Am. J. Hum. Genet. 64:1728–1738, 1999

Linkage-Disequilibrium Mapping of Disease Genes by Reconstruction of Ancestral Haplotypes in Founder Populations

S. K. Service,[1] D. W. Temple Lang,[2] N. B. Freimer,[1] and L. A. Sandkuijl[3]

[1]Neurogenetics Laboratory and Center for Neurobiology and Psychiatry, Department of Psychiatry, University of California, San Francisco; [2]Bell Labs, Lucent Technologies, Murray Hill, NJ; and [3]Department of Clinical Genetics, Erasmus University, Rotterdam, and Department of Human Genetics, Leiden University, Department of Medical Genetics, Groningen University, Groningen, the Netherlands

Summary

Linkage disequilibrium (LD) mapping may be a powerful means for genome screening to identify susceptibility loci for common diseases. A new statistical approach for detection of LD around a disease gene is presented here. This method compares the distribution of haplotypes in affected individuals versus that expected for individuals descended from a common ancestor who carried a mutation of the disease gene. Simulations demonstrate that this method, which we term "ancestral haplotype reconstruction" (AHR), should be powerful for genome screening of phenotypes characterized by a high degree of etiologic heterogeneity, even with currently available marker maps. AHR is best suited to application in isolated populations where affected individuals are relatively recently descended (<~25 generations) from a common disease mutation–bearing founder.

Introduction

Linkage mapping of disease loci is based on identification of marker alleles that are identical by descent (IBD) in patients—that is, are transmitted together with a nearby disease allele from a common ancestor or founder. In genetically isolated populations, a substantial proportion of apparently unrelated patients may share chromosomal regions IBD from a remote founder. In these populations it may be possible to apply mapping methods that take advantage of the fact that such patients are IBD for marker alleles primarily in the vicinity of the disease locus (Stam and Zeven 1981; Sanda and Ford 1986; Houwen et al. 1994). Formal analysis of IBD in population samples is usually based on the evaluation of linkage disequilibrium (LD)—that is, nonrandom association between individual marker alleles and disease alleles. LD methods have been employed in a large number of studies to fine-map disease loci by using patients from founder populations (Sheffield et al. 1995). Although there is growing interest in the employment of LD approaches for initial genome-screening studies of common diseases with complex inheritance patterns, to date only a small number of rare, autosomal recessive disorders have been mapped by such methods (Houwen et al. 1994; Friedman et al. 1995; Newport et al. 1996).

Although simple measures of association are still commonly used for the fine-mapping stage of the localization of disease genes (see the review by Devlin and Risch 1995), these two-point methods are useful only over relatively small genetic distances around a disease locus and therefore are not suitable for genome-screening studies. Newer methods that evaluate LD over more than one locus within a region have been proposed. These newer approaches are based on searching for genome regions—rather than alleles—that are shared IBD among affected individuals. These approaches use more information than two-point methods, and thus they may have greater power and could be suitable for genome screening.

Several methods have been proposed that formulate expectations about haplotype or allele frequencies under an alternative hypothesis that is based on the assumption that affected individuals are descended from a common founder (Kaplan et al. 1995; Terwilliger 1995; Devlin et al. 1996; Lazzeroni 1998). The methods of Kaplan et al. (1995), Devlin et al. (1996), and Lazzeroni (1998), however, are not likely to be useful for genome-screening studies of complex traits. Both Kaplan et al. and Devlin et al. concluded that their method would be effective only for mapping of diseases in which a single haplotype has a current high frequency in disease chromosomes; this situation is unlikely to occur for any complex trait. Lazzeroni states that her method is designed for fine-mapping—that is, for situations in which genes already have been mapped to a particular chromosomal region.

Received June 2, 1997; accepted for publication March 23, 1999; electronically published April 26, 1999.

Address for correspondence and reprints: Dr. N. B. Freimer, Neurogenetics Laboratory and Center for Neurobiology and Psychiatry, Department of Psychiatry, University of California, San Francisco, CA 94143. E-mail: nelson@ngl.ucsf.edu

0002-9297/99/6406-0027$02.00

In contrast, Terwilliger (1995) presents a method that is applicable to the genome-screening stage of a study, as well as to the fine-mapping stage, and, unlike the methods of Kaplan et al. and Devlin et al., this method makes no a priori assumption about which allele(s) is overrepresented on chromosomes from affected individuals. The statistic in Terwilliger's test can be used to evaluate LD at multiple contiguous marker loci.

Terwilliger's test does not examine haplotypes per se but combines the likelihood ratio (LR) statistics from separate LD analyses of each marker. This method does not make use of all of the information in a haplotype; for example, allele-frequency differences, between patients and controls, at contiguous loci may be nonsignificantly different; however, if the haplotype frequencies are compared, they may be significantly different between the comparison groups. Thus, intuitively, methods that directly evaluate LD by using haplotype data should be more powerful than methods that examine multiple loci without evaluation of haplotype sharing; but no direct assessment of this assertion has yet been conducted. We present here an LD-mapping method targeted to genome-screening studies in founder populations. This likelihood method makes direct use of haplotype information and assumes, as does Terwilliger's method, that, under the alternative hypothesis, affected individuals are related to a common founder. We term this method "ancestral haplotype reconstruction" (AHR). In our approach, the length of the haplotypes that is wholly or partially shared by affected individuals is used in the calculation of the likelihood.

In this paper we present the results of analyses using simulated data that show AHR to be a powerful method that is robust to high levels of etiologic heterogeneity and that thus is suitable for mapping of complex traits. Furthermore, we demonstrate the additional power that is obtained from a method that uses haplotype data to evaluate LD over a series of contiguous loci, by comparing AHR with a method that evaluates LD at each of the markers independently. Although we recognize that there are several multilocus LD-mapping methods, for this comparison we use Terwilliger's method because (1) its assumptions regarding relationship to a common founder are similar to those of AHR, (2) its use is not restricted to the fine-mapping stage, and (3) it makes no assumption regarding the ancestral allele at each marker.

Methods

Description of AHR

AHR computes expected haplotype probabilities, given a founder haplotype, and estimates the following four parameters: the most likely position of the disease locus (x); the proportion of chromosomes, in the sample of affected individuals, that is likely to have descended from a common founder (α); the separation time (in generations) from a common ancestor (g); and the marker-allele frequencies. These probabilities are conditional on disease status and predict the presence or absence of LD between markers on chromosomes with the disease mutation. The observed counts of different haplotypes in the sample are assumed to be distributed as a multinomial, and the haplotype probabilities are used to calculate the likelihood that this putative founder chromosome will give rise to the observed sample of disease haplotypes. These calculations are repeated for each of the H putative founder chromosome types. These H founder likelihoods are weighted by the probability that a haplotype will be observed in the population and are summed to create an overall likelihood for the chromosomal segment (Explicit formulation of the haplotype probabilities and likelihoods are described in the Appendix). The likelihood is maximized over x, g, and α and are compared with the null likelihood. Under the null hypothesis, marker-allele frequencies are estimated, and $\alpha = 0$. The parameters x, g, and α, then, are estimated iteratively by the maximum-likelihood method, and marker-allele frequencies are estimated by counting of chromosomes and the best estimate of α (see below). Since the parameters x and g are meaningless under the null hypothesis when $\alpha = 0$, the distribution of the LR is uncertain in this setting. This type of situation is a boundary problem that can arise in some LR statistics; the admixture test for homogeneity (Ott 1983) and Terwilliger's (1995) LR test, for example, both estimate parameters under the alternative that are unidentifiable under the null hypothesis. Faraway (1993) suggests that the null distribution of the LR in these cases has half its weight concentrated on 0 and the other half on a distribution that can be approximated as $\max(X_1,X_2)$ where X_1 and X_2 are independent χ_1^2 variables. We used the formula provided by Faraway (1993) to calculate the quantiles of this distribution so that one may calculate threshold LOD scores for different P values. This formula is a simple function of the χ^2 distribution.

AHR requires estimates of marker-allele frequencies in the population, in order to calculate these haplotype probabilities. As described above, marker-allele frequencies are treated as nuisance parameters and are estimated separately under the null and the alternative hypotheses, by counting of chromosomes in the observed data. Estimation of allele frequencies is tied to the estimation of the parameter α: as α approaches 1.0, frequencies are estimated primarily on the basis of normal chromosomes; as α approaches 0, information from chromosomes of patients is incorporated into the estimation of allele frequencies. Under the null hypothesis (where $\alpha = 0$), marker-allele frequencies are estimated on the

basis of both the normal and patient samples (see the Appendix).

Simulations

We tested the power of AHR on simulated data. Each simulated chromosome had three markers, each with four alleles. The disease gene was located halfway between the first two markers. Assuming that the disease mutation, on introduction into the population, was in complete LD with a particular haplotype, and given the marker-allele frequencies in the population and the number of generations since the disease mutation was introduced into the population, one can calculate the expected haplotype distribution in the current population of chromosomes carrying the disease mutation (see the Appendix). In simulations, true affected chromosomes were randomly selected from this expected haplotype distribution, and normal chromosomes were sampled from a haplotype distribution under the assumption of linkage equilibrium (LE). The population from which simulation samples were collected was assumed to be exactly 15 generations removed from a common ancestor (in contrast, in actual data sets, this would be an average). This value of g was used because it is approximates what we expect in many of the actual populations in which we envision that AHR will be useful (e.g., those of Quebec, Costa Rica, or the Ashkenazi Jews). We reasoned that, over this number of generations, much of the IBD with the original founder chromosome would likely have decayed through recombination. With regard to the region around a disease gene, however, some affected individuals should remain IBD over chromosomal segments that are large enough to be detected by genome screening with currently available markers—for example, segments of ~5 cM.

As with any inherited trait, the effect of a susceptibility allele for a complex trait can range from dominant to recessive. We presume that LD analysis for recessive traits will be more powerful than it is for dominant traits (Chapman and Wijsman 1998), for the following reasons. At a particular locus that is contributing to disease susceptibility (for both simple and complex traits), under recessive transmission, both parents must be transmitting alleles on "disease chromosomes." In simple dominant traits, the disease chromosome is identified as the chromosome inherited from the affected parent. Under dominant transmission, however, the presence of incomplete penetrance in complex traits makes it difficult to identify the disease chromosome; it may not be possible to differentiate, a priori, between the chromosome containing a susceptibility allele and the normal chromosome. Therefore, for dominant transmission of a complex trait, one would expect $\geq 50\%$ of the chromosomes to be actually nondisease chromosomes and to add noise to the IBD signal. This is true even when chromosomal phase can be unambiguously determined. To assess the power of AHR in a situation in which we could not differentiate between a disease and normal chromosome, simulations were performed under the assumption that individuals needed only one copy of the disease allele in order to become affected. In the simulation models, the penetrance, in affected individuals, of genotypes DD (i.e., two disease alleles at the disease locus) and DN was .80, and the probability of disease in individuals with genotype NN was varied to produce different levels of etiologic heterogeneity. Under etiologic heterogeneity we include phenocopies, locus heterogeneity, and allelic heterogeneity, since these will all have an identical impact on the probability of detection of a disease locus, when AHR is used. In simulations, the disease-gene frequency was .01. These transmission models were used only to simulate data; AHR does not require specification of the model.

In our simulations, we varied the amount of etiologic heterogeneity in the affected sample, for a given transmission model and disease-gene frequency, by changing the probability of disease in NN individuals. If the fraction of affected individuals with genotype DD is x and the fraction of affected individuals with genotype DN is y, then, in the sample of chromosomes forming the patient sample, $A = n[x + .5(y)]$ will be the proportion of chromosomes that are true disease chromosomes—that is, IBD from a common founder. In these simulations, we examined heterogeneity corresponding to $A = 50\%$, 30%, 20%, and 10%. For each level of heterogeneity, we simulated by using marker-map spacings of 1, 2, 4, 6, 8, and 10 cM. Each marker had four equally frequent alleles, in all simulations. Additional simulations were performed (at $A = 20\%$ and a marker density of 4 cM) to assess the effect of (1) different numbers of alleles at each marker (we examined markers with two and six equally frequent alleles per marker) and (2) the frequency of the associated alleles at each marker (we used markers with four alleles, and the frequency of the associated allele was 10% or 50%, with equal frequencies for the remaining alleles.

We tested AHR in comparison with Terwilliger's multipoint LD LR test (MLD). To put these simulation conditions in terms of λ (the parameter estimated by MLD), these combinations of map spacing and heterogeneity resulted in the true λ for markers closest to the disease locus to range from .46 ($A = .50$; 1-cM map) to .06 ($A = .10$; 10-cM map). For each set of simulation conditions, we generated 100 replicate populations, and, for each of these 100 populations, a sample of patient chromosomes and a sample of normal chromosomes were generated. Sample sizes of 200 and 400 chromosomes were investigated, and both AHR and MLD were applied to the same data. After the 100 populations were

generated, the power of both methods was evaluated in terms of the percent of replicates that resulted in LR statistics corresponding to one-sided P values $\leq .0001$ (the traditional threshold for statistical significance in mapping studies) and to P values $\leq .01$ (used as a threshold for detection of segments of possible interest in initial genome screening), the same thresholds used by Terwilliger (1995). When the formula provided by Faraway (1993) is used, these values correspond to LOD thresholds of 3.29 and 1.44, respectively. The false-positive rate for both methods was determined on the basis of two sets of simulations. In one set, 10,000 replicates were simulated under the null hypothesis of LE, for a sample size of 400 chromosomes, 3 markers, and two equally frequent alleles at each marker; in the second set, 1,000 replicates were simulated under the null hypothesis, for a sample size of 200 chromosomes, 10 markers, and four equally frequent alleles at each marker. The second set of simulations was designed to assess the robustness of the false-positive rate estimate both to changes in the number of markers tested and to the number of alleles per marker.

Results

The power of both AHR and MLD to detect a disease locus under different degrees of heterogeneity and with marker maps of varying density is presented in figure 1, for simulations using four equally frequent alleles at all markers. Both AHR and MLD perform well with low degrees of etiologic heterogeneity and with very densely spaced markers, and they perform poorly under high degrees of heterogeneity and with widely spaced markers. Under intermediate scenarios, however, AHR consistently outperforms MLD. This advantage of AHR is observed at significance thresholds that might be used in detection of regions of possible interest ($P < .01$), as well as at more-stringent thresholds of statistical significance ($P < .0001$). In the genome-screening stage of a mapping project, relatively few individuals are genotyped by use of spaced markers, and a low threshold for significance is employed to detect regions for more-intensive study. With sample sizes of 100 case and control individuals and a 6-cM map (this marker density is currently commonly considered for genome screens in populations), the power of AHR at significance level .01 was high for several conditions of etiologic heterogeneity: 100% for $A = 50\%$, 95% for $A = 30\%$, and 69% for $A = 20\%$. The power of MLD was also high for $A = 50\%$ and $A = 30\%$; however, the power for $A = 20\%$ with a 6-cM map was only 41%. For the highest heterogeneity level considered ($A = 10\%$), both methods have very low power with a 6-cM marker map. Genome screening with a slightly denser map (4 cM) affords a considerable increase in the power of AHR,

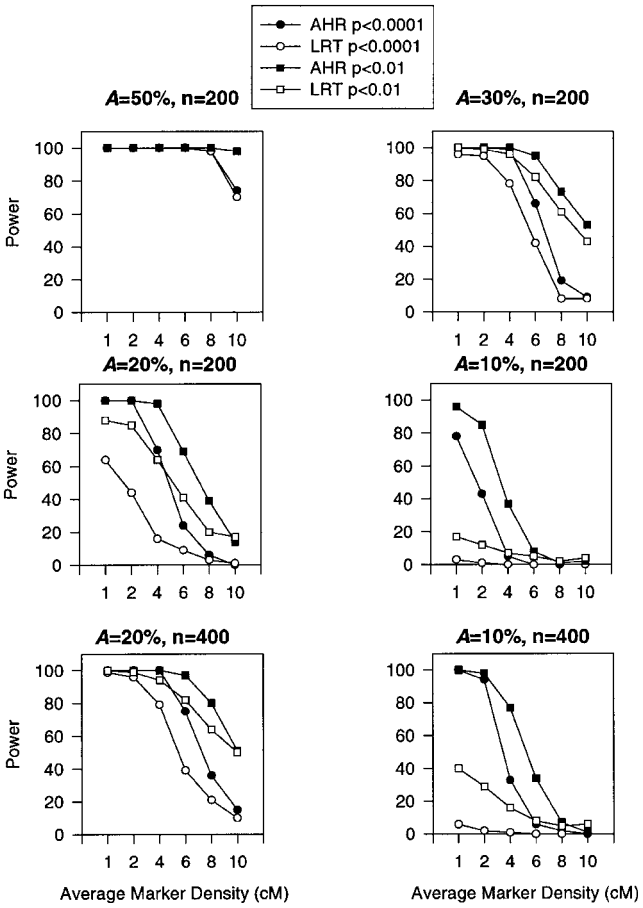

Figure 1 Percent of 100 simulation replicates that obtained results significant, at the $P = .01$ and $P = .0001$ levels, for MLD and AHR. Simulations were performed under the assumption that there are four equally frequent alleles at each marker and that the common ancestor of the affected individuals lived 15 generations ago. Sample sizes of 200 or 400 affected and normal chromosomes were used. Four different conditions of etiologic heterogeneity were examined. "A" denotes the percent of chromosomes in the affected sample that are true disease chromosomes descended from a common ancestor.

especially at higher levels of heterogeneity; for example, at $A = 20\%$, power at the .01 level was 98% for AHR, compared with 64% for MLD.

Genome areas that pass an initial threshold are usually then further explored, with a denser marker map and, possibly, with more individuals, in the hope that increased statistical significance will be obtained. With sample sizes of 100 cases and controls, marker-map densities $\leq$4 cM, and heterogeneity levels of $A \leq 20\%$, the power at the .0001 level was $\geq$70% for AHR. The power of MLD was substantially less when $A = 20\%$, for all marker maps considered. Increasing the sample size from 100 to 200 individuals (i.e., from 200 to 400 chromosomes) enables both methods to maintain higher power levels for wider marker-map spacing when there

are ≤80% normal chromosomes in the patient sample; however, at the highest heterogeneity level considered (90% normal chromosomes in the patient sample), the power of MLD is low even with a sample of 200 affected individuals. At this heterogeneity level and sample size, power at the .0001 level was 94% for a 2-cM map and 33% for a 4-cM map, when AHR was used, compared with 2% and 1%, respectively, when MLD was used.

Additional simulations (4-cM average map density, $A = 20\%$, 200 case and control chromosomes) demonstrate that, when either all markers have only two alleles or the frequency of the associated allele is high at all markers, the powers of MLD and AHR are very similar and quite low (fig. 2). Increased power is achieved with either an increase in the number of alleles per marker or a decrease in the frequency of the associated allele; and, in these situations, the power of AHR is greater than that of MLD.

In all simulations presented in figures 1 and 2, the true position of the disease locus was midway between two markers. When the true position was, instead, one-quarter of the distance from one marker to the next (making the disease locus three times closer to one marker than to the other), the power results for a single set of simulations (4-cM average marker density, $A = 20\%$, 200 case and control chromosomes, four equally frequent marker alleles) were virtually unchanged, for both methods.

In addition to having high power under different conditions, AHR produced relatively accurate estimates of both the heterogeneity parameter α and the disease location. The average estimate of α was similar to the true simulated level of heterogeneity, and, when a significant result was found, in the majority of cases the correct disease interval and founder haplotype were identified (table 1), whereas MLD identified the correct disease interval less often than did AHR. These results are expanded in figure 3, where power results for $A = 50\%$ and $A = 20\%$ are displayed for $P < .0001$ and where the correct disease interval was identified. Under conditions of low heterogeneity ($A = 50\%$), both methods are powerful for all marker maps; however, MLD is

Table 1

Average Estimates of α ($n = 600$) and Percent of Replicates with a Result Significant at $P = .0001$ That Identified Correct Disease Position and Correct Disease-Bearing Haplotype, for Four Levels of Etiologic Heterogeneity

	Average Estimated α, by	% (No.) of Replicates, with Significant Results, That Correctly Identify		
		Disease Position		
A	AHR	AHR	MLD	Disease Position and Founding Haplotype[a]
.50	.54	.97 (572)	.42 (568)	.95 (572)
.30	.29	.87 (393)	.48 (327)	.86 (393)
.20	.20	.72 (300)	.36 (137)	.72 (300)
.10	.13	.68 (126)	.75 (4)	.68 (126)

Note.—Simulations used four equally frequent alleles at each marker and sample sizes of 100 cases and 100 control individuals.

[a] Data are for AHR only (they are not applicable for MLD).

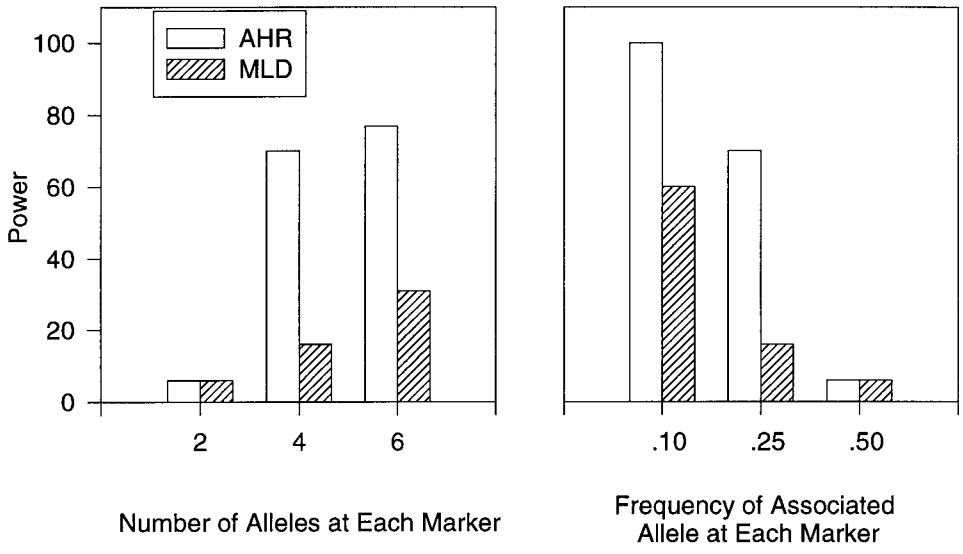

Figure 2 Effect of different numbers of alleles at each marker (*left*) and frequency of associated marker alleles (*right*) on the power of MLD and AHR. Simulations were performed under the assumptions that the marker map has an average density of 4 cM, that there are 200 case and control chromosomes, that there have been 15 generations since a common founder, and that the alleles are equally frequent (*left*) and that there are four alleles at each marker for (*right*). The results for four alleles (*left*) and frequency .25 (*right*) are from the same simulation. For each method, power is measured as the percent of 100 simulations replicates that obtain results significant at the $P = .0001$ level.

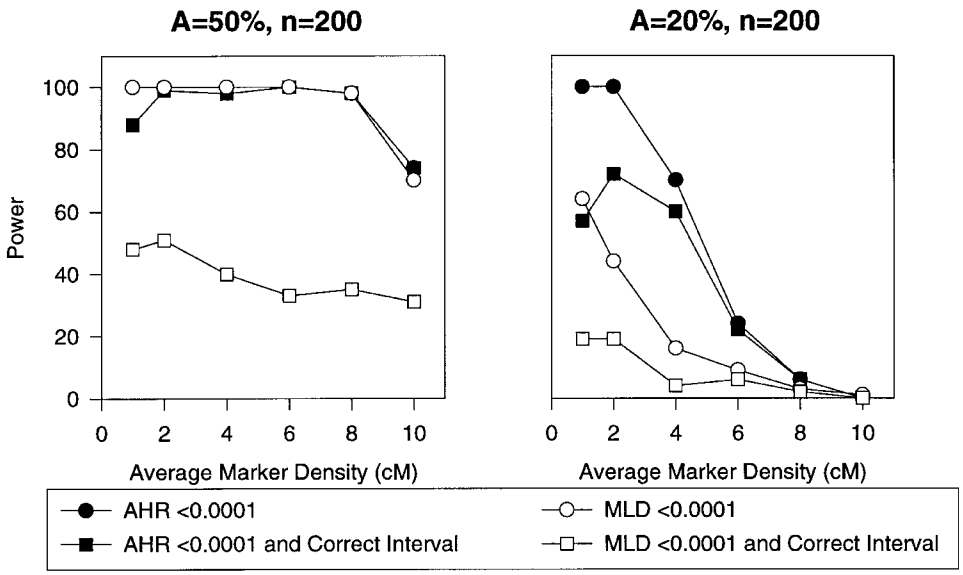

Figure 3 Comparison of the power of MLD and AHR, for two definitions of power: $P < .0001$ only (*circles*) and $P < .0001$ and correct disease interval identified (*squares*), for two different heterogeneity conditions. Simulations were performed under the assumptions that there are four equally frequent alleles at each marker and that the common ancestor of the affected individuals lived 15 generations ago. Sample sizes of 200 case and control chromosomes were used. The results indicated by circles are also presented in figure 1 and are replotted here to facilitate the comparison.

much less accurate. As discussed above, although denser marker maps afford an increase in power in heterogeneous conditions ($A = 20\%$), the signal is not as clearly localized. When less-dense marker maps are used, the LD signal is more clearly associated with the correct marker interval; however, overall power is lower (as is also shown in fig. 1).

Although an accurate, powerful test is desirable, it is also important to have a low type I–error rate. We evaluated the false-positive rates for AHR and MLD in 10,000 simulations, in the absence of a disease locus (100% normal chromosomes in the patient sample), using 200 affected individuals, two alleles at each marker, and equal allele frequencies. The fraction of false-positive results for nominal P values of .01, .001, and .0001 were .015, .0014, and .0002, respectively, for MLD, and .0075, .0005, and 0, respectively, for AHR. False-positive results from the set of 1,000 simulations using 10 markers (each with four equally frequent alleles) were evaluated only for AHR. The fraction of false-positive results for nominal P values of .01 and .001 were .008 and .001, respectively. Thus, the results for AHR conform reasonably well to the false-positive rate predicted by asymptotic theory.

Discussion

Increased attention is focusing on the possibility that genomewide association studies, using population samples, could be the approach of choice for the mapping of loci for complex traits (Risch and Merikangas 1996). It is likely that such population genetic–mapping studies will be implemented initially in relatively recently founded isolated populations (Freimer et al. 1997). Most previous LD-mapping studies in isolated populations involved small samples of individuals affected with simple recessive diseases (e.g., Houwen et al. 1994), in which, in some cases, statistical analyses may not even be necessary for isolation of the disease gene (Bull et al. 1998). Common diseases with a complex mode of inheritance have been a challenge to all mapping methods. Linkage studies have not worked well to identify susceptibility genes for such diseases, and current LD methods, even when employed in isolated founder populations, are likely to suffer loss of power from the high degree of etiologic heterogeneity expected for complex traits. Our simulations indicate that, under conditions of increased etiologic heterogeneity, AHR is more powerful than methods that look at loci independently.

AHR does not use a perfect model, and it makes simplifying assumptions regarding the pedigree and evolutionary history that connects the affected individuals, as do many approaches based on reconstruction of ancestral chromosomes. Rannala and Slatkin (1998) demonstrate that methods similar to AHR will produce maximum-likelihood estimates when the gene genealogy of the affected individuals is a "star" genealogy. A gene genealogy describes how present-day alleles are related

both to one another and to a common ancestor. In a star genealogy, the endpoints of the gene-genealogy tree are all equidistant from the root, implying that, at the locus in question, individuals are related only through one common ancestor and are essentially independent from one another. Such a genealogy may arise in rapidly growing populations (Slatkin and Hudson 1991). When the assumption of a star genealogy is approximately correct, as in our simulations, we have demonstrated AHR to be general enough to adequately represent the process that has produced the sample of affected individuals. Additionally, the use of linkage equilibrium for the null likelihood could be problematic for fine marker spacing (e.g., <1 cM) and/or a very young population (<6 generations), because of the potential for "background" LD (BLD) between markers—that is, LD unrelated to a disease phenotype (Peterson et al. 1995; Laan and Pääbo 1997). Since AHR is targeted to the genome-screening stage, where markers will be relatively far apart (the appropriate marker map spacing for a genome screen varies with the age of the population (e.g., see te Meerman et al. 1995), the impact of BLD may be minimal. When strong BLD exists in the absence of a disease locus, it may produce isolated false-positive results. When a true disease locus is nearby, BLD may add noise to the disease-associated LD signal and may weaken the power of the test. It is clear that, in order for LD mapping of disease genes to be applied widely (regardless of the method used), it will be necessary to obtain both a fuller assessment of the extent and strength of BLD in various populations and a greater understanding of the factors that influence BLD. We have previously suggested that the pattern of LD surrounding disease genes in isolated populations may be distinct from the pattern of BLD (Freimer et al. 1997). Empirical studies comparing AHR with methods used for analysis of BLD both in regions surrounding disease genes and in anonymous genome regions (in individuals who carry disease alleles and in controls) will help to clarify these issues.

Appropriate thresholds for the establishment of significant linkage have been proposed (Lander and Kruglyak 1995) and have been formulated to account for the multiple-testing situation that is a consequence of genome screening. In this formulation, as the relatedness among affected individuals decreases, the number of independent tests (and, consequently, the significance threshold) increases; for example, Lander and Kruglyak (1995) indicate a LOD threshold of 3.6 for sibs and 3.8 for second cousins. Consider, for instance, a single pair of affected siblings. Although the number of genetic variations that could be genotyped in these two patients might be in the millions (if the technology permits), there are only a limited number of recombinations that have occurred in the four meioses that separate these subjects from their parents. A limited number of tests would therefore detect all chromosomal segments that these individuals share or fail to share; additional genotyping would not provide any more information. Similarly, for a set of affected sib pairs or affected distant relatives, there are only a limited number of variants that would need to be genotyped in order to delineate all chromosomal segments shared by the patients, and the exact number required would depend on both the number of patients and their degree of relationship. Accordingly, the number of independent statistical tests that can be performed on these patients' genotypes will be limited. Clearly, the upper limit on the number of independent statistical tests is equal to the number of variants that could potentially be tested in the genome, presumably >3 million. To correct for the probability of false-positive results in that many independent tests, one would need to aim for a testwise significance level of 1.71×10^{-8}. This situation, in which all genetic variants in the genome would yield virtually independent tests, results in a sample of patients that might correspond to the testing of completely unrelated individuals. Whenever related individuals are tested, a less rigorous correction for multiple testing would be required, with a lower bound given by the correction required for very close relatives, such as siblings, in which a testwise significance level of 2.2×10^{-5} corresponds approximately to a genomewide false-positive rate of .05. We will make no attempt here to accurately define the testwise significance level required for each data set, but, on the basis of what has been shown above, it will be clear that this significance level will always be between 2.2×10^{-5} and 1.71×10^{-8}. For LD analysis, Durham and Feingold (1997) have recently used the Poisson clumping heuristic to overcome multiple-testing problems in searches for genome segments shared IBD in "unrelated" individuals. For a given study, a highly conservative guideline would be to estimate the total number of meioses in the data set (the number of chromosomes multiplied by the estimated number of generations since a common founder) and, on the basis of this, to calculate the average segment size shared by chance (in the absence of a disease locus) by all individuals (Durham and Feingold 1997). On the basis of this, one could estimate an upper bound to the number of possible independent tests, and a Bonferroni correction for this number of tests could be employed; for example, 50 individuals (100 chromosomes) separated by 20 generations have undergone 2,000 meioses, resulting in an average segment size of 0.1 cM, 33,000 potential tests, and a testwise significance level of 1.5×10^{-6}. This is likely to be extremely conservative, because, over the course of 20 generations, it is very unlikely that each segment would be independent from the adjacent segments. In the comparisons presented in this paper, any correction for multiple testing would be the same for both AHR and MLD, since both methods

move a hypothetical disease locus through a fixed map of markers (for the sake of simplicity, we used the .0001 threshold, to be consistent with the results of Terwilliger [1995]).

Oftentimes in a genome screen, an initial threshold less than that considered for genomewide statistical significance is used to identify interesting regions for follow-up. In practice, saturating such areas with denser sets of markers and increasing the size of the sample tested permits one to distinguish between true and false positives. This is both why it is important to consider the power differences that we derive from different sample sizes and marker maps and why we show power under thresholds for screening, as well as those for follow-up studies. It is worth noting that examination of haplotypes will not afford an increase in power in all situations. When the marker map is not very dense (i.e., 8 or 10 cM), the power of AHR is very similar to the power of MLD, which does not use information from haplotypes (fig. 1). Of course, the marker-map spacing that is ideal for haplotype detection will vary with the separation of patients from their common ancestor, and these actual power results are specific to our simulation conditions, in which the affected individuals are separated by 15 generations from their common disease-bearing founder. Had simulations been performed with g larger or smaller, it is likely that the power curves would be shifted along the X-axis in figure 1 (average map density [in cM])—to the left (finer map spacing needed) or the right (higher power with a less dense map), respectively.

For several isolated founder populations that are currently being used to map disease genes, it is estimated that a high proportion of affected individuals are approximately this distantly removed (i.e., $<\sim 25$ generations) from a small number of common ancestors. A number of these populations are now sufficiently large (e.g., those of Costa Rica, Quebec, Iceland, and the Ashkenazi Jews) that it is feasible to collect 200 affected individuals (as in our simulations) or even to obtain much larger samples. In the employment of AHR, it is important to keep in mind that the need to construct haplotypes in cases and controls dictates extra sampling and genotyping efforts, in that parents or other relatives must be collected to establish chromosomal phase.

It is often assumed that genomewide association studies will not be feasible until extremely dense marker maps are available. This assumption is probably correct in heterogeneous populations, such as that of the United States (Risch and Merikangas 1996), or even in more-homogeneous populations that may be very old, such as that of Sardinia or Finland; however, the marker density of current genetic maps is such that it is now realistic to undertake genome-screening studies in younger founder populations, such as those mentioned in the preceding paragraph, at a marker spacing appropriate for AHR analyses; for example, assays are now available for automated genotyping of ~1,000 microsatellite markers at a density of $\leq$4 cM; in our laboratory we are performing genome screens for complex diseases, using population samples from Costa Rica and from the Ashkenazi Jews (Escamilla et al. 1996; Mathews et al. 1997). The conditions in these screening studies approximate those for which we have shown AHR to be a powerful analysis method, even in conditions of etiologic heterogeneity. A great deal of current attention is focused on the development of genetic maps based on single-nucleotide polymorphisms (SNPs) (Wang et al. 1998). Our simulation results indicating low power, of both AHR and MLD, for biallelic markers spaced at a density of 4 cM suggest that very dense SNP maps will be needed even in recently founded populations, if complex traits are to be mapped by use of biallelic markers and LD methods. Chapman and Wijsman (1998) also have stated that extremely dense marker maps are needed for genome screens with biallelic markers, and they suggest that, even with multiallelic markers, single-marker LD testing may be feasible (in terms of power) only for simple Mendelian diseases. Our results demonstrate that multiple-marker LD tests such as AHR can be powerful for traits with complex transmission, and we anticipate that, under conditions of etiologic heterogeneity, at denser marker spacing, AHR will outperform methods such as MLD, even when biallelic markers are used.

Our simulations examined a range of heterogeneity conditions that probably encompass the level of heterogeneity to be expected in many common, complex diseases; for example, in a variety of populations, the APOE gene has been implicated in Alzheimer disease (AD)—with association between the $\epsilon 4$ allele of this gene and increased risk of AD. Samples of patients from different types of AD families report a .20–.50 frequency of the $\epsilon 4$ allele (van Gool et al. 1995). Even at the lower end of this range, the power of AHR with 100 patients is good (fig. 1).

In summary, the examination of linkage disequilibrium in founder populations is a potentially powerful way to map and localize disease genes. Complex models of transmission are not required, and sampling from a founder population makes use of the many historical recombinations between markers and disease, enabling one to pinpoint interesting regions around disease loci. Extreme heterogeneity in the patient sample has been a challenge to all LD methods. With sample sizes consisting of as few as 100 individuals, AHR has the potential to perform well under conditions of etiologic heterogeneity. In addition to identifying segments potentially containing disease loci, the method gives information about the ancestral haplotypes on which the disease mutation arose.

Acknowledgments

We would like to thank M. Slatkin for CPU time and to thank J. Terwilliger, M. Slatkin, and C. Van Duijn for helpful discussion. M. Slatkin, S. Buttrey, V. Carlton, S. Blower, L. A. McInnes, L. Kruglyak, and J. Neuhaus made useful comments on the manuscript. Software (for UNIX systems) to perform these analyses is available from S.K.S. (e-mail: service@ngl.ucsf.edu).

Appendix

Consider a chromosome with three markers, A–C, in which the original disease mutation occurred between A and B. At the time when the mutation was introduced, there were n_a alleles at marker A, n_b alleles at marker B, and n_c alleles at marker C. On the chromosome containing the disease-mutation markers A–C carried allele x. The probability that, after g generations, an affected individual carrying the original disease mutation would still have allele x at markers A–C is

$$\begin{aligned}(1-\theta_1)^g(1-\theta_2)^g(1-\theta_3)^g + (1-\theta_1)^g(1-\theta_2)^g \\ \times[1-(1-\theta_3)^g]f(x_C) + (1-\theta_1)^g[1-(1-\theta_2)^g] \\ \times f(x_B)f(x_C) + [1-(1-\theta_1)^g](1-\theta_2)^g(1-\theta_3)^g \\ \times f(x_A) + [1-(1-\theta_1)^g](1-\theta_2)^g[1-(1-\theta_3)^g] \\ \times f(x_A)f(x_C) + [1-(1-\theta_1)^g][1-(1-\theta_2)^g] \\ \times f(x_A)f(x_B)f(x_C)\ ,\end{aligned} \tag{A1}$$

where θ_1 is the recombination fraction between the disease and marker A, θ_2 is the recombination fraction between the disease and marker B, θ_3 is the recombination fraction between markers B and C, g is the number of generations since founding (i.e., since the mutation was introduced into the population), and the population frequencies of the x-allele at markers A, B, and C are $f(x_A)$, $f(x_B)$ and $f(x_C)$, respectively. Equation (A1) includes terms for the possibility of recombination between the markers and the disease locus, with the x allele at the markers then being IBS rather than IBD. The probabilities that an affected individual with the original mutation will have haplotypes other than the founding haplotype x-x-x are calculated similarly. The method as currently implemented examines three-marker haplotypes as described above; however, extension to more markers is straightforward.

These probabilities assume (1) that there is no interference in recombination and (2) that the same marker alleles are present now as were present g generations ago, in similar frequencies. If, for example, marker A has n_a alleles, marker B has n_b alleles, and marker C has n_c alleles, where $(n_a) \cdot (n_b) \cdot (n_c) = H$, then these probabilities form an $H \times H$ transition matrix $\mathbf{T}$, with row i containing the probabilities that founder haplotype i gave rise to each of the H different haplotypes in g generations. The rows of this transition matrix sum to 1. In simulations, the haplotype frequencies in the population of chromosomes carrying the disease mutation were formulated on the basis of these transition probabilities, under the assumption that the disease arose on a haplotype with the "1" allele at each of the three markers.

Because not all chromosomes in the patient sample may be true disease chromosomes from a common founder, this basic procedure has been modified to deal with heterogeneity in the sample of patient chromosomes. Some fraction, α, of the chromosomes in the patient sample will be associated with this chromosomal segment, and $(1-\alpha)$ will not be associated with it. We examine α in steps of .02, from 1.0 to .0, and for each step in α we produce a new transition matrix. If we call the transition matrix calculated under the alternative hypothesis (in which the disease locus is hypothesized to be between markers A and B) "$\mathbf{T}$" and call the transition matrix calculated under the null hypothesis of linkage equilibrium "$\mathbf{T}_n$," then this new transition matrix is calculated as

$$\mathbf{T}^* = \alpha\mathbf{T} + (1-\alpha)\mathbf{T}_n\ . \tag{A2}$$

Once these transition probabilities are estimated, the likelihood that a particular founder chromosome will give rise to the observed sample of patient haplotypes in g generations is easily estimated; for example, if one assumes that the disease mutation arose on a chromosome with haplotype K, then the likelihood (L_K) that this chromosome was the founder of the present-day sampled patient chromosomes is given by the multinomial

$$L_K = \prod_{i=1}^{H} (P_{K,i})^{Y_i}(le_i)^{N_i}\ , \tag{A3}$$

where i indexes the H potential haplotypes for the two markers, $P_{K,i}$ is the probability that the ancestral disease chromosome haplotype K gave rise to a haplotype of type i in g generations (entry K,i of transition matrix $\mathbf{T}^*$ from eq. [A2]), le_i is the linkage-equilibrium probability of haplotype i, and Y_i and N_i are the observed numbers of haplotype i in the sample of patient and normal chromosomes, respectively ($\Sigma_i[Y_i]$ is the number of chromosomes from affected individuals in the samples to be analyzed, and $\Sigma_i[N_i]$ is the number of normal chromosomes in the samples to be analyzed). The likelihood in

equation (A3) assumes that all individuals are independent. Although one might consider that, after many generations of separation from a common ancestor, these individuals are independent, they are, in fact, related through a complex and unknown pedigree. The simplification of considering that individuals the are independent makes the likelihood much more tractable to computation (for an alternative approach using single biallelic markers, see the work of Rannala and Slatkin [1998]). In practical application of AHR, it is important that individuals not be too closely related to each other, to minimize the possibility of detection of IBD regions that do not contain disease-susceptibility genes. In population genetic–mapping studies in our laboratory, we stipulate that affected individuals should not share ancestry for at least four generations (although the vast majority of affected individuals will be more distantly related to one another). The expected amount of chromosome sharing among such affected individuals, in the absence of a disease gene (Durham and Feingold 1997), is below the resolution that we would anticipate for genome-screening studies using AHR.

The H likelihoods are then summed, and they are weighted by the probability that a particular haplotype will be observed in the population, to produce an overall likelihood:

$$L = \sum_{i=1}^{H} f_i L_i \ , \qquad \text{(A4)}$$

where f_i is the frequency of haplotype i in the population. The haplotype frequencies f_i are estimated under the assumption of linkage equilibrium. The a posteriori most likely "founder" disease chromosome can be identified as the haplotype that adds the most weight to this sum.

The estimation of marker-allele frequency is tied to the estimation of the parameter α, and the frequencies are reestimated with each increment of α. The frequency of allele x on marker A would be estimated as

$$f(x_A) = \frac{N_D f(x_A)_D (1-\alpha) + N_N f(x_A)_N}{N_D (1-\alpha) + N_N} \ ,$$

where N_D and N_N are the number of sampled patient and normal chromosomes, respectively, and the frequency of allele x on marker A, as estimated on the basis of patient chromosomes, is $f(x)_D$ and, as estimated on the basis of normal chromosomes, is $f(x)_N$.

The likelihood in equation (A4) is compared with the null likelihood, which is generated under the assumption that $\alpha = 0$. This LR is then maximized over α, position of the disease locus (x), and g.

Equations (A3) and (A4) parallel the approach taken by Terwilliger (1995) for two-point analysis of LD between disease and a single marker locus. When examining multiple marker loci, Terwilliger sums the log likelihoods of equation (A4) for the different markers (at common values of α and g), whereas AHR examines entire haplotypes. In equations (A3) and (A4), $P_{K,i}$, le_i, and f_i refer to haplotype frequencies in AHR but are allele frequencies in Terwilliger's method.

References

Bull LN, van Eijk MJT, Pawlikowska L, DeYoung JA, Juijn JA, Liao M, Klomp LWJ, et al (1998) A gene encoding a P-type ATPase mutated in two forms of hereditary cholestasis. Nat Genet 18:219–224

Chapman NH, Wijsman EM (1998) Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. Am J Hum Genet 63:1872–1885

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

Devlin B, Risch N, Roeder K (1996) Disequilibrium mapping: composite likelihood for pairwise disequilibrium. Genomics 36:1–16

Durham LK, Feingold E (1997) Genome scanning for segments shared identical by descent among distant relatives in isolated populations. Am J Hum Genet 61:830–842

Escamilla M, Spesny M, Reus V, Gallegos A, Meza L, Molina J, Sandkuijl L, et al (1996) Use of linkage disequilibrium approaches to map genes for bipolar disorder in the Costa Rican population. Am J Med Genet 67:244–253

Faraway JJ (1993) Distribution of the admixture test for the detection of linkage under heterogeneity. Genet Epidemiol 10:75–83

Freimer NB, Service SK, Slatkin M (1997) Expanding on population studies. Nat Genet 17:371–373

Friedman TB, Liang Y, Weber JL, Hinnant JT, Barber TD, Winata S, Arhya IN, et al (1995) A gene for congenital, recessive deafness *DFNB3* maps to the pericentromeric region of chromosome 17. Nat Genet 9:86–91

Houwen RHJ, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, Sandkuijl LA, Freimer NB (1994) Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet 8:380–386

Kaplan NL, Hill WG, Weir BS (1995) Likelihood methods for locating disease genes in nonequilibrium populations. Am J Hum Genet 56:18–32

Laan M, Pääbo S (1997) Demographic history and linkage disequilibrium in human populations. Nat Genet 17: 435–438

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lazzeroni LC (1998) Linkage disequilibrium and gene mapping: an empirical least-squares approach. Am J Hum Genet 62:159–170

Mathews CA, DeMille MMC, Herrera LD, Reus VI, Lowe TL, Spesny M, van de Wetering BJM, et al (1997) Population genetic mapping studies of Tourette syndrome. Am J Hum Genet Suppl 61:A401

Newport MJ, Huxley CM, Huston S, Hawrylowicz CM, Oostra BA, Williamson R, Levin M (1996) A mutation in the interferon -γ-receptor gene and susceptibility to mycobacterial infection. N Engl J Med 335:1941–1949

Ott J (1983) Linkage analysis and family classification under heterogeneity. Ann Hum Genet 47:311–320

Peterson AC, Di Rienzo A, Lehesjoki AE, de la Chapelle A, Slatkin M, Freimer NB (1995) The distribution of linkage disequilibrium over anonymous genome regions. Hum Mol Genet 5:887–894

Rannala B, Slatkin M (1998) Likelihood analysis of disequilibrium mapping, and related problems. Am J Hum Genet 62:459–473

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Sanda AI, Ford JP (1986) Genomic analysis. I. Inheritance units and genetic selection in the rapid discovery of locus linked DNA markers. Nucleic Acids Res 14:7265–7283

Sheffield VC, Nishimura DY, Stone EM (1995) Novel approaches to linkage mapping. Curr Opin Genet Dev 5: 335–341

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129:555–562

Stam P, Zeven AC (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30:227–238

te Meerman GJ, Van Der Meulen MA, Sandkuijl LA (1995) Perspectives of identity by descent (IBD) mapping in founder populations. Clin Exp Allergy 25:97–102

Terwilliger JD (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am J Hum Genet 56: 777–787

van Gool WA, Evenhuis HM, van Duijn CM (1995) A case-control study of apolipoprotein E genotypes in Alzheimer's disease associated with Down's syndrome: Dutch study group on Down's syndrome and aging. Ann Neurol 38: 225–230

Wang DG, Fan J, Siao C, Berno A, Young P, Sapolsky R, Ghandour G, et al (1998) Large scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082