Am. J. Hum. Genet. 52:135–143, 1993

Linkage Analysis of "Necessary" Disease Loci versus "Susceptibility" Loci

David A. Greenberg

Department of Psychiatry, Mount Sinai Medical Center, New York

Summary

The association of some diseases with specific alleles of certain genetic markers has been difficult to explain. Several explanations have been proposed for the phenomenon of association, e.g., the existence of multiple, interacting genes (epistasis) or a disease locus in linkage disequilibrium with the marker locus. One might suppose that when marker data from families with associated diseases are analyzed for linkage, the existence of the association would assure that linkage will be found, and found at a tight recombination fraction. In fact, however, linkage analyses of some diseases associated with HLA, as well as diseases associated with alleles at other loci located throughout the genome, show significant evidence *against* linkage, and others show loose linkage, to the puzzlement of many researchers. In part, the puzzlement arises because linkage analysis is ideal for looking for loci that are *necessary,* even if not sufficient, for disease expression but may be much less useful for finding loci that are neither necessary nor sufficient for disease expression (so-called susceptibility loci). This work explores what happens when one looks for linkage to susceptibility loci. A susceptibility locus in this case means that the allele increases risk but is neither necessary nor sufficient for disease expression. It might be either an allele at the marker locus itself that is increasing susceptibility or an allele at a locus in linkage disequilibrium with the marker. This work uses computer simulation to examine how linkage analyses behave when confronted with data from such a model. The results show that if the probability of having the disease with the associated allele is less than 10 times greater than the probability of having the disease without the allele, then it may be difficult to detect linkage, even in data sets consisting of 30 nuclear families with at least two people affected. On the basis of these results, it would appear that doing linkage analysis on risk-factor data may not yield additional information about *linkage* in the usual sense but may help distinguish between different hypotheses to explain the association. As the difficulties inherent in linkage analysis for analyzing common disease become more apparent, investigators will turn to association analysis to find genes involved in disease. Before embarking on a search for a susceptibility locus, one must weigh the potential importance of such a gene against the possibility that the locus detected via association analysis may represent a relatively minor contribution to disease causality.

Introduction

The association of some diseases with certain alleles of marker loci is a poorly understood phenomenon. In some conditions, such as narcolepsy, virtually every patient with the disease has the associated HLA haplotype, which is found in only 20% of the general population (Langdon et al. 1984; Billiard and Seignalet 1985). The more common type of association is seen in conditions such as thyroid disease, where 50% of the patients have an allele that is found in 20% of the population (Farid et al. 1979; Allanic et al. 1980). What must be explained is the presence of the marker allele, in increasing proportions, in patients with the disease. Because only a fraction of the patients have the associated allele, the marker allele itself is generally thought not to be the agent involved in the disease. A frequent explanation of the phenomenon of associa-

Received February 12, 1992; final revision received September 22, 1992.

Address for correspondence and reprints: Dr. David A. Greenberg, Department of Psychiatry, Unit 1229, Mount Sinai Medical Center, New York, NY 10029.

0002-9297/93/5201-0016$02.00

tion, particularly in connection with the HLA region, is that the marker allele itself is not causing the disease but that an allele at a locus in linkage disequilibrium with the marker locus is. The observation that only a portion of the patients have the allele is then explained by recombination that has taken place since the HLA marker allele or the disease allele first appeared. However, this explanation has seldom been proved. Furthermore, some studies have found significant evidence *against* linkage for an associated trait (Go et al. 1987; Roman et al. 1992). I will show that failing to find linkage when an association exists, or even finding significant evidence against linkage, may indicate that the disease-related locus is not necessary for disease expression and is merely a *susceptibility* locus. This in turn implies that linkage analysis may be a disadvantageous way to look for susceptibility loci, not merely a less powerful way than association analysis (Bodmer 1984). I will also suggest that, because association studies are very sensitive, the effects that association detects may be minor or far removed from the proximate cause of disease. Because so much work has been done in regard to association and the HLA system, I will refer most often to HLA-disease association, but this work applies to all association phenomena.

In the present work, I distinguish two fundamental concepts: a necessary disease locus versus a susceptibility locus. By the term "necessary disease locus" I mean one whose disease allele is necessary for disease expression; that is, all affected people must have that allele. By the term "susceptibility locus" I mean one that increases susceptibility or risk for the disease but that is not necessary for disease expression. In other words, an allele at some locus makes it more likely that a person will become ill with a disease, but the presence of that allele is not the determining factor in disease expression; it merely lowers the threshold for disease. The true determining factor could be either environmental (e.g., a virus) or genetic, for example, the susceptibility locus may alter the penetrance for the main locus. Under this model, the marker allele itself could conceivably be the susceptibility allele.

These two different models of a disease locus—the necessary disease locus and the susceptibility locus—may lead to similar outcomes on association tests but to very different outcomes on linkage tests. If a *necessary* disease locus exists in linkage disequilibrium with a marker locus (e.g., the HLA region), then some allele of the marker locus will be found more frequently in the patient population than in the general population. The disease will be associated with an allele of the marker locus. The marker alleles will *also* segregate together with the disease within families, and linkage analysis will then show that genetic linkage exists between the disease and the HLA region. Thus, in the case of an associated necessary disease locus, not only is a marker allele associated with the disease, but the marker locus and the disease locus show positive results on linkage analysis. In contrast, if the locus involved is a *susceptibility* locus, then, depending on how much additional risk is conferred by the susceptibility allele and how much risk there is without that allele, one may or may not observe a segregation of the disease with the marker.

When we do linkage analysis, most of us imagine that we are looking for a gene locus that is the cause of the disease, a locus that is *necessary* for disease expression, even if not sufficient. ("Not sufficient" would mean that other factors, either genetic or environmental, in addition to the disease gene, need to be present for disease expression. These would be modeled by, for example, reduced penetrance in the linkage analysis.) Susceptibility genes, on the other hand, are neither necessary nor sufficient for disease expression. What does it mean to look for linkage to such loci, and is it even possible to find such loci by using linkage analysis?

Unfortunately, when confronted with an association, we do not know a priori whether the association points to a necessary disease gene or to a susceptibility gene. Furthermore, it turns out that it may be fruitless to do linkage analysis on data that come from a susceptibility locus model, and several differences in the expected outcome of linkage analysis exist, depending on which model is the true one. Cox and Spielman (1989), whose results prefigured those presented here, showed that an increase in risk will not necessarily show up as an increase in affected sib haplotype sharing.

In the present work, I investigate the following questions: Under what conditions is it possible to detect linkage, when the data, in fact, come from a susceptibility model? Can we determine parameters that will allow us to predict when it will be fruitful to do linkage analysis on an associated disease? How can we test whether the susceptibility model or the linkage disequilibrium model is correct for a given set of data?

The goal of the current work is to see how linkage analysis results behave when the data derive from a susceptibility locus. I will show that if the associated risk factor only *increases risk* but is not essential for disease expression, then evidence for linkage can be difficult to detect, or the evidence may even disprove

linkage. This suggests that attempts at finding some risk factors for disease through linkage analysis may not be fruitful.

Methods

Basic Assumptions

The simulation model assumes that the gene in question is a *susceptibility* gene, not a necessary one. I assume that the marker allele itself or a locus in complete linkage disequilibrium with the marker increases susceptibility to the disease. The presence of either one or two copies of the allele confers the same risk. The risk of disease, given the associated allele, is called "α." If the associated allele is not present, then the risk of having the disease is β. Thus, if $\beta = 0$, the model is identical to a dominant mode of inheritance with reduced penetrance, the penetrance being α, and the locus is necessary for disease expression. The ratio of α/β is almost identical to what is generally called "relative risk."

Simulation of Family Data

Nuclear families were simulated, with the number of offspring specified according to a well-characterized distribution (Cavalli-Sforza and Bodmer 1971, pp. 310–313; Greenberg 1984). Each parent was given a genotype by randomly choosing two alleles (from A, B, C, or D) at the marker locus. Except as noted, the gene frequency of allele A, the disease-associated allele, was .10. The frequency of the other alleles was .3 each. The alleles were then allowed to segregate randomly in the offspring. For each individual, if allele A was present, either homozygously or heterozygously, the probability of being affected was α, where α was set to .07, .15, or .5. If allele A was not present, then the probability of being affected was β, where β was set to values of 0–.05 (and higher when $\alpha = .5$). A uniform-random-number generator was used to assign affectedness status.

Ascertainment Model

In actual data collection in both linkage and association studies, families with more than one affected person are often preferentially ascertained. Thus, for these simulations, only families with at least two affected family members were included in the analyses.

Analysis

The association model described above is similar to a dominant mode of inheritance with reduced penetrance. When $\beta = 0$, the model is exactly a dominant model with penetrance equal to α. When $\beta \neq 0$, the entire population is at risk for the disease, but the risk is lower for those without the associated allele.

When $\beta \neq 0$, the model might be viewed as a dominant model with sporadics. When investigators do linkage analysis of associated diseases, they generally do not assume the existence of sporadics. Thus, initially, no sporadics were assumed in these analyses. The linkage analyses assumed a dominant mode of inheritance. In preliminary calculations, the linkage analysis assumed a range of analysis penetrance values. The highest lod score occurs when the assumed penetrance is the same as α, no matter what the value of β is. This is what is seen in the case of classical linkage analysis (i.e., when no association is present) (Greenberg 1989), even in the presence of heterogeneity (Durner and Greenberg 1992). Preliminary calculations showed that, at the low penetrances used to generate the data, assuming a penetrance of .1 for all values of α gave results almost identical to those generated when the "true" penetrance was used. Thus, for all analyses, except when $\alpha = .5$, the analysis penetrance was set to .1.

For each value of α and β, 600 nuclear families were generated. These 600 families were broken into 20 30-family data sets. A data-set size of 30 families is a good approximation to the usual data-set size in the literature. Lod scores were calculated at recombination fraction (θ) values of .01, .05, .1, .2, .3, .4, and .5. Each data set was analyzed separately using LIPED (Ott 1974), and the means and SDs of the lod scores over the 30 data sets at the seven values of θ were calculated. The means were then plotted.

Results

Bear in mind while interpreting the results that when data from the above-described association model are analyzed, β, which is the probability of having the disease without the associated allele, has an effect on linkage analysis that is similar to the effect of recombination or a "sporadic rate" in actual linkage. Thus, if allele A is segregating in a family, and if a family member without A is affected, then the linkage analysis interprets the non-A–affected person as a recombinant. This has the effects of lowering the lod score and raising the estimate of θ. As β increases, more apparent recombinants appear, decreasing the evidence for linkage. Remember, however, that the data are simulated from an association model. The disease does not

strictly segregate with allele A, and β is not a recombination fraction.

In figures 1–3, we see that as the probability of being affected without the associated allele (β) increases, the maximum of the mean lod score curve (henceforward referred to simply as "the maximum lod score") rapidly decreases. For example, when α, the probability of being affected given allele A, is .15, and $\beta = 0$, the maximum lod score is 3.2. When β increases to .01, that is, when the probability of being affected without the associated allele is only 1% compared with 15% with the associated allele, the maximum lod score drops to less than 2.0 and the estimate of θ increases. Relative risks (α/β) of less than about 10 yield barely positive lod scores. Figures 1 and 2 show the lod score versus θ curves for $\alpha = .07$ and .15, respectively. Note how quickly the maximum value of the lod score falls as β increases.

Also examined was the more extreme situation where α, the probability of developing the disease given the associated allele, was .5. For this simulation, the frequency of the associated allele in the population was set to .05. (Assuming a higher allele frequency would make it harder to detect linkage.) Again, only families with two or more affected members were ascertained. Because the risk is so high, linkage is easier to detect, but the ability to detect linkage when the associated allele is not the sole determining factor ($\beta \neq 0$) still falls quickly as β increases. At α/β ratios less than 10, detection of linkage again becomes difficult.

The analyses reported thus far assumed that $\beta = 0$, even when the true, or generating, $\beta \neq 0$. In order to show the effect of assuming $\beta \neq 0$, the analyses were rerun where $\alpha = .5$ and the correct values of β were assumed. The results, presented in figure 4, show that the value of the maximum lod score was virtually the same whether β was set to the generating value or to 0. What changed when the analysis β was set equal to the generating value (i.e., a nonzero sporadic rate in the linkage analysis) was the estimate of θ at which the maximum lod score occurred. When the analysis $\beta = 0$, the θ values at the maximum lod score increased as the generating β rose, as is seen in the figures. When the analysis β equaled the generating value, the estimate of θ was .01 (the lowest value of θ assumed) for all values of β. Thus, assuming a nonzero sporadic rate in the analysis does not appear to increase the evidence for linkage but does change (i.e., "improve") the estimate of θ.

Discussion

The reader will have noticed that the model used to simulate the data can be viewed as a sporadic model or as a heterogeneity model; those families that have the associated allele can be viewed as having one form of disease; those without, another form. From the mathematical viewpoint, these results are not unexpected: If a disease has a relatively high sporadic rate and if it is analyzed for linkage by assuming a sporadic rate of zero, then this model misspecification will lead to an overestimate of θ. However, while this may be mathematically correct, it may not be biologically meaningful; rather, the diseases may be identical in patients with and without the associated allele. In fact, this is the view taken by most investigators. They do not allow for sporadics when analyzing linkage between a disease and an associated marker. Moreover, as we have shown here, even including a nonzero sporadic rate β does not increase our understanding of what is happening biologically (see fig. 4). Assuming a nonzero sporadic rate improves the estimate of θ but has little effect on the maximum lod score. That is, when the sporadic rate is high, there simply is not very much linkage information in the data, regardless of whether $\beta = 0$ is used.

Investigators involved in studying diseases in which there are associations have little evidence that there is a difference between patients with and without the associated allele. Some diseases, such as insulin-dependent diabetes, have associations with alleles at multiple marker loci (Field 1991). While it is possible that each separate association represents a different disease with a different etiology, the parsimonious explanation is that there is only one disease but that the marker allele itself, or a very near neighbor, is increasing the susceptibility. Thus, in the absence of more biological evidence, it cannot be assumed that lack of evidence for linkage is due to heterogeneity or sporadics.

It has been suggested that the existence of an association assures that linkage will be found (Hodge et al. 1979). There has also been discussion about how linkage disequilibrium could bias parameter estimates in linkage analysis (Hodge and Spence 1981; Clerget-Darpoux 1982). However, this explanation as it is generally expressed presupposes that whatever allele is influencing disease expression is a major factor in the ultimate expression of the disease. Furthermore, under this model of association, the associated marker

allele *itself* cannot be the disease allele; otherwise, all affected people would have the marker allele.

Previous investigators have remarked on the possibility of confounding linkage and association (Hodge and Spence 1981; Clerget-Darpoux 1982). The current results suggest that linkage will be difficult to detect if α/β is about 10 or less, and lod scores will actually be negative for many values of θ if α/β is much less than 10. (As stated above, the α/β ratio is identical to what epidemiologists call "relative risk.") When α/β is less than about 10 and data are analyzed with $\beta = 0$, the lod scores will be *lowest* at a tight recombination fraction, which is difficult to explain if one is assuming linkage disequilibrium with a necessary locus. Thus, the danger is less one of confounding linkage and association than of misinterpreting what the two distinct analyses—linkage analysis and association analysis—say about any given set of data.

Linkage analysis tells us whether a marker locus is in the proximity of a disease locus by looking for cosegregation of the disease with alleles at the marker locus. Linkage analysis does not take into account *which* alleles are cosegregating. An association study, on the other hand, looks at whether patients have a given *allele* at a marker locus more frequently than does the general population. Thus, association analysis is very sensitive to the risk for disease that is conferred by alleles at individual loci. Linkage analysis is insensitive to the effects of individual alleles but puts a great deal of weight on the movement of alleles from generation to generation.

The hypothesis of a necessary disease locus in linkage disequilibrium with a marker locus might be distinguishable from the susceptibility-locus model. If a necessary disease locus is in linkage disequilibrium with the marker locus, then linkage analysis of families with and without the marker allele should yield similar lod scores for families of similar structure. This is so because, with linkage disequilibrium to a necessary disease locus, the disease locus is truly linked to the marker locus, an allele of which will segregate with the disease. Thus, regardless of whether the associated *allele* is present in the family, the disease will segregate with *some* allele at the marker *locus* within a family. Thus, families could be subdivided on the basis of whether the index case has the associated allele A. If these two subgroups of families show different linkage results, it can be argued that the marker is not near a necessary disease locus (Greenberg and Hodge 1992). This is seen, for example, in the case of insulin-dependent diabetes mellitus, where families having DR3/4 in the index case show tight linkage, whereas the remaining families show loose linkage (Morton et al. 1983).

Keeping in mind these distinctions between linkage analysis and association analysis helps one to interpret the results of the present study. Given data that come from a susceptibility model, unless a great deal of disease risk is conferred by the allele or risk factor, looking for linkage to that locus or to the risk factor may yield little new information beyond the fact of the association itself. Within the families, the individuals who do not have the risk factor but who have the disease will be interpreted by the linkage analysis as recombinants. That is why the estimate of θ is so high. This is exactly what is seen, for example, in multiple sclerosis (Tiwari et al. 1980) or in thyroid disease (Roman et al. 1992).

Another implication of these results concerns the familiality, or recurrence risk, of a disease. If α, the probability of having the disease given the allele, is arbitrarily low, but β, the probability of having the disease without that allele, is zero, then detection of association, with even a minimal sample size, is virtually certain. However, detection of linkage under such circumstances is still difficult. Families with multiple affected cases would be rare, and the low penetrance would mean that most family members carrying the associated allele would be unaffected. With such a low risk, families will not appear to be segregating a genetically caused condition and would be unlikely candidates for linkage analysis. In such circumstances, association is a very sensitive detector of genetic effects.

If linkage disequilibrium between a necessary disease locus and a marker locus is the cause of the association, how does one explain families that have multiple affected family members but in which the risk factor—the associated allele—is present in only some of the affected family members? One cannot invoke a high θ, since that presupposes that the actual risk factor is not the allele itself or a locus in linkage disequilibrium with it but is a factor that is, at best, loosely linked. However, if one assumes actual, but loose, linkage, then one cannot explain the association, since, according to that explanation, one would anticipate very tight linkage.

One possible way out of this conundrum is to assume that another locus elsewhere in the genome provides the stronger genetic influence on the disease.

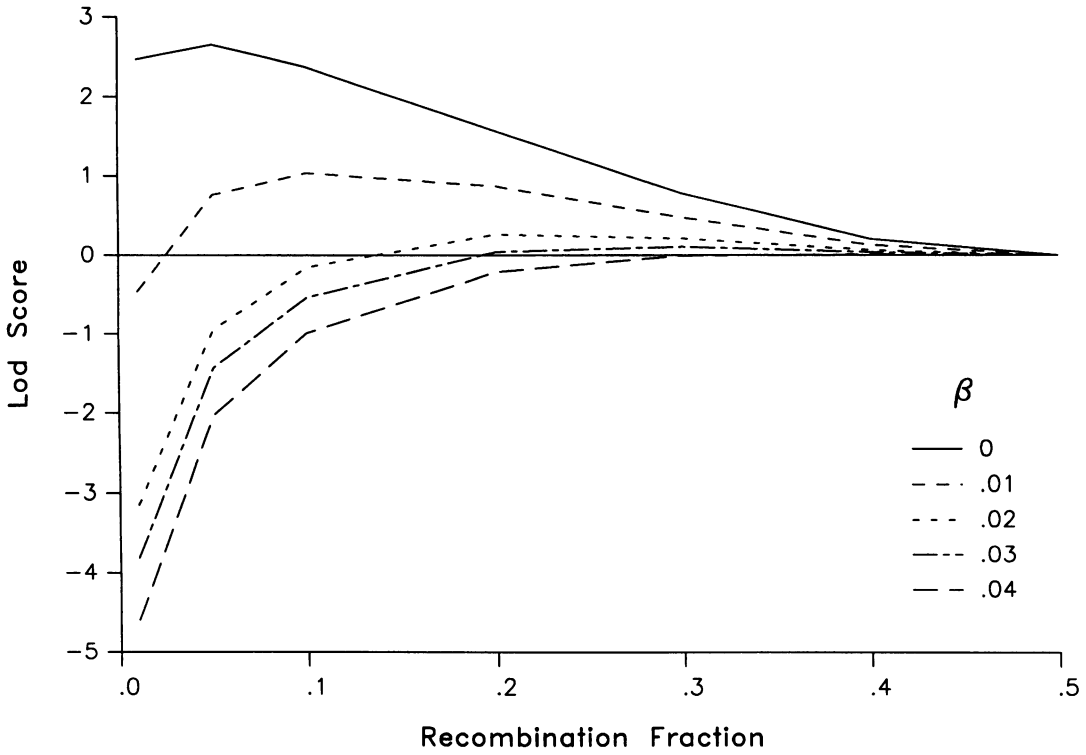

Figure 1 Graph of lod score vs. recombination fraction when $\alpha = .07$. The different curves represent the different values of β.

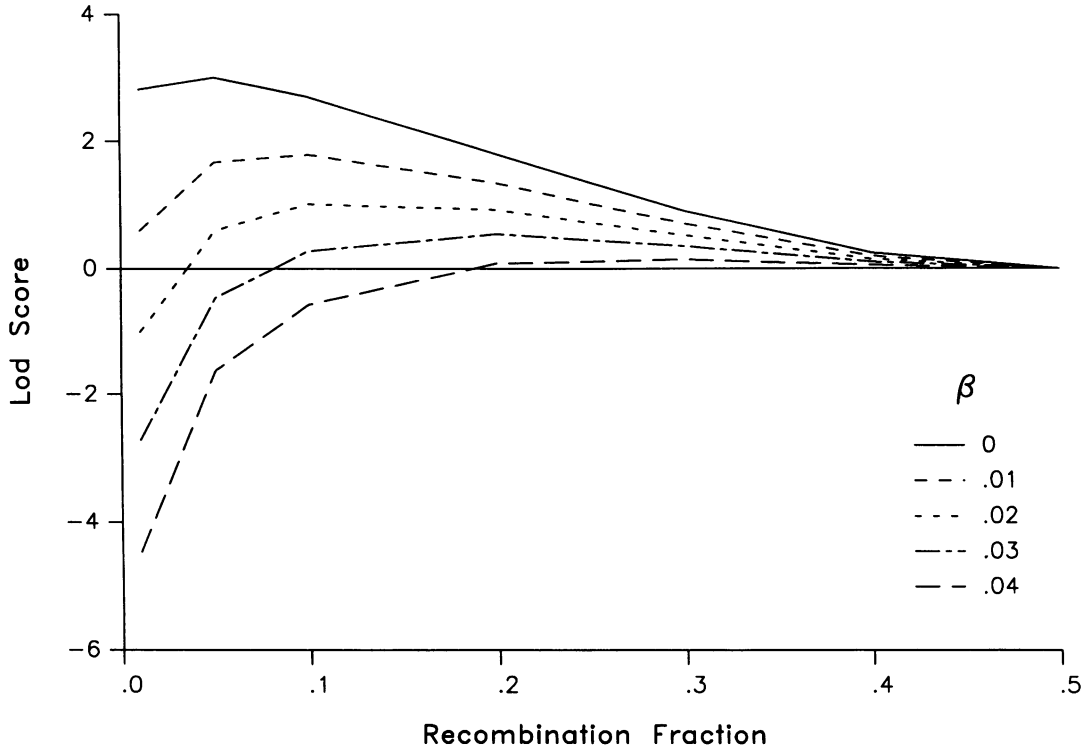

Figure 2 Same as in fig. 1, except that $\alpha = .15$

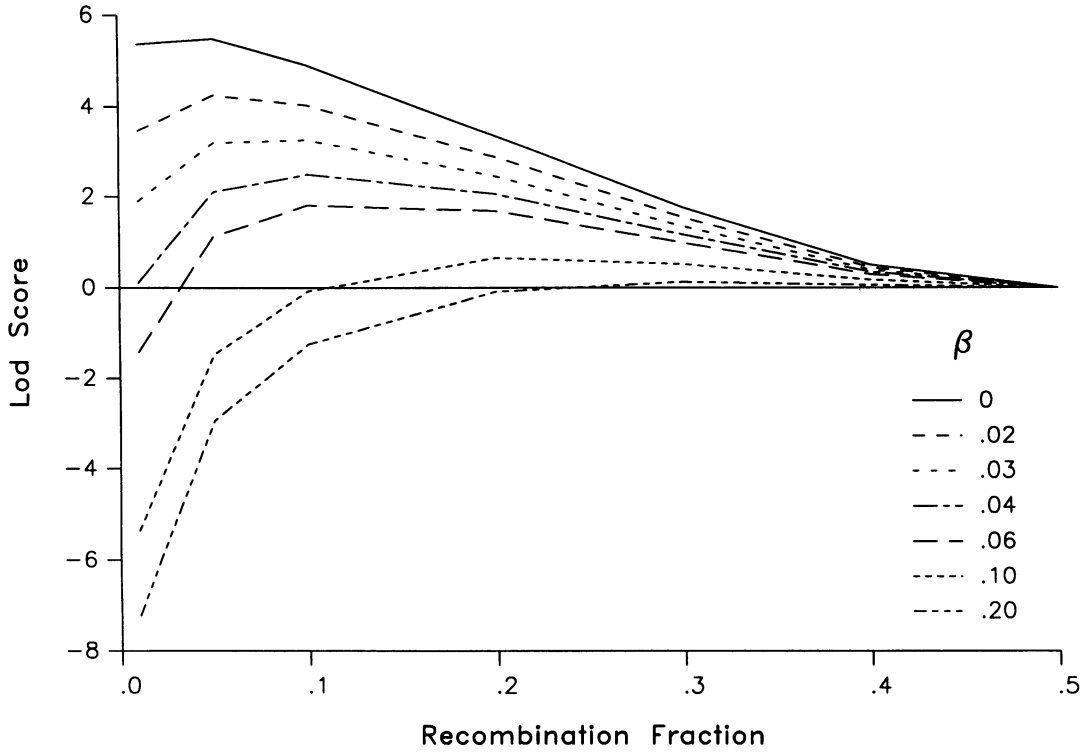

Figure 3 Same as in fig. 1, except that $\alpha = .5$. The frequency of the associated allele in the general population was .05 (see text).

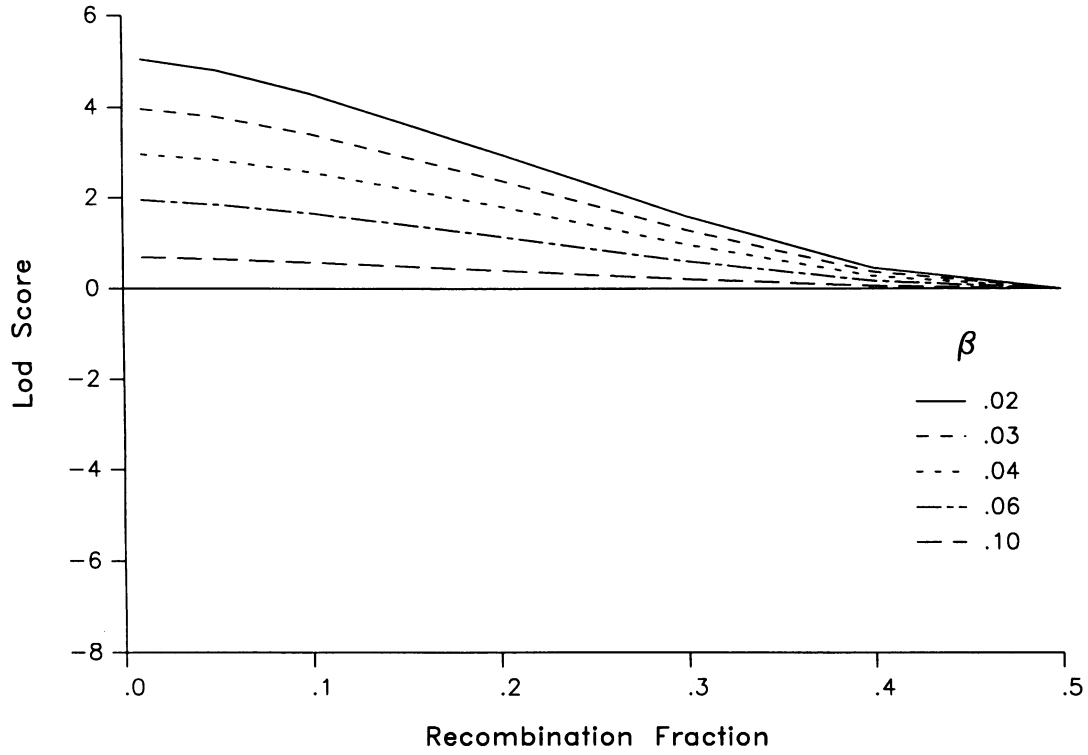

Figure 4 Same as in fig. 3, except that a sporadic rate equal to the appropriate β was assumed in the linkage analysis

This would explain the high recurrence risk. The association would then exist by virtue of the fact that the associated allele confers a risk but is not ultimately responsible for the disease. Two-locus models have been discussed for such HLA-associated diseases as coeliac disease (Greenberg and Rotter 1981) and diabetes (Thomson 1980), which has a relatively low recurrence risk in sibs (Wagener et al. 1982) but which is strongly associated with DR alleles 3 and 4 (Tiwari and Terasaki 1985).

Whatever the advantages of two-locus models for HLA-related disease, strict epistasis—i.e., where both genes are necessary for disease expression—will not explain all linkage results with HLA-associated disease. In work on thyroid disease (Roman et al. 1992), invoking a second locus will not explain the evidence against linkage of autoimmune thyroid disease with the HLA locus because the familial risk is high, even though the HLA region appears not to cosegregate with the disease. In a thorough study of multiple sclerosis, Tiwari et al. (1980) combined data from 10 separate studies. The combined data showed that there was strong evidence in favor of linkage to HLA, but the lowest estimate of θ was .15, a figure far higher than one would expect in light of the fact that the HLA region is only about 4 cM in length and that the disease locus would have to be within the HLA region, for disequilibrium to be the explanation for the association. Epistasis cannot be the whole story because an epistatically interacting two-locus model shows minimal bias in $\hat{\theta}$ (Greenberg and Hodge 1989; Vieland et al. 1992). In the case of multiple sclerosis, the observed $\hat{\theta}$ is quite different from that expected. In fact, the graphical results for the different data sets presented by Tiwari et al. (1980) bear a strong resemblance to the results of the simulations shown here. The estimate of θ was high, and, although not shown on the graphs here, the maximum lod score varied with respect to assumed penetrance, in a way similar to that of the simulations.

Clerget-Darpoux (1982) also looked at the behavior of the lod score in the presence of an association, but for that work the association was explicitly due to a necessary disease locus in linkage disequilibrium with the marker locus. Clerget-Darpoux found that unless one takes account of the association explicitly, by modifying phase probabilities, the lod score is greatly underestimated. In the case where the association is due to linkage disequilibrium with a necessary disease locus, taking account of the association means explicitly allowing for the fact that the associated allele is in coupling with the disease allele more often than 50% of the time, i.e., incorporating phase probabilities. Since the Clerget-Darpoux method was derived by assuming a "necessary" and sufficient locus, it would be inappropriate to apply it to data such as those being examined here.

Association analysis is a more sensitive detector of subtle risk factors, which may be genetic, than is linkage analysis. But an association study may not be able to show definitively the movement of the risk factor from one generation to the next, especially if there are major environmental factors involved. Linkage analysis is suited to determine the inheritance of disease and the location of the loci but may be a poor tool to use to look for risk or susceptibility factors. Association studies may be much better than linkage analysis to look for the disease locus, when the penetrance is low. As more of the human genome is mapped, this will become more practical. But the effects that are detected by association may not be crucial for disease expression. This is shown in work by Rigby et al. (1991). Those authors determined that, in rheumatoid arthritis, only a 1.61-fold increased risk to sibs, over the risk to the general population, is due to the HLA region, of an observed risk of 3.90. Whether one considers such an increase in risk strong enough to justify devoting resources in its pursuit may depend very much on the particular disease, on the characteristics of the marker, and on the knowledge of the chromosomal region.

This raises the question, How strong should the susceptibility or association be before one actively pursues the gene? The first step in such a pursuit is often linkage analysis. The results here suggest that linkage analysis may, in some cases, be useless in finding the gene locus in many cases where association exists but that it may be useful in showing whether the gene locus in question is necessary for the disease expression.

Instead of there being a confounding of linkage and association, there often appears to be no evidence of linkage when one expects there to be, if one assumes that the disease locus is necessary for disease expression. At the same time, if a susceptibility allele is the reason for the association, linkage analysis is likely to yield little more information than is already shown by the fact of the association itself.

Acknowledgments

This work was supported in part by NIH grants DK31775, NS27941, MH45212, and MH48858 and by a

fellowship in the neurosciences from the Klingenstein Foundation. My thanks to Dr. Susan Hodge for helpful discussions.

References

Allanic H, Fauchet R, Lorcy Y, Heim J, Gueguen M, Leguerrier A, Genetet B (1980) HLA and Graves' disease: an association with HLA-DRw3. J Clin Endocrinol Metab 51:863–867

Billiard M, Seignalet J (1985) Extraordinary association between HLA-DR2 and narcolepsy. Lancet 1:226–227

Bodmer WF (1984) The HLA system, 1984. In: Albert ED, Baur MP, Mayr WR (eds) Histocompatibility testing 1984. Springer, Berlin, pp 11–22

Cavalli-Sforza LL, Bodmer WF (1971) The genetics of human populations. WH Freeman, San Francisco

Clerget-Darpoux F (1982) Bias of the estimated recombination fraction and the lod score due to an association between a disease gene and a marker gene. Ann Hum Genet 46:363–372

Cox NJ, Spielman RS (1989) The insulin gene and susceptibility to IDDM. Genet Epidemiol 6:65–69

Durner M, Greenberg DA (1992) Effect of heterogeneity and assumed mode of inheritance on lod scores. Am J Med Genet 42:271–275

Farid NR, Sampson L, Noel EP, Bernard JM, Mandeville R, Larsen B, Marshall W (1979) A study of human D locus related antigens in Graves' disease. J Clin Invest 63: 108–113

Field LL (1991) Non-HLA region genes in insulin dependent diabetes mellitus. Baillieres Clin Endocrinol Metab 5: 413–438

Go RCP, Allercon GS, Acton RT, Koopman WJ, Wittor VJ, Barger BO (1987) Analyses of HLA linkage in white families with multiple cases of seropositive rheumatoid arthritis. Arthritis Rheum 30:1115–1123

Greenberg DA (1984) Simulation studies of segregation analysis: application to two-locus models. Am J Hum Genet 36:167–176

——— (1989) Inferring mode of inheritance by comparison of lod scores. Am J Med Genet 34:480–486

Greenberg DA, Hodge SE (1989) Linkage analysis under "random" and "genetic" reduced penetrance. Genet Epidemiol 6:259–264

——— (1992) Identity-by-descent (IBD) distributions in affected sib pairs (ASPs) in presence of a disease-marker association can help distinguish between models. Am J Hum Genet Suppl 51:A150

Greenberg DA, Rotter JI (1981) Two-locus models for gluten sensitive enteropathy: population genetic considerations. Am J Med Genet 8:205–214

Hodge SE, Spence MA (1981) Some epistatic two-locus models of disease. II. The confounding of linkage and association. Am J Hum Genet 33:396–406

Hodge SE, Spence MA, Cederbaum SD (1979) Hypertrophic cardiomyopathy. N Engl J Med 301:442–443

Langdon N, Welsh K, van Dam M, Vaughan RW, Parkes D (1984) Genetic markers in narcolepsy. Lancet 2:1178–1180

Morton NE, Green A, Dunsworth T, Svejgaard A, Barbosa J, Rich SS, Iselius L, et al (1983) Heterozygous expression of insulin-dependent diabetes mellitus (IDDM) determinants and the HLA system. Am J Hum Genet 35:201–213

Ott J (1974) Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am J Hum Genet 26:588–597

Rigby AS, Silman AJ, Voelm L, Gregory JC, Ollier WER, Khan MA, Nepom GT, et al (1991) Investigating the HLA component in rheumatoid arthritis: an additive (dominant) mode of inheritance is rejected, a recessive mode is preferred. Genet Epidemiol 8:153–157

Roman SH, Greenberg DA, Rubinstein P, Wallenstein S, Davies TF (1992) Genetics of autoimmune thyroid disease: lack of evidence for linkage to HLA within families. J Clin Endocrinol Metab 74:496–502

Thomson G (1980) A two locus model for juvenile diabetes. Ann Hum Genet 43:383–398

Tiwari JL, Hodge SE, Terasaki PI, Spence MA (1980) HLA and the inheritance of multiple sclerosis: linkage analysis of 72 pedigrees. Am J Hum Genet 32:103–111

Tiwari J, Terasaki P (1985) HLA and disease association. Springer, New York

Vieland VJ, Hodge SE, Greenberg DA (1992) The adequacy of single-locus approximations for linkage analyses of oligogenic traits. Genet Epidemiol 9:45–59

Wagener DK, Sacks JM, LaPorte RE, MacGregor JM (1982) The Pittsburgh study of insulin-dependent diabetes mellitus: risk for diabetes among relatives of IDDM. Diabetes 31:136–144