Am. J. Hum. Genet. 62:690–697, 1998

Significance Levels in Complex Inheritance

Newton E. Morton

Human Genetics, University of Southampton, Princess Anne Hospital, Southampton, United Kingdom

Summary

A LOD score $\geqslant 3$ is necessary but not sufficient to make a linkage test reliable, and this applies to complex inheritance as well as to major loci. Factors that affect this threshold are considered here. A LOD score as small as 2 is suggestive but is unreliable except as confirmation of either a significant linkage or a strong candidate locus. A threshold as great as 4 is unnecessarily conservative if multipoint tests are used sensibly. Marker density is not a major factor, and biases in the evaluation of LOD scores—especially inadequate allowance for estimation of nuisance parameters in multiple models—are paramount. Allelic association increases resolution for oligogenes within a candidate region and remains the only practical method to locate polygenes. The method sketched here combines multipoint linkage and allelic association to test efficiently for a regional candidate locus.

Introduction

The past year has seen vigorous debate about significance levels appropriate for complex inheritance (Curtis 1996; Risch and Botstein 1996; Witte et al. 1996), partly in response to guidelines proposed for the interpretation and reporting of linkage results (Lander and Kruglyak 1995). However, important considerations have been overlooked, and so it seems worthwhile to set out the argument systematically, both as it has been applied for 40 years to major loci (Morton 1955) and as it now applies to complex inheritance. Numerical results are based on the human genome, but the logic is general.

The Assumptions

We are interested in a trait that is significantly heritable. The trait is often related to disease, and this is reflected by use of terms such as "affection," but the same principles apply to nonclinical variation. The trait is defined by phenotypes that may be affection status, a continuous variable, or a polychotomy of normal ("diathesis") and/or affected ("severity") (Morton et al. 1991). Inference of heritability is based on family resemblance that significantly exceeds population frequencies. Statistical significance must be high enough to justify additional analysis, and the nominal significance level is usually much less than .01. Power to reject the null hypothesis of no heritability usually approaches 1, and the null hypothesis for traits subjected to genetic analysis is rarely true. Therefore heritability is seldom in doubt, even if confounding with family environment is not excluded. However, heritability is only a necessary—not sufficient—basis for detecting a contributory locus, which has become the raison d'être for genetic analysis in complex inheritance.

Sewall Wright (1968) introduced the term *leading factors* for incompletely penetrant genes with relatively large and therefore potentially recognizable effects on the trait. The remaining *minor factors* are more numerous but of individual effect so small that their isolation by linkage analysis may be prohibitively difficult and costly. *Factor* is nowadays an uncommon synonym for *gene,* and it is usual to replace "leading factor" by *oligogene* and "minor factor" by *polygene,* which in some cases may be allelic with major genes (fig. 1). Whereas major genes have effects so large that their parameters can be estimated with sufficient accuracy by segregation analysis, this is unreliable for oligogenes and hopeless for polygenes, which can be identified only by allelic association (linkage disequilibrium) with candidate loci recognized through either structure, function, or alleles with greater effect. Allelic association can be detected over small distances that are rarely >1 cM, unless there is a favorable combination of few founders and generations. Linkage provides a complementary strategy for oligogenes and major genes, since it can identify a candidate region.

Success in detecting some oligogenes raises the hope—but does not provide certainty—that every heritable trait has one or more oligogenes. For a particular oligogene we may entertain either a "parametric" model, which specifies allele number (usually two), allele frequency, and penetrances, or a "nonparametric" model,

Received November 5, 1996; accepted for publication December 16, 1997; electronically published March 4, 1998.

Address for correspondence and reprints: Dr. Newton E. Morton, Human Genetics, University of Southampton, Level G, Princess Anne Hospital, Coxford Road, Southampton SO16 5YA, United Kingdom. E-mail: nem@soton.ac.uk

0002-9297/98/6203-0024$02.00

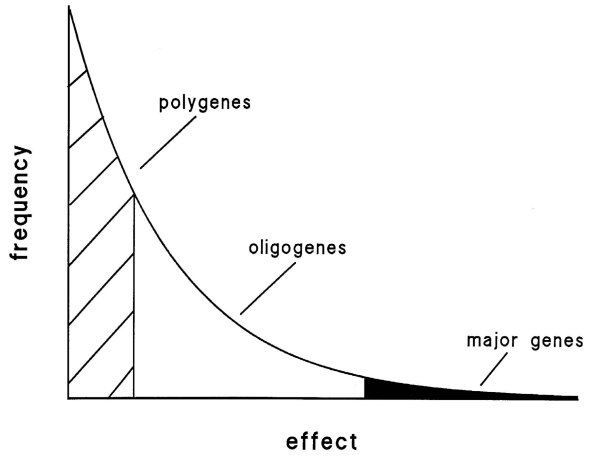

Figure 1 Allelic classes: the postulated inverse relation between frequency of a contributory locus and its effect (Wright 1968). The values of 0.1 and 1.5 relate to the β model in which the risk for affection in the child of an affected individual is $\lambda_0 = \lambda e^{\beta}$, where λ is the population prevalence (Morton 1996).

which recognizes, on some plausible liability scale, only one or more variance components that are functions of unspecified parameters. I assume that the test is expressed as a *l*ogarithm of *od*ds (LOD), either for single loci or, preferably, for multiple loci. Most of this report is devoted to linkage, first for major genes and then for oligogenes, but allelic association is considered below (see Allelic Association).

In addition to the trait, there are n marker loci chosen to localize one or more oligogenes, for which some of the markers may be candidates. However, most markers are not candidates. Ideally, markers should be uniformly spaced, highly polymorphic, and without null alleles, and each allele should be distinguishable without error and with complete penetrance.

Besides trait(s) and markers, there is a sampling procedure and a model that determine how an inference is reached. This is the source of nearly all controversy about significance levels. Is sampling fixed size or sequential? Is n specified or virtually infinite? What is the distribution of gene effects? What logic determines acceptable type 1 and type 2 errors, and what confidence can be placed in the specification of these errors? An appropriate LOD-score threshold is necessary but not sufficient to obtain reliable results (Risch and Botstein 1996).

Classical Linkage

These problems were faced many years ago when traits of interest were determined by major genes and the number of potential markers was small. A major gene is sufficient to cause affection with high penetrance or has a megaphenic effect on a quantitative trait (i.e., the displacement is greater than the SD), but different loci may act in different families. A locus at which at least one allele is a major gene for the trait is called a *major locus*. If there are sporadic cases, linkage analysis may be restricted to multiplex families. Even then the genetic parameters can be estimated by segregation analysis with little error, and so a parametric model is the method of choice. A simple argument showed that a LOD $\geqslant 3$ is required to provide a prima facie case for linkage (Morton 1955).

To generalize, let f be the expected number of linkages falsely asserted by the procedure (type 1 errors), and let t be the expected number of true linkages detected by the procedure. Then the *reliability* is

$$\rho = t/(f + t) \ . \qquad (1)$$

This is the conditional probability that a significant linkage be true (i.e., not a type 1 error). A test with low reliability is unacceptable, since it will lead to claims of nonexistent loci and to fruitless attempts at replication. Therefore we require, for credible assertion of linkage, that ρ approach 1, which implies that t/f be large. Contrary to recent discussion, it is not sufficient that f be small (Lander and Kruglyak 1995), since t may be smaller.

In the simplest case there is only one major locus and a random marker. If the procedure has a significance level α to assert linkage falsely and has a power P to detect linkage if present, the expected numbers of significant tests are $f = \alpha(1 - \phi)$ and $t = \phi$P, where ϕ is the probability that the major locus is on the same chromosome (i.e., is syntenic) with a random marker. Whereas P is a function of sample size, ϕ is independent. At a time when there were no critical observations on human chromosomes (and the diploid number was still thought to be 48), the available data on mouse, maize, and *Drosophila* suggested $\phi = .05$ for human autosomes (Morton 1955). The basic formula is $\phi = \Sigma_I L_i^2 / (\Sigma L_i)^2$, where L_i is proportional to the number of expressed loci on the ith chromosome. If gene number is proportional to physical length, then the estimate for autosomes is .054 (Renwick 1969), becoming .051 if the X chromosome is included, as is appropriate for complex inheritance. Recent estimates follow that convention. If gene number is proportional to genetic length (as suggested by Haldane and Smith [1947]), the corresponding estimate is .051. For the 25,000 loci in the location database *ldb* (which includes nonexpressed loci and is biased by unequal effort devoted to different chromosomes), $\phi = .053$ (Collins et al. 1996*a*). A reasonable consensus is .052.

Of course, many syntenic pairs of loci are too far apart

to show linkage. Haldane and Smith (1947) suggested, "chiefly from a comparison with the known linkage values of *Drosophila,*" that it may not be a bad approximation to assume, for syntenic loci, that the recombination fraction θ has a uniform distribution from 0 to $\frac{1}{2}$. This was confirmed analytically under interference for chromosomes of length 100 cM, but longer chromosomes have an excess (and smaller chromosomes have a deficiency) near $\theta = .5$ (Morton 1955). Over all chromosomes, the human genome has a nearly uniform distribution, with a pole at .5, justifying $g(\theta)$ and P below.

Every procedure can give an estimate of the power $p(\theta)$ to detect a recombination fraction $\theta < \frac{1}{2}$. The distribution of θ for two random loci is approximately

$$g(\theta) = \begin{cases} 2\phi & \text{for } 0 < \theta < \dfrac{1}{2} \\ 1 - \phi & \text{for } \theta = \dfrac{1}{2} \end{cases} ,$$

from which mean power for two loci on the same chromosome may be obtained by integrating over the uniform distribution from 0 to .5, since $P = 2\int_o^{.5} p(\theta)d\theta$. It is therefore unnecessary to partition $g(\theta)$ arbitrarily into a part where power is high and a complementary part where it is low. When the only useful markers were blood groups and isozymes, null alleles and relatively low heterozygosity reduced power in any particular study. Since strong evidence for linkage usually required several samples, it seemed useful to assume sequential sampling. If only the recombination fraction θ is estimated, an inequality on the maximum LOD $\hat{Z}$,

$$P(\hat{Z} > \log A \mid H_0) < 1/A , \qquad (2)$$

provides a tight bound on the significance level α even in the small samples that were then usual. The inequality holds for fixed sample size but is more conservative (Haldane and Smith 1947; Wald 1947; Morton 1955). In the limit for large samples, $(2\ln_e 10)\,\hat{Z}$ has a χ^2 distribution with q df if q parameters are estimated, except in certain cases where parameter constraints make the number of df smaller than the number of parameters estimated. For linkage, these cases are admixture with unlinked loci, where the test is one sided, and the Δ model that is confined to a "possible triangle," both of which are best applied (if at all) subsequent to linkage detection (Collins et al. 1996*b;* Morton 1996). Equation (2) is invalid for $q > 1$, except under conditions noted in the Discussion.

Morton (1955) determined the expected LOD, power, and average sample size for double-backcross sib pairs of unknown phase and a single disease locus with complete penetrance, concluding that P values of .28–.71 are realistic. A test with high power to detect $\theta < .2$ was estimated to have $P = .56$. On the basis of these considerations the values in table 1 seem reasonable. Most statistically significant results at $\alpha = .05$ are false, as are a large fraction at $\alpha = .01$. All but a few percent of results significant at $\alpha = .001$ are true. This corresponds to a LOD of 3 in a sequential test but to a LOD of only 2.07 in large-sample theory. Use of a LOD of 3 to define significance corresponds to $\alpha = .0001$ in large-sample theory (Morton 1955) and therefore to a reliability of .996. Rao et al. (1978) found, for 1,665 pairs of loci, a reliability of .991 at $\log A \geqslant 3$. With rare exceptions, linkage at a correctly computed LOD $\geqslant 3$ is true.

Of course, this argument leaves the investigator free to respond to a smaller but suggestive LOD by tests of allelic association in the candidate region, identification of candidate locus by structure or function, recognition of homology with a candidate region in another mammal, accelerated sequential sampling, typing of additional markers, or in other ways. However, linkage evidence should not be considered more than suggestive if the LOD is <3, and confirmation requires replication or other evidence.

Multiple Markers

The simplest extension of this theory is to one major locus and n markers segregating independently and analyzed separately. The probability of at least one spurious linkage is $1 - [1 - \alpha(1 - \phi)]^n$, and the probability of detecting the major locus is $[1 - (1 - \phi)^n]P$. If n is small, then the expected number of false linkages is $\sim n\alpha(1 - \phi)$ and the expected number of true linkages is $\sim n\phi P$, and so ρ is nearly independent of n. If n is large, then, for adequate sample size, the expected number of true linkages approaches 1, and so ρ is $\sim 1/(1 + n\alpha)$, or .99 for $n = 100$ and a LOD of 3 in large samples. If there are r major loci, detection of linkage with any one of them will increase effort on the remainder, making P approach 1. Then ρ becomes $\sim r/(r + n\alpha)$, or .99 for $r = 3$, $n = 300$, and a LOD of 3 in large samples. These considerations explain why multiple markers and candidates have not invalidated the canonical LOD-3 criterion for major loci.

At this point it is instructive to consider the argument for an indefinitely dense genome scan, which assumes that it is cost effective and can be described by the Ornstein-Uhlenbeck model for Brownian motion (Lander and Botstein 1989). Neither assumption is self-evident.

At present, a dense genome scan is prohibitively expensive. The current standard aims at a highly informative marker every 10 cM, beyond which it is preferable to increase sample size rather than marker density. Sample collection is costly, but increasing density is not confirmatory, since it is not independent evidence. Data errors are conserved, and information about previously uninformative meioses is confounded with amplification

of type 1 errors: greater marker density for meioses that by chance give reduced recombination can only strengthen that spurious evidence, because of the part-whole correlation. Besides that fatal flaw, the strategy of seeking confirmation by dense mapping in the original sample has two other disadvantages: the information per additional marker is small because much of the information is extracted by the initial sparse markers, and, arguably, it requires a higher significance level. Lander and Kruglyak (1995, p. 244) have claimed that a hierarchical search has "essentially the same [false positive rate] as if a dense map had been used throughout the genome (D. Siegmund, personal communication)," but it is not clear whether that is a proof or a conjecture. In either event, it is not germane to confirmation by sparse markers in independent samples. It is generally agreed that confirmation of preliminary evidence for linkage requires independent samples, and conclusive evidence of linkage ordinarily precedes dense mapping. Allelic association provides a stronger argument for dense mapping in a candidate region (see the section titled "Allelic Association," below), since it gives independent evidence. However, candidate regions identified by a genome scan are typically 10–30 cM in length (Barnes et al. 1996), and so further refinement by linkage is still desirable.

The assumptions underlying the Ornstein-Uhlenbeck model are also arguable. Let X be a threshold for a standardized normal deviate x with density $f(x)$ corresponding to the one-sided significance level

$$\alpha = \int_X^\infty f(x)dx \ .$$

Lander and Botstein (1989) assumed that multilocus linkage tests are accurately described by Brownian motion, that testing continues until the process is stationary (without excluding regions that are resoundingly negative), that the central-limit theorem holds and so $X^2 = (2\ln 10)\hat{Z}$, that $f(X)/X \sim \alpha$, and that the genome scan really is infinitely dense. Given all these assumptions, the equivalent number of equally spaced, independent markers is $n' = 2\mu GX^2$, where G is the genome length in morgans and μ is the recombinational autocorrelation function for a pair of relatives. Equivalence signifies that the expected number of type 1 errors in such a genome scan is $n'\alpha$. Values of μ have been tabulated by Lander and Kruglyak (1995; also see table 1), ranging from 1 for parent-offspring to 16/5 for second cousins. They accepted a map length of $G = 34.5$ morgans, citing Gyapay et al. (1994), who gave $G = 36.9$, but recent evidence favors $G = 38$ (Collins et al. 1996*b*). Calculations (table 2) predict that, on an infinitely dense scan, loci separated by a few centimorgans behave as independent, since n' is large relative to the genome length (3,800 cM), whereas in a sparse scan the LODs for loci several centimorgans apart are significantly correlated, as expected. Given highly polymorphic markers, it is not clear how increased density could make closely linked markers effectively independent. This paradox is explained by the fact that Ornstein-Uhlenbeck theory applied to linkage enumerates local maxima, so that there can be an indefinitely large number of points exceeding a given threshold within a few centimorgans, whereas neighboring points fall short by a small amount. This is an unrealistic representation of a linkage test and exaggerates the LOD required for significance.

Multipoint analysis is becoming the standard for linkage detection. It leads to partition of a chromosome into a small number of regions that are believed to contain, at most, one candidate locus, and so only the global maximum in each region is sought, with local maxima being ignored. The power of multiple markers can be increased, and the probability of at least one type 1 error can be reduced, by multipoint analysis. For example, in the β model the likelihood is a function of only two parameters: S is the location of a disease gene within the region, and β is the logarithm of the relative risk $\lambda_o/\lambda = \lambda_m/\lambda_o$, where λ is the population risk and the subscripts "o" and "m" denote offspring and MZ twins, respectively (Morton 1996). This model has been found to be the most powerful nonparametric test for single loci (Collins et al. 1996*b*) and multipoint analysis (Lio and Morton 1997). The number of independent tests is no greater than the number of candidate regions (Zeng 1994). Under these conditions the effective number of markers is so small that single-locus theory cannot be misleading, and so Ornstein-Uhlenbeck theory in its present form is irrelevant.

The application of Ornstein-Uhlenbeck theory or a simulation-based analogue could be refined to ignore local maxima near a regional maximum and thus could provide a faithful description of dense mapping, but it cannot provide an estimate of reliability unless it makes additional assumptions including the distribution of effects: the model considers only type 1 errors, and so it does not provide the assurance of table 1 that significant linkages are reliable when type 2 errors are considered (eq. [1]).

Distribution of Effects in Complex Inheritance

Wright (1968) did not specify a density for gene effects, but some attempt must be made if the power of a linkage test is to be discussed. By contrasting "a few essential genes" with "a host of quantitatively varying series," he implied that the density is monotonic decreasing on the effect, as in figure 1. Accepting this as plausible, I tentatively suppose that the density for all contributory genes is exponential: $f(\beta) = \omega e^{-\omega\alpha}$, where

Table 2

Critical Values for Significance in an Infinitely Dense Scan

		n' FOR[a]		
Z	χ^2	Double Backcrosses: $\mu = 1$	Sibs: $\mu = 2$	Second Cousins: $\mu = \frac{16}{5}$
1.3	5.991	455	911	1,457
2.0	9.210	700	1,400	2,240
3.0	13.816	1,050	2,100	3,360
4.0	18.421	1,400	2,800	4,480
5.0	23.026	1,750	3,500	5,600

NOTE.—$G = 38$; $\chi^2 = 2Z\ln 10$; and $n' = 2\mu G\chi^2$.

[a] Such that the mean number of type 1 errors is $n'\alpha$ as $k \to \infty$.

$\beta > 0$ is an effect that is proportional to the additive genetic variance on the exponential scale and is a function of gene frequency, dominance, and displacement (Morton 1996). Then oligogenes correspond to values of β greater than some arbitrary threshold τ that defines detectable genes. The mean value of β in the interval $\tau < \beta < \infty$ is $\bar{\beta} = 1/\omega + \tau$. There is little chance of detecting a value of β as small as .1, even in large samples (Morton 1996), and so τ is taken as .1. Evidence from insulin-dependent diabetes (IDDM) suggests that $\bar{\beta}$ lies in the range .25–.32 (Collins et al. 1996*b*), giving an interval of 4.5–6.6 for ω. A rough estimate of the number of oligogenes r may be obtained as four times the relative recurrence risk, or ~10 for IDDM (Collins et al. 1996*b*).

Obviously these suggestions are speculative unless they can be confirmed by data on other diseases. If they are tentatively accepted, then results for major genes can be carried over to oligogenes, which are associated with substantial power P for the highly polymorphic markers now in use. Although power to detect an oligogene may be less than for a major gene, the number of oligogenes is often greater. It would seem that a LOD of 3 provides adequate evidence for linkage if the assumptions under which the LOD was calculated are valid and the sample size is reasonably large.

Trait Associations

Associations of traits are relevant to complex disease. The following argument is heuristic, but not rigorous for $\pi > 0$. If the disease is defined by m associated traits (e.g., IgE and skin-prick score for atopy, or narrow and broad definitions for schizophrenia), and if the mean absolute correlation between traits is π, then the equivalent number of independent traits is $1 + (m - 1)(1 - \pi) = m(1 - \pi) + \pi$, and the equivalent number of markers is $\xi = n[m(1 - \pi) + \pi]$. Assuming that power approaches 1 for multiple traits, reliability of a significant linkage when there are r oligogenes and each marker is analyzed separately is $\rho = r/(r + \xi\alpha)$. The probability of at least one type 1 error in ξ tests is $1 - (1 - \alpha)^{\xi} \doteq \xi\alpha$, and so the effect of m associated traits can be fully compensated by increasing the critical LOD to $3 + \log \xi$. Transforming the traits so that they are uncorrelated gives $\pi = 0$ and $\xi = nm$, and the above argument is exact.

Allelic Association

Allelic association resulting from pleiotropy or linkage disequilibrium is more challenging. Misleading association due to population stratification can be avoided by suitable controls. Allelic association can have greater power than linkage when the marker is a candidate, but power declines much more rapidly with genetic distance. Both observation and theory suggest that an oligogene in a population near equilibrium must be considerably <1 cM from the marker if allelic association is to be detected in a sample of reasonable size (Lawrence et al. 1994). Furthermore, the total number of alleles for highly polymorphic markers is an order of magnitude greater than the number of markers. If a total of v alleles are tested in n loci, the number of independent tests is $k = v - n$. The combination of high k and low P makes a LOD of 3 inadequate to assure high reliability for allelic association with a random marker. One way to circumvent this is to restrict tests of allelic association to (*a*) strong candidate loci selected a priori from structure or function and (*b*) markers selected a posteriori from linkage, taking $3 + \log k$ as the critical value. Another tactic is to sample populations with high linkage disequilibrium due to admixture or inbreeding, but then founder effects and assortative mating extend even to unlinked genes. Neglect of pedigree structure increases noise, and replication is difficult. Meanwhile, allelic association with sparse random markers is a minefield in which a LOD of 3 or even 4 is likely to be a type 1 error.

Obviously, this caveat should not discourage tests of allelic association within a candidate region, where they enhance the resolution of linkage and provide the only practical method to locate polygenes. Therefore, tests of allelic association give a rationale for markers denser than linkage can justify. This led Risch and Merikangas (1996) to suggest that a dense genome scan would become practical, in effect making the whole genome a candidate region. If $3 + \log k$ is used, the critical LOD would be 9 for $k = 1,000,000$. It would require enormous samples to give such evidence, unless the gene effect were so large that linkage with a much smaller number of markers would also be feasible. The Risch-Merikangas dilemma can be avoided by multilocus tests in a small number of regions, which could be the same as those used for linkage. Likelihood over a region can

be expressed as a function of only two parameters: location S, which is the same as for multilocus linkage, and a parameter of exponential decline, $\in$. A plausible model uses the Malecot formula for isolation by distance, $\Delta = (1 - \mathrm{L})M \exp[\in \delta_i(S_i - S)] + \mathrm{L}$, where Δ is a measure of allelic association, L is a bias due to spurious association when alleles are pooled into two classes, $M = 1$ if there is a unique susceptible haplotype and negligible mutation, S_i is the physical location of the ith marker locus, and

$$\delta_i = \begin{cases} 1 & \text{if } S_i < S \\ -1 & \text{else} \end{cases} .$$

Malecot theory has been implemented by Collins and Morton, including both efficient combination with linkage evidence and extension to the transmission/disequilibrium test (Spielman and Ewens 1996), where the apparent segregation frequency is $(1 + \Delta)/2$. This approach reduces the number of tests to one per chromosome region, regardless of how many markers and alleles are typed within the region. The current formulation uses multiple pairwise logic, assuming independence of marker pairs. A goodness-of-fit test is provided for this assumption. Given a sufficiently dense map, power to detect allelic association is higher than for linkage. Under these conditions the canonical LOD of 3 is justified, instead of 9 as contemplated by Risch and Merikangas (1996). The cost of mapping is therefore reduced by a factor of three, whether measured in dollars or in time and effort. A trustworthy physical map is the sine qua non for this method, illustrating a principle well known to geographers and classical geneticists but not self-evident to molecular biologists—that is, that exploration without a good map is possible but costly. Unfortunately, there is no international effort, except as a by-product of genome sequencing, to create a map at the required density.

Efficient mapping by multipoint allelic association in candidate regions does not replace a genome scan with linkage. In yeast and other lower organisms, a large proportion of sequenced loci belong to no known gene family and have no known function. So long as the human genome is incompletely sequenced, many candidates will not be precisely located. We must anticipate that only a small fraction of candidates may be both recognized and localized, especially for behavioral traits that have no certain correspondence in other organisms. It seems likely that, in the foreseeable future, linkage will retain its utility for oligogenes, with allelic association increasing the resolution for known candidate loci and suspected candidate regions. Whether this speculation is true or false is irrelevant, since the theory sketched here allows efficient combination of the two independent sources of information into a single likelihood ratio testing for a regional candidate.

Discussion

Reliability of a statistical test depends on a sampling procedure with specified power. Results do not apply to a sample with unspecified power, for which reliability as a special case of equation (1) is $\rho = \phi \mathrm{L}(\theta) / [\phi \mathrm{L}(\theta) + (1 - \phi)\mathrm{L}(.5)]$, where ϕ is the probability of synteny and L (θ) is the likelihood of the sample as a function of the known recombination fraction θ. If θ has the uniform distribution from 0 to .5, L (θ) is replaced by $\Lambda = 2\int_0^{.5} \mathrm{L}(\theta)\, d\theta$. In small samples, L ($\theta$) for a small value of θ may be much greater than Λ. Then if a small estimate is used for θ, reliability will be low, as observed by Skolnick et al. (1984) and Genin et al. (1995). At least two—and usually more—samples are required for confirmation of oligogenic linkage. Power and average sample size are specified in sequential analysis, which minimizes the number of observations for a given risk of error (Wald 1947; Morton 1955). The investigator should beware of samples with low power—and, therefore, low reliability.

Given a sampling procedure with good power, a LOD of 3 provides convincing evidence of linkage unless the assumptions of the test are violated. Risch and Botstein (1996) have argued that dense markers and heterogeneity are insufficient to cause low reliability for claims of linkage in complex inheritance, and the fundamental problem is low power of oversimplified models, especially single-locus models. Earlier work had suggested good power for single-locus models (Clerget-Darpoux et al. 1986), but generality against complex alternatives has not been demonstrated. "Nonparametric" tests based on identity by descent make minimal assumptions and, on the basis of present evidence, are not significantly oversimplified in large samples with known gene frequencies (Collins et al. 1996*b;* Morton 1996).

Risch (1991) identified multiple testing procedures as a major source of error. If g models are specified a priori and the one with highest LOD is selected, a conservative correction is to take $3 + \log g$ as the critical value. However, it is not uncommon to maximize the LOD by estimation of parameters other than θ, and then this correction is no longer conservative. Three different strategies should be distinguished. The first (which is certainly valid in large samples and which, I believe, retains the small-sample exactness of eq. [2]) maximizes L (θ) and L (.5) with respect to each model. The second (which is liable to abuse) estimates fewer parameters for L (.5) than for L ($\hat{\theta}$)—say, p and $p + q$, where p may be 0 (Greenberg 1989). In large-sample theory, the LOD $\hat{Z}_q$ obtained in this way corresponds to a χ^2 of

$(2\ln 10)\hat{Z}_q$ with q df, except for constraints discussed after equation (2). There is a χ_1^2 with the same large-sample level of significance, and there is an equivalent $\hat{Z}$ of $\chi_1^2/(2\ln 10)$. For example, suppose that $q = 5$ and $\chi^2 = 15.086$, with nominal significance .01 and $\hat{Z}_q = 3.28$. At the same nominal level of significance, $\chi_1^2 = 6.635$ and $\hat{Z} = 1.44$. This is the value that should be used in metanalysis. Failure to allow for the way in which the parameters were chosen has inflated a modest LOD by nearly 2 (much greater than log $q = .70$) and has created spurious significance. An example with $q = 5$ is not extreme, since a parsimonious two-locus model has at least six segregation parameters, in addition to one or two for recombination. Undoubtedly many investigators would, in all innocence, fiddle more than six parameters if they had a computer program to do it. Although the proposed correction may be adequate, the first strategy is preferable.

Maximizing the LOD rather than L (θ) is far worse—in Risch's (1991, p. 1064) words "beyond statistical correction." In extreme cases, phenotypes may be reclassified, but usually the genetic model is manipulated (Clerget-Darpoux et al. 1986; Hodge et al. 1997). The rigorous mathematical theory for maximum likelihood has no parallel for maximal LODs, except in the trivial case $p = 0$, $q = 1$, when they are identical. The maximal LOD (MOD) approach was introduced to compensate for computer programs, such as LINKAGE, that are incapable of estimating segregation and linkage parameters simultaneously and do not allow for complex inheritance or incomplete ascertainment. Advocates try arbitrary sets of single-locus parameters. A large proportion give likelihoods so small that they are incompatible with the data and so do not contribute to g, which therefore may be as small as 2 (Hodge et al. 1997). This is an empirical result for one simulation, and it need not hold generally. Moreover, the best estimates found in this way are neither meaningful nor efficient. If the ascertainment scheme violates parametric analysis, a nonparametric method is better.

Whereas major loci lend themselves to LODs with small-sample exactness, LODs for complex inheritance are usually derived from statistics that have a χ^2 distribution in large samples. This introduces some error, which can be addressed by simulation. If α is the significance level estimated by simulation with the corresponding χ_1^2, then $Z = \chi^2/(2\ln 10)$ is an appropriate LOD. Approach to large-sample theory is somewhat slower when $s(s-1)/2$ pairs are made from $s > 2$ relatives, which Fisher (1935) showed to be asymptotically valid and efficient. This has been confirmed by Collins and Morton (1995) through simulation. Nevertheless, confusion persists that multiple pairs are "wrong" and that they should be either weighted by $1/s$ or replaced by $s - 1$ cyclical pairs (Daly and Lander 1996). Of course that can be done, but it is inefficient and still does not give small-sample exactness. This favors parametric methods that do not reduce data to pairs of relatives, but a realistic theory that includes two or more loci, allowance for ascertainment, and multiple markers has yet to be implemented. Identity-by-state methods should be avoided, since the combination of multiple pairs and homozygous parents makes them unreliable even in large-sample theory, and they provide no analog for a β or parametric model.

Multipoint analysis for a valid model introduces only the complication that location and effect are estimated simultaneously under linkage, whereas location cannot be estimated under H_0. When the estimate of effect is >0, we have $p = 0$ and $q = 2$, whereas $q = 1$ when the estimate is $\leqslant 0$. Metanalysis requires that effect and location be averaged over samples, weighting by information. Under a fixed-effects model, this gives $q = 1$. Under the null hypothesis the expected value of effect must be 0. Therefore, methods that either constrain estimates to positive values or do not estimate an effect are unsuitable for metanalysis.

Currently, the type 1 error for claims of linkage and association in complex inheritance is unacceptably high. It can be controlled by fastidious avoidance of the worst methods, attention to significance levels, and independent replication if both negative and positive data are publicly available for metanalysis. This conflicts with industrial secrecy as genome screening gravitates to venture capital. Unless the necessity for open verification in multiple samples is appreciated, the pursuit of profit will lead, paradoxically, to unprofitable cloning and eventual clinical trial of type 1 errors.

Forty years ago the fundamental problem in mapping major loci (apart from a scarcity of markers) was identified as excessive reliance on large-sample theory and modest significance levels. Complex inheritance adds much greater uncertainty about genetic parameters and therefore gives more opportunity for error, both by reliance on large-sample theory and, especially, in choice of a model. This is not an insuperable problem unless it is conjoined with low power and inadequate allowance for estimation of nuisance parameters in multiple models.

References

Barnes KC, Neely JD, Duffy DL, Freidhoff LR, Breazeale DR, Schou C, Naidu RP, et al (1996) Linkage of asthma and total serum IgE concentration to markers on chromosome 12q: evidence from Afro-Caribbean and Caucasian populations. Genomics 37:41–50

Clerget-Darpoux F, Bonaiti-Pellie C, Hochet J (1986) Effects of misspecifying genetic parameters on lod score analysis. Biometrics 42:393–399

Collins A, Frezal J, Teague J, Morton NE (1996*a*) A metric

map of humans: 23,500 loci in 850 bands. Proc Natl Acad Sci USA 93:14771–14775

Collins A, Maclean CJ, Morton NE (1996*b*) Trials of the beta model for complex inheritance. Proc Natl Acad Sci USA 93: 9177–9181

Collins A, Morton NE (1995) Nonparametric tests for linkage with dependent sib pairs. Hum Hered 45:311–318

Curtis D (1996) Genetic dissection of complex traits. Nat Genet 12:356–357

Daly MJ, Lander ES (1996) The importance of being independent: sib pair analysis in diabetes. Nat Genet 14:131–132

Fisher RA (1935) The detection of linkage with "dominant" abnormalities. Ann Eugenics 6:187–201

Genin E, Martinez M, Clerget-Darpoux F (1995) Posterior probability of linkage and maximal lod score. Ann Hum Genet 59:123–132

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

Gyapay G, Morissette J, Vignol A, Dib C, Fizames C, Bernardi G, Lathrop M, et al (1994) The 1993–94 Généthon human genetic linkage map. Nat Genet 7:246–339

Haldane JBS, Smith CAB (1947) A new estimate of the linkage between genes for haemophilia and colour-blindness in man. Ann Eugenics 14:20–31

Hodge SE, Abreu PC, Greenberg DA (1997) Magnitude of type I error when single-locus linkage analysis is maximized over models: a simulation study. Am J Hum Genet 60: 217–227

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

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

Lawrence S, Beasley R, Doull I, Begishvili B, Lampe F, Holgate ST, Morton NE (1994) Genetic analysis of atopy and asthma as quantitative traits and ordered polychotomies. Ann Hum Genet 58:359–368

Lio' P, Morton NE (1997) Comparison of parametric and nonparametric methods to map oligogenes by linkage. Proc Natl Acad Sci USA 94, 5344–5348

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318

——— (1996) Logarithm of odds (lods) for linkage in complex inheritance. Proc Natl Acad Sci USA 93:3471–3476

Morton NE, Shields DC, Collins A (1991) Genetic epidemiology of complex phenotypes. Ann Hum Genet 55:301–314

Rao DC, Keats BJB, Morton NE, Yee S, Lew R (1978) Variability of human linkage data. Am J Hum Genet 30:516–529

Renwick J (1969) Progress in mapping human autosomes. Br Med Bull 25:65–73

Risch N (1991) A note on multiple testing procedures in linkage analysis. Am J Hum Genet 48:1058–1064

Risch N, Botstein D (1996) A manic depressive history. Nat Genet 12:351–353

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

Skolnick MH, Thompson EA, Bishop DT, Cannon LA (1984) Possible linkage of a breast cancer-susceptibility locus to the ABO locus: sensitivity of LOD scores to a single new recombinant observation. Genet Epidemiol 1:363–373

Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 59:983–989

Wald A (1947) Sequential analysis. Wiley, New York

Witte JS, Elston RC, Schork NJ (1996) Genetic dissection of complex traits. Nat Genet 12:355–356

Wright S (1968) Evolution and the genetics of populations. Vol 1: Genetic and biometric foundations. University of Chicago Press, Chicago

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468