determinants* are part of the *HLA* system. They may include alleles at a hypothetical locus for IDDM susceptibility, *HLA* haplotypes with pleiotropic effects on IDDM susceptibility, or complementation of *HLA* haplotypes and immune response factors. In any case, recombination is presumably no greater than between the *A* and *D* loci of the *HLA* system. *Unlinked determinants* include family environment, polygenes, and epistatic loci that recombine freely with the *HLA* system.

On the hypothesis of *independence*, the linked and unlinked determinants do not interact. Close relatives who are affected are likely to be etiologically identical, as if there were a proportion M of pedigrees with tight linkage and gametic disequilibrium, the remainder $1 - M$ having gametic equilibrium and no linkage [3].

On the hypothesis of *interaction*, the linked and unlinked determinants have synergistic effects on the penetrance scale (although they may be additive on the hypothetical scale of liability to affection). Close relatives who are affected are likely to share some but not all IDDM determinants, simulating loose linkage but not a simple mixture of linked and unlinked pedigrees.

Segregation analysis of affection status has low power to resolve different models, especially when there is selection for multiplex pedigrees. Resolution comes from supplementary tests on HLA phenotypes of patients and the distribution of HLA identity by descent for affected sib-pairs (APPENDIX). Thomson and Bodmer [4] pioneered this approach for dominant and recessive models without sporadic cases. We have extended it to intermediate models in which dominance on the liability scale lies between 0 (recessivity) and 1 (dominance). Intermediate liability models imply dominance deviations on the penetrance scale. Penetrance may approach but never reaches the limits 0, 1 for any genotype. For each penetrance vector and gene frequency, coupling frequencies are estimated. For single antigens, a nonzero χ^2 indicates that phenotype frequencies are not correctly predicted even when coupling takes an extreme value of 0 or 1 (table 1). A larger calculation loop can maximize likelihood over different gene frequencies and penetrance vectors.

RESULTS

The best-fitting recessive model in supplementary tests has a high gene frequency (.210). It fits most distributions well, but considerably underpredicts the frequencies of *BF*F1* homozygotes among patients and of nonidentity by descent among affected sib-pairs (table 2). The fit is much improved by an intermediate model that attributes 1/3 of adult patients (but a smaller fraction of affected children) to heterozygosity for an *HLA*-linked IDDM determinant (table 3). Under the intermediate model, the coupling frequency (i.e., the proportion of *HLA* haplotypes that carry linked IDDM determinants) varies from nearly 0 for *DR2* to 1 for

TABLE 1

CHARACTERISTICS OF THREE MODELS FOR IDDM (PREVALENCE = .0045)

MODEL	GENE FREQUENCY q	DISPLACEMENT t	DOMINANCE* d	PENETRANCE			P(GENOTYPE|AFFECTION)		
				$G'G'$	GG'	GG	$G'G'$	GG'	GG
Best recessive	.210	2.62	0	.102	0	0	1.0	0	0
Best intermediate	.166	2.24	.5	.114	.005	0	.700	.295	.005
Spielman et al. [5]	.030	4.60	.74	.710	.065	0	.147	.853	0

* Expressed as liability, not penetrance. Probabilities less than .001 are reported as zero.

TABLE 2

SUPPLEMENTARY TESTS ON DATA FROM LITERATURE [8]

FACTORS	PHENOTYPE +	PHENOTYPE −	GENE FREQUENCY (p)	VALUES OF χ^2 Recessive	Intermediate	Spielman et al. [5]
B7	99	584	.131	0	0	0
B8	318	555	.104	0	0	0
B15	207	666	.070	0	0	0
B18	103	558	.046	0	0	0
DR2	44	752	.136	0	0	38.26
DR3	402	394	.125	0	0	0
DR4	501	295	.116	0	0	0
	PHENOTYPE T_1T_1	PHENOTYPE T_1T_2	PHENOTYPE T_2T_2	p_1		
*BF*F1*	8	176	1262	.012	4.88	1.85	4.07
*BF*S1*	0	42	1304	.009	0.67	0.66	0.59
*BF*F*	28	299	1119	.185	2.15	2.30	5.34
*C2*B*	2	46	386	.020	0.22	0.24	0.73
	PHENOTYPE *XX*	PHENOTYPE *AX*	PHENOTYPE *A'X*	PHENOTYPE *AA'*			
B8/B15	78	70	46	19	0.50	0.63	3.91
DR3/DR4	13	46	106	81	2.34	5.21	54.03
	IDENTITY BY DESCENT 2IBD	IDENTITY BY DESCENT 1IBD	IDENTITY BY DESCENT 0IBD				
Affected sib pairs	154	97	12		11.26	4.84	8.76
Total					22.02	15.73	115.69

TABLE 3

PREDICTION OF RISKS UNDER BEST INTERMEDIATE MODEL

AGE	PREVALENCE	PENETRANCE			P(GENOTYPE|AFFECTION)		
		G'G'	*GG'*	*GG*	*G'G'*	*GG'*	*GG*
0–4.....	.0006	.0190	.0003	0	.875	.125	0
35–39	.0062	.1470	.0074	.0001	.658	.334	.007

*BF*F1*. Gametic disequilibrium is as great as for the *A* and *B* loci, supporting tight linkage (table 4).

Spielman et al. [5] proposed an intermediate model with more dominance, a lower gene frequency, and a greater proportion of affected due to heterozygosity for linked determinants. At their values, the fit to the data of table 2 is poor ($\chi^2_9 = 115.69$). A higher gene frequency and higher proportion of cases due to homozygosity or overdominance are required to fit the DR2 and DR3/DR4 phenotype frequencies.

Although the fit of our model to B8/B15 and DR3/DR4 phenotype distribution is acceptable, this is achieved by inflating coupling frequencies above the estimates obtained for single antigens. When coupling frequencies are constrained to the latter values (i.e., .358 for B8, .330 for B15, .444 for DR3, and .644 for DR4), χ^2 increases by 9.07 for B8/B15 and by 18.57 for DR3/DR4. This indicates that these heterozygotes show overdominance (viz., the morbid risk is greater than the geometric mean of the homozygotes). Apparently, there are at least two qualitatively different *HLA*-linked IDDM determinants, one associated with DR3 and the other with DR4, and the statistical analysis confounds overdominance with recessivity. This poses unsolved analytical problems, which so far have been addressed only in

TABLE 4

ESTIMATES OF PARAMETERS FOR INTERMEDIATE MODEL

Haplotype	Coupling frequency c_i	Relative risk	Gametic disequilibrium Δ
B7.....	.079	0.52	−.011
B8.....	.358	2.34	.020
*B8**	.423	...	...
B15	.330	2.00	.011
*B15**	.443	...	...
B18	.321	1.88	.007
DR2.....	.004	0.17	−.022
DR3.....	.444	3.36	.035
DR3†	.437	...	...
DR4.....	.644	6.14	.055
DR4†	.809	...	...
*BF*F1*	1.000	5.62	.010
*BF*S1*	.315	1.76	.001
*BF*F*.....	.098	0.59	−.013
*C2*B*.....	.548	3.08	.008

* Estimated from *B8/B15* joint distribution.

† Estimated from *DR3/DR4* joint distribution.

TABLE 5

TESTS OF HYPOTHESES IN LINKAGE ANALYSIS (INTERMEDIATE MODEL)

Hypothesis	Data	χ^2_1
$\theta = .5$	Green et al. [1]	140.06
	Dunsworth et al. [2]	78.02
$M = .1 \mid \theta = 0$	Green et al. [1]	0
	Dunsworth et al. [2]	0.26
$\theta = 0$	Pooled	6.85

a paper that anticipated heterozygous expression by assuming both susceptibility determinants completely dominant to the normal haplotype, no recombination, and no effect of unlinked loci, and was considered by the authors to represent "exploratory modeling, not hypothesis testing" [6]. We do not see how to make the extension to hypothesis testing without simplifying the model and partitioning the data.

The intermediate model was entered in segregation analysis [7]. An excellent fit is obtained providing that there is an additional source of variation among sibships. The specified major locus accounts for 35% of the liability variance, and other inherited factors for 32%. This is preliminary evidence for unlinked IDDM determinants, for which linkage analysis is definitive. Linkage is overwhelmingly significant, and the data cannot be accounted for by a mixture of completely linked determinants in proportion M (table 5). Either linkage is loose, or linked and unlinked IDDM determinants interact. Estimates of coupling frequencies are consistent with samples of cases and controls (table 6).

At these values, there is striking difference between pedigrees with a *DR3/DR4* propositus and the remainder (table 7). In both samples, the first group shows no recombination: unlinked determinants are of little importance. The residual group favors loose linkage, and heterogeneity with the *DR3/DR4* class is highly significant. Given the previous results, there can be no doubt that interactive, unlinked determinants are important for non-*DR3/DR4* disease, and their random assortment with regard to the *HLA* locus simulates loose linkage. Stockert et al. [8] reported that the murine G_{1X} thymocyte antigen depends on the H-2 system in exactly the same way, and the unlinked determinants include loci epistatic with the linked determinant.

TABLE 6

ESTIMATES OF PARAMETERS FROM LINKAGE ANALYSIS (INTERMEDIATE MODEL, $M = 1$)

			COUPLING FREQUENCIES			
DATA	$\hat{\theta}_m$	$\hat{Z}$	*DR2*	*DR3*	*DR4*	Other
Green et al. [1]	.036	30.43	*	.407	.825	.032
Dunsworth et al. [2]	0	17.07	.005	.375	.488	.065

NOTE: Ratio of female to male map distance was taken as $k = 1.8$ (Morton and Lalouel [3]).
* Not distinguished from "other."

TABLE 7

LOD SCORES FOR LINKAGE AT ESTIMATES OF TABLE 6

Data	No. pedigrees	Propositus	$\hat{\theta}_m$	$\hat{Z}$	Z(0)	Z(.01)	Z(.05)	Z(.1)	Z(.2)
Green et al. [1].....	40	*DR3/DR4*	0	26.17	26.17	25.57	23.12	19.86	13.25
	52	Other	.109	8.10	2.90	4.29	7.09	8.08	6.86
Dunsworth et al. [2].....	29	*DR3/DR4*	0	10.50	10.50	10.26	9.24	7.90	5.23
	89	Other	.041	6.96	6.57	6.75	6.95	6.44	4.37

* NOTE: Lods computed by the LINKAS program (Morton and Lalouel [3]).

This interpretation is favored by the distribution of lod scores by DR phenotype (table 8). The excess of *DR3/DR4*, significant in both data sets, is specific for the linked group. Other genotypes, especially known *DR* heterozygotes, often invoke apparent recombination. It may be that nearly all apparent recombinants between IDDM and the *HLA* system involve IDDM heterozygotes.

DISCUSSION

This study illustrates principles that we believe to be important for understanding *HLA*-related diseases. First, the best model is intermediate, but with substantially greater gene frequency (.166) and less dominance (.5) than suggested by Spielman et al. [5]: rejection of one set of parameters is not evidence against the intermediate hypothesis, which may apply to many *HLA*-associated diseases. Second, under such a model the proportion of cases due to homozygosity for the linked determinant should be greater with early onset and among familial cases: correlation with age of onset is more powerful than dichotomy by a particular age of onset. The report of Svejgaard et al. [9] supports this prediction, whereas that of Barbosa et al. [10] does not. Failure of this prediction could implicate two or more linked IDDM determinants directed perhaps against different pathogens, one associated with *DR3* and another with *DR4*, which jointly determine age of onset. *HLA* types of mature-onset IDDM are of special interest, but have been little studied.

Partition of data by *HLA* type of the propositus is a powerful method to resolve etiological heterogeneity, but raises some technical problems in linkage analysis. Morton [11] showed that lod scores are not affected by selection on the main (disease) locus, but that a bias is introduced if main and test loci are selected simultaneously in offspring. The results hold under gametic disequilibrium and incomplete penetrance. There is no bias if pedigrees are classified by *HLA* types of the parents of the propositus.

To determine if we are justified in partitioning data by *HLA* type of propositus, consider selection of affected sib-pairs from a double backcross mating, $GgTt \times ggtt$, where G is a disease determinant and T is the test (*HLA*) locus. Let λ be the frequency of coupling to T among G haplotypes in double heterozygotes. Then the probability that the propositus receives GT from the informative parent is $\lambda(1 - \theta) + (1 - \lambda)\theta = \lambda + \theta - 2\lambda\theta$. Omission of this term from the lod score has no

TABLE 8

DISTRIBUTION OF PEDIGREES BY APPARENT RECOMBINATION AND DR PHENOTYPE OF PROPOSITUS

RECOMBINATION	DATA	DR PHENOTYPE			
		3/4	3/– or 4/–	3/X or 4/X	X/X
0	Green et al. [1]	35	14	10	1
	Dunsworth et al. [2]	29	20	36	3
> 0	Green et al. [1]	5	9	16	2
	Dunsworth et al. [2]	0	3	21	6

NOTE: 3/– means 3/3 or 3/blank; X means not 3 or 4. χ^2_1 for 3/4 vs. others: Green et al. [1], 15.49; Dunsworth et al. [2], 13.11; heterogeneity, 1.60.

effect if $\lambda = \frac{1}{2}$, and is negligible if $\lambda >> \theta$. This will generally be true, since high risk for disease corresponds to $\lambda \geq \frac{1}{2}$, $\theta \rightarrow 0$. We conclude that ascertainment bias when data are subdivided by *HLA* type of propositus is negligible.

As a check on this inference, we analyzed the data assuming gametic equilibrium ($c_i = q$). Estimates of recombination θ_m are not systematically higher or lower on this falsified hypothesis, in agreement with calculations of Clerget-Darpoux and Goldin [12] for tight linkage. Although significant, lod scores and χ^2 are only about half as great as when gametic disequilibrium is allowed for. This shows that computer programs like LIPED that assume gametic equilibrium are inefficient for heterogeneity analysis, even when the segregation parameters can be confirmed from other evidence.

Paradoxically, identity by descent may be discordant with linkage when there is gametic disequilibrium. Consider a pair of doubly concordant affected sib-pairs from a double backcross mating. There are two possible lod scores when selection is only through affection:

$$\text{with probability } \frac{\lambda(1-\theta)^2 + (1-\lambda)\theta^2}{(1-\theta)^2 + \theta^2} \quad \text{the score is } \log 4[\lambda(1-\theta)^2 + (1-\lambda)\theta^2] ,$$

$$\text{with probability } \frac{(1-\lambda)(1-\theta)^2 + \lambda\theta^2}{(1-\theta)^2 + \theta^2} \quad \text{the score is } \log 4[(1-\lambda)(1-\theta)^2 + \lambda\theta^2] .$$

At $\theta = 0$, the first score is negative if $\lambda < \frac{1}{4}$, which appears to give evidence against linkage. However, the expected value of the score for a doubly concordant pair when $\theta = 0$ is $\log 4 + \lambda \log \lambda + (1-\lambda)\log(1-\lambda)$, which is a minimum at $\lambda = \frac{1}{2}$. Thus, evidence for tight linkage from doubly concordant sib-pairs is strengthened by gametic disequilibrium, even though in a minority $1 - \lambda$ of cases the lod score is negative.

Suspicion that closely linked cell surface properties may interact in selection is supported by our inference of high coupling frequencies between *DR3/DR4* and linked IDDM determinants. Presumably high-risk haplotypes can be better characterized by two or more *HLA* loci than by specificities at a single locus. For diseases with suspected viral or autoimmune etiology and/or primary association with the *DR* locus, joint characterization by *DR* and the closely linked *BF*, *C2*, *C4*, and *B* loci would contribute to recognition of high-risk haplotypes ("supratypes"), reflected by coupling frequencies near unity. Besides more precise specification of risk, this would give many degrees of freedom for resolving etiological heterogeneity through association, segregation, and linkage studies of DR BF C4 C2 B phenotypes.

Recently, two studies provided independent (but equivocal) evidence for interaction of *HLA*-linked and unlinked IDDM determinants. Weitkamp [13] reported that double identity by descent for *HLA* haplotypes is more frequent in pairs of affected sibs when other sibs are normal, as if the number of IDDM factors segregating increases with the number of affected sibs. His test does not distinguish between recessivity and epistasis, but inference of multiple, interactive loci

was supported by Hodge et al. [14], who obtained lod scores of 1.99 and 2.83 under two models for linkage to the *JK* system. Although suggestive of linkage, these lods are not conventionally significant and have not been confirmed [15]. If a locus interactive with *HLA* is closely linked to *JK*, recombination with that marker should appear greater when the propositus is *DR3/DR4*.

Since any misspecification of the complex pattern for inheritance of IDDM tends to give spurious evidence of recombination between major locus and closely linked susceptibility determinant, apparent recombinants must be examined with suspicion, especially if the "recombinant" is phenotypically normal and therefore perhaps a nonpenetrant carrier.

APPENDIX

PROBABILITIES FOR SUPPLEMENTARY TESTS

The major locus parameters of segregation analysis are dominance (d), displacement between homozygotes (t), and gene frequency (q), which together with indicator-specific prevalences give the probability a_j that an individual of genotype j be affected [7, 16]. Linkage analysis uses other parameters, including vectors of haplotype frequencies $\{p_i\}$ and corresponding coupling frequencies $\{c_i\}$, where $c_i p_i$ is the frequency of the haplotype GT_i, G is an allele at the main locus favoring affection, and T_i is the ith haplotype in the HLA system [3]. By definition, if the main and test loci are in linkage equilibrium, $c_i = q$ for all i. We assume panmixia in the general population, but this does not necessarily give Hardy-Weinberg frequencies in patients. For simultaneous transmission of G and T, recombination is assumed negligible (i.e., $\theta < .01$). Only two alleles are considered at the main locus, with no effect of the test locus on disease liability unless $c_i = 0$ or 1 (in which case the test locus is said to be pleiotropic and a linked disease locus is not defined).

There is every reason to formulate supplementary tests consistently with segregation and linkge analysis. In terms of these parameters and derived probabilities (table 9), the risks for affection are defined in terms of probabilities denoted by capital letters:

General population	B
Child of affected	V/B
Sib of affected.....	$(U + 2V + B^2)/4B$
MZ cotwins of affected.....	U/B .

All risks but the last are fitted in segregation analysis, and so all acceptable models give similar predictions. An estimate of twin concordance requires fastidious ascertainment correction and may be inflated by family environment unique to MZ twins, spoiling inference of residual genetic variability.

The first evidence of disease association usually involves a dominant antigen like B8 or DR3. If p is the corresponding gene frequency with coupling c, the probability that a patient has the antigen is $(E + F)/B$, where c' is the probability that a haplotype *not* carrying the antigen has the disease gene. A detected association leads to examination of other phenotypes at the test locus, such as *B8/B15* and *DR3/DR4*. The probability that a patient be T_1T_2 is D/B.

The above two cases permit estimates of $c_1 = c$, which may be compared with the estimate from linkage analysis. Also gametic relative risks and linkage disequilibrium Δ may be calculated (under the given parameter set). More interestingly, a solution in the range $0 < c < 1$ supports linkage rather than pleiotropy.

TABLE 9

DEFINITIONS

p_i	Frequency of the ith test locus haplotype T_i	$0 < p_i < 1$,	$\sum_i p_i = 1$
q	Frequency of a tightly linked gene for disease susceptibility G	$0 < q < 1$	
c_i	Coupling frequency (the conditional probability of G in the T_i haplotype)	$0 \leq c_i \leq 1$,	$\sum c_i p_i = q$
a_2	Probability that GG is affected	$0 \leq a_2 \leq 1$	
a_1	Probability that $G\gamma$ is affected	$0 \leq a_1 \leq 1$	
a_0	Probability that the normal homozygote $\gamma\gamma$ is affected	$0 \leq a_0 \leq 1$	

$c_i p_i$ ..... Frequency of the GT_i haplotype

$B = q^2a_2 + 2q(1-q)a_1 + (1-q)^2a_0 \equiv$ prevalence in general population

$U = q^2a_2^2 + 2q(1-q)a_1^2 + (1-q)^2a_0^2 \equiv P$(both affected|two genes identical by descent)

$V = q^2a_2[qa_2 + (1-q)a_1] + q(1-q)a_1[qa_2 + a_1 + (1-q)a_0] + (1-q)^2a_0[qa_1 + (1-q)a_0] \equiv P$(both affected one gene identical by descent)

$D = 2p_1p_2[c_1c_2a_2 + (c_1 + c_2 - 2c_1c_2)a_1 + (1-c_1)(1-c_2)a_0] \equiv P$(affected, T_1T_2)

$E = p^2[c^2a_2 + 2c(1-c)a_1 + (1-c)^2a_0] = P$(affected, TT)

$F = 2p(1-p)[cc'a_2 + (c + c' - 2cc')a_1 + (1-c)(1-c')a_0] = P$(affected, TT'), $c' = (q - cp)/(1-p)$

$W = (1-p_1-p_2)^2[(c')^2a_2 + 2c'(1-c')a_1 + (1-c')^2a_0] = P$(affected, XX, $X \neq T_1$ or T_2)

$X = p_1^2[c_1^2a_2 + 2c_1(1-c_1)a_1 + (1-c_1)^2a_0] + 2p_1(1-p_1-p_2)[c_1c'a_2 + (c_1 + c' - 2c_1c')a_1 + (1-c_1)(1-c')a_0]$

$= P$(affected, T_1X), $c' = (q - c_1p_1 - c_2p_2)/(1-p_1-p_2)$

$Y = p_2^2[c_2^2a_2 + 2c_2(1-c_2)a_1 + (1-c_2)^2a_0] + 2p_2(1-p_1-p_2)[c_2c'a_2 + (c_2 + c' - 2c_2c')a_1 + (1-c_2)(1-c')a_0] = P$(affected, T_2X)

A χ^2 goodness-of-fit test is available if there are three or more test locus phenotypes. At the *BF* locus within the *HLA* system there is codominance, giving phenotypes *F1F1*, *F1X*, and *XX*, for example, where $X = F, S$, or *S1*. The probability of the first class is E/B, and of the heterozygote, F/B, where p, c are parameters of the *F1* haplotype. If c is estimated from the data, 1 degree of freedom (df) remains for the test of goodness of fit. An acceptable model tends to give a small χ^2. Possible deviations from the model include overdominance, with the heterozygote at greatest risk for disease.

The rest of the *HLA* system has been plagued by dominance, so that a *B8/B8* homozygote can sometimes be deduced in families but never recognized phenotypically. Overdominance has been claimed for some studies, but the test is complicated by ascertainment bias and expected violation of Hardy-Weinberg frequencies in patients. An exact expression, if there is no overdominance, is

$$\begin{aligned} P(XX) &= W/B \\ P(T_1X) &= X/B \\ P(T_2X) &= Y/B \ , \end{aligned}$$

where

$$c' = \frac{q - c_1p_1 - c_2p_2}{1 - p_1 - p_2}$$

is the coupling frequency for the X haplotype that is neither T_1 nor T_2. If c_1 and c_2 are both estimated, there is 1 df for testing goodness of fit.

The most familiar supplementary test is based on sib-pairs with 2, 1, or 0 haplotypes identical by descent (IBD), ignoring possible effects of polygenes and other familial factors on penetrance in sibs of probands. To avoid classification bias, parents should be typed, and sibs or children may be helpful. We assume no recombination between G and T and no selection on the test locus: this may be violated if, say, *DR3/DR4* probands were deliberately chosen for study. The probabilities are

$$\begin{aligned} P(\text{two haplotypes IBD}) &= U/(U + 2V + B^2) \ . \\ P(\text{one haplotype IBD}) &= 2V/(U + 2V + B^2) \ . \end{aligned}$$

Since the probabilities are functions only of assumed parameters, free of the c_i, there are 2 df for testing goodness of fit.

Estimation is by Newton-Raphson iteration with exact derivatives. Both Pearsonian and likelihood ratio χ^2 are calculated. As measures of association, we use the phenotypic relative risk (PRR), the gametic relative risk (GRR), the gametic disequilibrium parameter Δ, defined expected frequencies a, b, c, d in the corresponding 2×2 table as

$$\text{RR} = ad/bc$$

$$\Delta = \frac{a}{n} - \frac{(a + b)(a + c)}{n^2} \ , n = a + b + c + d \ ,$$

with standard errors that neglect imprecision in parameters not estimated in the supplementary test. Since this extends the methods of Thomson and Bodmer [4], the computer

program is called THOMBOD [17]. A possible extension to two-locus epistasis has not yet been made. The economical assumption of additive effects on the liability scale is compatible with any degree of epistasis, depending on displacements and the threshold for affection, whereas alternative formulations of epistasis are either arbitrary or involve so many parameters as to be indeterminate [18, 19].

REFERENCES

1. GREEN A, MORTON NE, ISELIUS L, SVEJGAARD A, PLATZ P, RYDER LP: Genetic studies of insulin-dependent diabetes mellitus: segregation and linkage analyses. *Tissue Antigens* 19:213–221, 1982
2. DUNSWORTH TS, RICH SS, MORTON NE, BARBOSA J: Heterogeneity of insulin dependent diabetes—new evidence. *Clin Genet* 21:233–236, 1982
3. MORTON NE, LALOUEL J-M: Resolution of linkage for irregular phenotype systems. *Hum Hered* 31:3–7, 1981
4. THOMSON G, BODMER WF: The genetic analysis of HLA and disease associations, in *HLA and Disease*, edited by DAUSSET J, SVEJGAARD A, Copenhagen, Munksgaard, 1977, pp 84–93
5. SPIELMAN RS, BAKER L, ZMIJEWSKI CM: Gene dosage and susceptibility to insulin-dependent diabetes. *Ann Hum Genet* 44:135–150, 1980
6. HODGE SE, ROTTER JI, LANGE K: A three-allele model for heterogeneity of juvenile onset insulin-dependent diabetes. *Ann Hum Genet* 43:399–409, 1980
7. LALOUEL J-M, MORTON NE: Complex segregation analysis with pointers. *Hum Hered* 31:312–321, 1981
8. STOCKERT E, BOYSE EA, SATO H, ITAKURA K: Heredity of the G_{1X} thymocyte antigen associated with murine leukemia virus: segregation data simulating genetic linkage. *Proc Natl Acad Sci USA* 73(6):2077–2081, 1976
9. SVEJGAARD A, PLATZ P, RYDER LP: Insulin-dependent diabetes mellitus, in *Histocompatibility Testing 1980*, edited by TERASAKI PI, Los Angeles, UCLA Tissue Typing Laboratory, 1980, pp 638–656
10. BARBOSA J, BACH FH, RICH SS: Genetic heterogeneity of diabetes and HLA. *Clin Genet* 21:25–32, 1982
11. MORTON NE: Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277–318, 1955
12. CLERGET-DARPOUX F, GOLDIN LR: Bias of the estimated recombination fraction and lod score due to an association between a disease gene and a marker gene. *Am J Hum Genet* 33:135A, 1981
13. WEITKAMP LR: HLA and disease: predictions for *HLA* haplotype sharing in families. *Am J Hum Genet* 33:776–784, 1981
14. HODGE SE, ANDERSON CE, NEISWANGER K, ET AL.: Close linkage between IDDM and the Kidd blood group. *Lancet* ii:893–895, 1981
15. BARBOSA J, RICH SS, DUNSWORTH TS, SWANSON J: Linkage disequilibrium between insulin-dependent diabetes and the Kidd blood group Jk-b allele. *J Clin Endocrinol Metab* 55:193–195, 1982
16. MORTON NE, MACLEAN CJ: Analysis of family resemblance. III. Complex segregation of quantitative traits. *Am J Hum Genet* 26:489–503, 1974
17. MORTON NE, RAO DC, LALOUEL J-M: *Methods in Genetic Epidemiology*. New York, S. Karger, 1982
18. THOMSON G: A two locus model for juvenile diabetes. *Ann Hum Genet* 43:383–398, 1980
19. CLERGET-DARPOUX F, BONAÏTI-PELLIÉ C, DESCHAMPS I, HOR J, FEINGOLD N: Juvenile insulin-dependent diabetes: a possible susceptibility gene in interaction with HLA. *Ann Hum Genet* 45:199–206, 1981