Am. J. Hum. Genet. 65:876–884, 1999

Replication of Linkage Studies of Complex Traits: An Examination of Variation in Location Estimates

Seth B. Roberts,[1] Charles J. MacLean,[2] Michael C. Neale,[2] Lindon J. Eaves,[1,2] and Kenneth S. Kendler[1,2]

Departments of [1]Human Genetics and [2]Psychiatry, Virginia Institute of Psychiatric and Behavioral Genetics, Medical College of Virginia of Virginia Commonwealth University, Richmond

Summary

In linkage studies, independent replication of positive findings is crucial in order to distinguish between true positives and false positives. Recently, the following question has arisen in linkage studies of complex traits: at what distance do we reject the hypothesis that two location estimates in a genomic region represent the same gene? Here we attempt to address this question. Sampling distributions for location estimates were constructed by computer simulation. The conditions for simulation were chosen to reflect features of "typical" complex traits, including incomplete penetrance, phenocopies, and genetic heterogeneity. Our findings, which bear on what is considered a replication in linkage studies of complex traits, suggest that, even with relatively large numbers of multiplex families, chance variation in the location estimate is substantial. In addition, we report evidence that, for the conditions studied here, the standard error of a location estimate is a function of the magnitude of the expected LOD score.

Introduction

The identification of genetic loci that contribute to risk for complex medical and psychiatric disorders is a major goal of human geneticists (Collins 1995). To date, this goal has met with limited success. Efforts to identify such loci are likely hampered by factors such as incomplete penetrance, genetic heterogeneity, and the presence of phenocopies (Ott 1990). When positive results are found, they are often only weakly positive. Independent replication of these findings is crucial (Lander and Kruglyak 1995).

Recently, it has become clear that, in linkage studies of complex disorders, determining whether a given study has replicated an initial study's findings is not a trivial task. For instance, when two or more studies of a disease find suggestive evidence in a particular region of the genome, there may exist a large degree of variation in the specific position that gives maximum evidence for linkage (the location estimate). A priori, this variation might represent chance variation around a single genetic signal, the presence of multiple genetic signals, or one or more false-positive signals.

This issue may be illustrated by recent findings from linkage studies of schizophrenia. Figure 1 shows the location estimates from seven different investigations (Moises et al. 1995; Schwab et al. 1995; Straub et al. 1995; Arolt et al. 1996; Brzustowicz et al. 1997; Maziade et al. 1997; Riley and Williamson 1997). Distances shown are approximate. Two of these studies found evidence for linkage to schizophrenia-related phenotypes rather than to schizophrenia per se. One study (Arolt et al. 1996) found linkage for eye-tracking dysfunction, a putative schizophrenia endophenotype that has been shown to be associated with genetic liability to schizophrenia (Levy et al. 1993). Another study found linkage for severity of positive psychotic symptoms (Brzustowicz et al. 1997). The locations of the positive findings from the seven studies are widely scattered along chromosome 6p, covering a region of ~50–60 cM. Given the large degree of variation in position, could these findings have resulted from the same susceptibility locus? Can we consider the 6p linkage to schizophrenia a finding that has been "replicated" multiple times?

Similar findings have arisen in linkage studies of other complex traits. For example, two studies of type I diabetes (Field et al. 1996; Mein et al. 1998) report evidence for a susceptibility locus on chromosome 14q, but the location estimates are ~33 cM apart. Independent evidence for linkage of multiple sclerosis to chromosome 19 has been obtained by two groups (Sawcer et al. 1996; Kuokkanen et al. 1997), but the location estimates are

Received November 6, 1998; accepted for publication July 7, 1999; electronically published August 9, 1999.

Address for correspondence and reprints: Dr. Kenneth S. Kendler, Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, P.O. Box 980126, Richmond, VA 23298-0126. E-mail: kendler@hsc.vcu.edu

0002-9297/1999/6503-0032$02.00

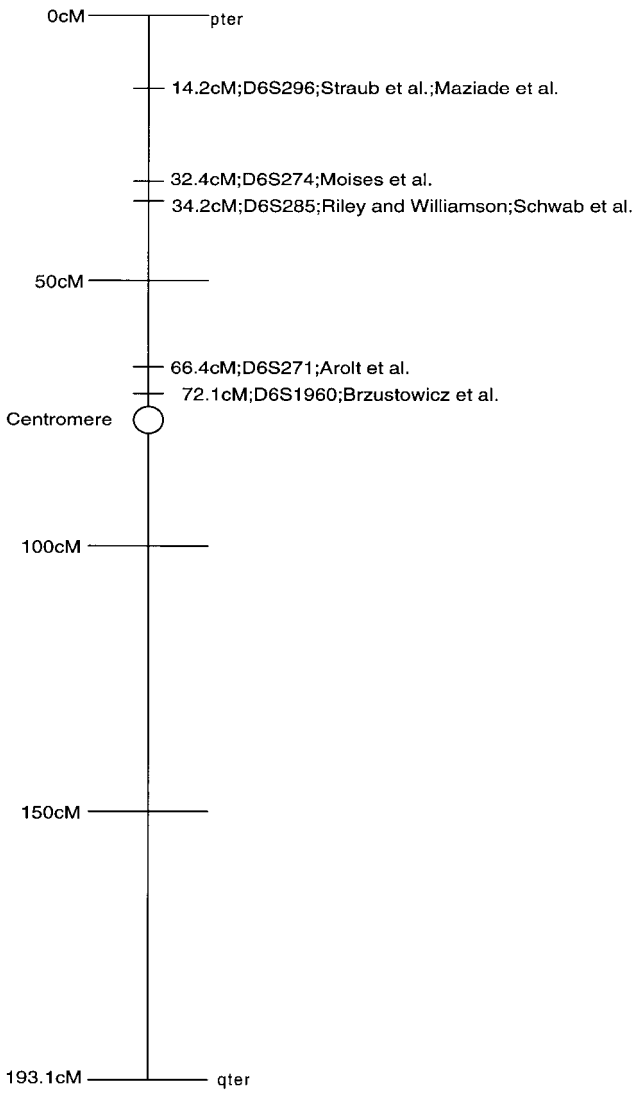

Figure 1 Positions that give maximum evidence for linkage along chromosome 6, in recent linkage studies of schizophrenia. The markers were placed relative to one another, on the basis of the Généthon map and the map of the Weber V6 screening set (Research Genetics) for marker D6S1960.

separated by ~7 cM. Linkage studies of psoriasis have yielded evidence for linkage to chromosome 4 (Matthews et al. 1996; Nair et al. 1997), with location estimates separated by 40 cM. Other examples of such variation in findings from linkage studies could be adduced.

In this report, we address the issue of variation in the position that gives maximum evidence for linkage—that is, the location estimate—in studies of complex traits. Using computer simulations, we have studied the distribution of the location estimate arising from repeated linkage studies of a single "disease." The linkage studies focus on a simulated chromosome that contains a susceptibility locus. The disease embodies characteristics such as incomplete penetrance of susceptible genotypes, heterogeneity, and phenocopies.

Methods

Three steps were repeatedly performed by computer: (1) simulation of a set of genotypes and derivation of phenotypes consistent with a "typical" complex trait, (2) analysis of the simulated data, and (3) recording of the location estimate (i.e., determination of the position of the peak multipoint LOD score). In this way, sampling distributions were constructed for location estimates.

The genotypes simulated were marker and disease genotypes, linked to each other in specific ways. Thus, the simulated data represented a "scan" of a small chromosome, containing a disease-susceptibility locus. Figure 2 contains a picture of the simulated chromosome, showing the relative positions of the markers and of the disease-susceptibility locus. The 13 markers, equally spaced at ~10.1-cM intervals, cover the length of the chromosome. The disease-susceptibility locus was placed between markers 5 and 6, closer to marker 5, ~41.4 cM from the left end of the chromosome.

Table 1 lists parameter values and other conditions used in our study. All markers had four alleles of equal frequency, corresponding to a PIC score of ~.7. Genotypes at two disease-susceptibility loci were simulated. One (locus A) corresponded to the (linked) locus shown on the framework of markers. The second disease-susceptibility locus (locus B) was unlinked to any other loci; this was introduced to mimic the effects of genetic heterogeneity. This concept of heterogeneity is based on the admixture model (Smith 1963). All pedigree structures were the same. Each family consisted of two parents and two offspring (both affected; see below), with genotype and phenotype information available for all four family members.

The only reason for our simulation of two disease-susceptibility loci was to create a system with genetic heterogeneity. We are interested in the precision of the location estimate for locus A, the locus on the simulated

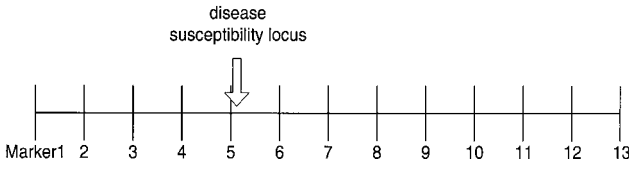

Figure 2 Structure of the simulated chromosome. All intermarker distances are ~10.1 cM, giving a total length of ~121.2 cM. The disease-susceptibility locus is 1 cM from marker 5, at ~41.4 cM. Note that marker 1 is at 0 cM, marker 2 is at 10.1 cM, etc.

Table 1

Parameters for Simulation and Analysis

Characteristic	Description
Marker loci	All have four alleles, each with frequency .25
Family structure	Nuclear family (parents and two children, both affected)
Locus A	Two alleles, disease-allele frequency .0033
Locus B	Two alleles; disease-allele frequency .0099
α	~.25
Disease prevalence	.03

chromosome. We analyzed the data by using a one-locus representation of the two-locus system, with a parameter, α, standing for the proportion of families linked to locus A. Evidence suggests that the power to detect linkage in a two-locus system analyzed as though it were a one-locus system is very close to the power when the true model is assumed, if the correct mode of inheritance is specified at the linked locus (Greenberg and Hodge 1989; Vieland et al. 1992; Durner et al. 1999). Moreover, the one-locus approach reflects the method by which most linkage analyses of complex traits are currently conducted. We set the penetrance vector for this one-locus treatment of our system at .8, .4, and .02. We also decided that 25% of the families (α = .25) containing a "genetic" form of the disease would be affected because of locus A—that is, the locus linked to the simulated chromosome. We then chose disease-allele frequencies for loci A and B and a penetrance matrix for the true two-locus system that were approximately consistent with both the aforementioned requirements for α and the penetrance vector for the one-locus treatment. The disease-allele frequencies chosen for locus A and locus B are shown in table 1. The penetrance matrix that we constructed is shown in table 2.

Genotypes were simulated for large numbers of individuals by the Genometric Analysis Simulation Program (GASP) (Wilson et al. 1996), which provides an environment for simulation of marker and disease-susceptibility genotypes under a variety of specified conditions. The output of GASP was modified by an SAS program (SAS Institute 1990). This SAS program used both the penetrance matrix of table 2 and the simulated genotypes at the disease-susceptibility loci to derive diagnoses for each individual. Only families with both children affected were kept for subsequent analysis. The final set of families, with phenotypes and genotypes at the marker loci, for all individuals, was saved to a file. The genetic-analysis program GENEHUNTER (Kruglyak et al. 1996) was then used to analyze the data.

Each set of families was analyzed by GENEHUNTER, with the penetrance vector for the one-locus treatment of our system, a disease-allele frequency of .0132 (the sum of the disease-allele frequencies at loci A and B), and the correct value of α (i.e., 25%). These parameters give the correct population prevalence, ~3%. Multipoint LOD scores were calculated every 2 cM along the simulated chromosome. The position that gave the highest, or peak, LOD score for each analysis step was recorded; this was the location estimate. The entire multipoint-LOD curve was also saved.

A single run thus comprised one round of simulation, appropriate modification of the simulated data, and analysis. We studied the distribution and variation of the location estimate across four different groups, each group consisting of 500 runs. The only difference across groups was the number of families retained for analysis per run. Groups 1, 2, 3, and 4 consisted of 200, 400, 800, and 1,600 nuclear families/run, respectively.

Results

Figure 3 shows the frequency distributions of the location of the peak LOD score obtained for each of the four groups. Visual inspection reveals a trend as the number of families per run increases. The frequency of observations in the tails of the distribution decreases (i.e., the height of the "flat" portion of the distribution decreases) while the frequency of observations in the central part of the distribution increases (i.e., the central part of the distribution becomes taller and somewhat more narrow).

We examined three ways of quantifying the amount of variation in the location estimate for these groups of simulations. Method 1 simply reports the observed standard errors (SEs) of the frequency distributions shown in figure 3. Table 3 contains these results. However, when this method is used, the results are strongly influenced by the tails of each distribution. As outlined in Appendix A, the SEs become partly a function of the length of the simulated chromosome. In brief, a longer simulated chromosome is equivalent to providing more chances for the multipoint LOD score to randomly reach a height that is greater than that achieved because of the signal from the disease locus. Location estimates from these random peaks in the multipoint-LOD curve may occur very far from the true disease locus and thus grossly inflate the calculated SE.

Table 2

Penetrance Matrix for Joint Genotypes at Disease-Susceptibility Loci

	Penetrance for Joint Genotype		
	AA	Aa	aa
BB	.8	.8	.8
Bb	.8	.8	.4
bb	.8	.4	.02

The second method (method 2) for quantifying the amount of variation in each group is to model each distribution in figure 3 as a mixture of two distributions: (1) a normal distribution centered on the true disease locus and (2) a rectangular, or flat, distribution over the entire chromosome. Then the SE of the normal component can be calculated. Appendix B explains this procedure. The results are shown in table 3.

A third method (method 3) for quantifying the variation in location estimates involves the use of the entire multipoint-LOD curve for each simulation run, not just the location of the maximum LOD score. We generated an "average," or expected, multipoint curve for each of the four simulation groups by plotting the mean multipoint LOD at each location along the chromosome. Quadratic functions were fitted to several points of expected LOD score around the maximum, one for each simulation group. As outlined Appendix C, the curvature (second derivative) of these functions at their maxima is related to the SE of the location estimate. The SEs for location that are derived in this way are given in table 3.

Inspection of table 3 reveals that, with 200, 400, or 800 families, method 1 produces much larger SEs than do methods 2 or 3. With 1,600 families, method 1 and method 2 produce similar SEs, with method 3 producing a smaller SE. With 200 families, estimates from methods 2 and 3 are similar, but thereafter method 3 produces smaller SEs.

Table 3

SEs from Three Methods of Quantifying the Variation in Location Estimates

No. of Families/Run	SE (95% CI[a]) (cM)		
	Method 1	Method 2	Method 3
200	25.76 (100.98)	7.45 (29.20)	7.92 (31.05)
400	18.51 (72.56)	6.45 (25.28)	5.60 (21.95)
800	9.17 (35.95)	5.50 (21.56)	3.84 (15.05)
1,600	4.05 (15.88)	4.18 (16.39)	2.73 (10.70)

[a] Calculated as 2(1.96 SE).

Discussion

We have studied the variation of location estimates in simulated linkage studies of a complex disorder. Our findings suggest that the variability in position is substantial for complex disorders, with 95% confidence in-

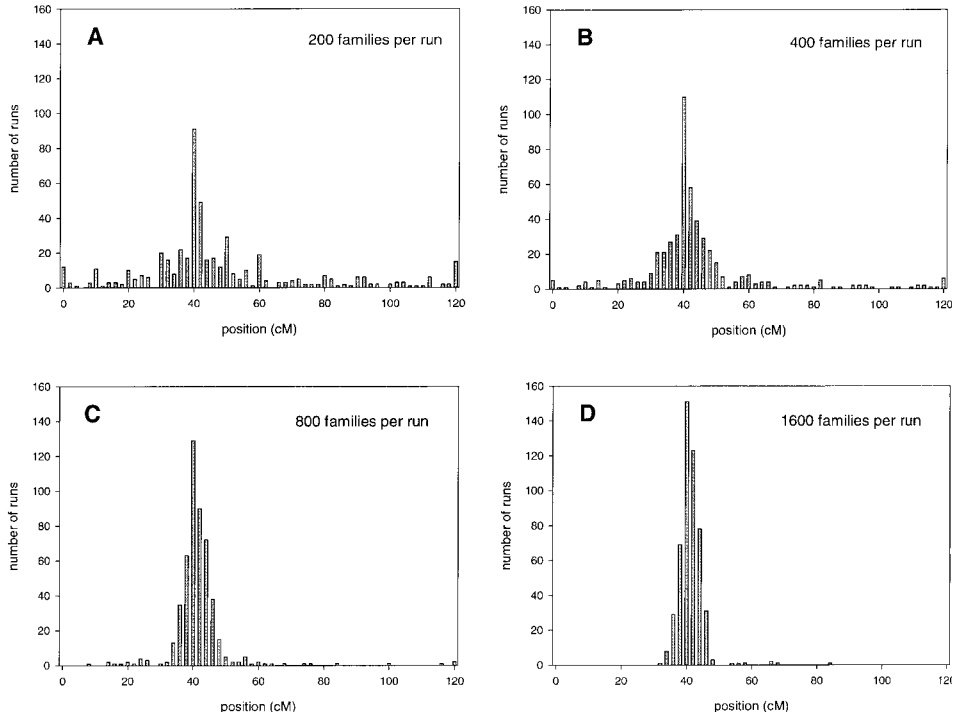

Figure 3 Frequency distributions of location estimates. The distribution of location estimates is shown for each of four groups: 200 families/run (*A*), 400 families/run (*B*), 800 families/run (*C*), and 1,600 families/run (*D*). The *X*-axis represents the position along the chromosome. The *Y*-axis represents the number of runs that gave location estimates at the corresponding position. The true position of the disease locus is 41.4 cM. Multipoint LOD scores were calculated every 2 cM.

tervals (95% CIs) covering 10s of cM in samples consisting of relatively large numbers of families. Most existing linkage studies use sample sizes less than that of our smallest group, 200 families. For instance, in the linkage studies of schizophrenia that were cited above, the largest sample size is 265 families—although many of these studies contain families that are larger than those simulated in our study. Our study suggests that, if the linkage "signal" from a complex-trait susceptibility locus is weak (because of incomplete penetrance, heterogeneity, etc.), the location estimate may be many centimorgans from the true disease locus. This emphasizes the need for other, complementary approaches—such as studies of linkage disequilibrium—to narrow the chromosomal regions of interest that have been identified in linkage analysis. On the other hand, it seems at least plausible that some of the findings that implicate chromosome 6p in linkage studies of schizophrenia may in fact be detecting the same susceptibility locus, despite the large variation in the location estimates observed. The degree of variation in location estimates is consistent with what one might expect for linkage studies of a weak susceptibility locus.

We have reported three different methods for quantifying the variation derived from our simulations. Comparison of the SEs shown in table 3 reveals that the methods give quite different results. Which is the best quantitative summary of the results? Method 1 (in which the SEs are from the "raw" distributions) includes the effect of random local maxima in the multipoint-LOD curve, which corresponds to the effect of false positives in actual linkage studies. This is always a reality faced by investigators in linkage analysis. On the other hand, the more interesting and relevant results are probably those which eliminate the effect of the length of the simulated chromosome. Investigators who wish to know the precision of a location estimate are implicitly assuming that the location estimate is for an actual disease-susceptibility locus, not for a local, random maximum (a false positive) in the multipoint-LOD curve. Because method 2 (mixture model) and model 3 (Fisher's expected information) reduce the influence of local maxima, they are probably better for quantifying the variation than is method 1. Methods 2 and 3 produce SEs that are closer to each other but that are still appreciably different for larger sample sizes. Estimates of SEs produced by method 2 are partially dependent on how the data are binned (i.e., 12 equal intervals were used to calculate the SEs in table 3; use of 24 or 6 intervals would give markedly different estimates for the SE [data not shown]). Moreover, the method of calculating the curvature of the average LOD curves has the advantage of incorporating the entire multipoint curve from each simulation. Thus, we believe that method 3 is probably the best method for quantifying the results of our simulations. It is important to note that SEs produced by this method represent the precision of location estimates after elimination of the effects of random, local maxima in the multipoint-LOD curve. These SEs may be quite different from those derived empirically from the simulated sampling distributions.

Our work confirms previous findings of large variation in location estimates from linkage studies of complex traits (Darvasi et al. 1993; Kruglyak and Lander 1995; Hauser and Boehnke 1997; Hovatta et al. 1998). In one simulation experiment (Hovatta et al. 1998), the mean deviation of location estimates from a true disease locus was anywhere from 6.8 to 9.5 cM, with 200 sib pairs and a relatively weak genetic signal. This would correspond to 95% CIs covering 10s of cM. In another study, investigators examined the variation in location estimates for genes mapped by use of experimental backcross populations (Darvasi et al. 1993). Even with 500 animals, CIs covered many 10s of cM for genes of moderate to small effect. In a third set of simulations, location estimates occurred as far as 20 cM from the true locus (Hauser and Boehnke 1997).

The fact that increasing the number of families leads to a more precise estimate of location and produces higher LOD scores is not surprising, since more families provide more genetic information about the disease locus in question. Changing other parameters that influence the amount of genetic information should have similar effects on variation in location estimates. We have conducted a series of auxiliary simulations to demonstrate this.

First, we simulated a set of groups in which the trait was increasingly penetrant. As expected, the more penetrant the trait, the smaller the SE of the location estimate. One completely recessive case was also simulated. The SE of the location estimate was smaller than that from the completely dominant case. Next, we simulated a set of groups in which we varied α from its initial value of .25 up to 1.0. The SEs for groups with higher values of α decline dramatically, to a much greater extent than did the SEs for increasingly penetrant disease loci. We also ran a simulation group with parameters set exactly as those for the group with $\alpha = 50\%$, except that marker spacing was every 5 cM. This was to examine the effects that a denser set of markers would have on the precision of the location estimate. The SE of the location estimate for this group was 3.30 cM (by method 3), comparable to the SE of 3.72 cM for the analogous group with markers every 10 cM. This is consistent with previous findings that suggest that decreasing intermarker spacing below the "resolving power" (i.e., 95% CI) does not greatly increase the precision of the location estimate (Darvasi et al. 1993). Finally, four groups of simulations using the nonparametric LOD score (Risch 1990; Kruglyak and Lander 1995) in the analysis step, instead of

the parametric LOD score, were run. When compared with the analogous groups analyzed by parametric LOD score, the nonparametric LOD–analyzed groups revealed both less evidence for linkage overall and larger SEs in the location estimates.

To summarize the results of the auxiliary simulations, we constructed a graph (fig. 4) of the expected LOD score of each group at the disease locus (as an indication of the amount of genetic information) versus the SE of the location estimate, derived by method 3 in Appendix C. As discussed above, changing the parameters so that genetic information is increased (i.e., so that there is a higher expected LOD score) results in a corresponding decrease in the SE of the location estimate.

Interestingly, each group of simulations seems to fall on the same curve. We found empirically that the curve given by

$$\text{SEL} = 7.0181(\text{expected LOD})^{-0.5881}, \qquad (1)$$

where SEL is the SE of the location estimate by method 3 and expected LOD is the expected LOD score at the disease locus for all runs in a group, fits the set of points well (fig. 4). The curve suggests that, given the presence of a disease-susceptibility locus, the expected LOD score of a study implies the associated SE of the location estimate. In other words, the genetic information implied by any combination of conditions (α, penetrances, number of families, etc.) and summarized by the expected LOD score at the disease locus across repeated studies is all that is needed to determine the variation in the location estimate. Darvasi and Soller (1997) also have derived a mathematical relationship between genetic information and variation in location estimates; however, their relationship uses the effect size of a quantitative-trait locus and sample size to derive the size of the 95% CI for a location estimate.

It might seem that relation (1) would provide a simple method for determining the CIs for published linkage studies; one could take the reported LOD score at a location estimate and simply read off the corresponding SE, using equation (1). However, this not the case; the reason is that the proposed method makes a critical, unwarranted assumption—namely, that the reported LOD score has the same value as the expected LOD score at the disease locus. In reality, the reported LOD score is only an imperfect estimate of the expected LOD score. We used our simulations to confirm this idea. First, we stratified each run on the basis of the magnitude of its maximum LOD score. For each of the strata, we calculated the SE of the location estimate. Although there definitely was a correlation between the magnitude of the observed LOD score and the SE of location estimate, the correlation was far from perfect. In fact, it was sufficiently imperfect that CI estimates based on this principle would be unreliable.

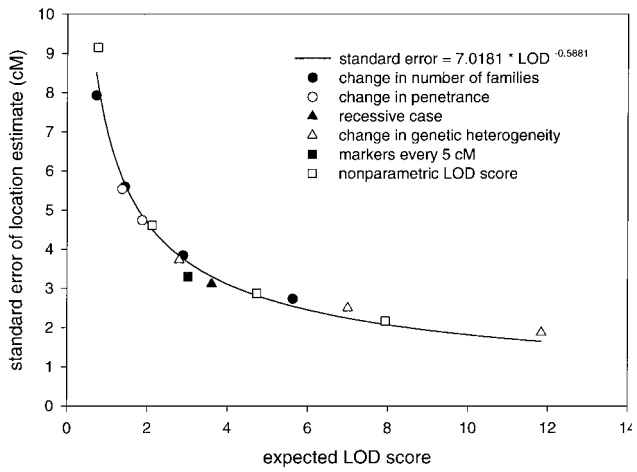

Figure 4 Relationship between the average or expected LOD score at the disease locus and the standard error of the location estimate. The blackened circles represent the four groups discussed in the text, each differing in the number of families per run. In another set of simulations, penetrance was varied; these groups are represented by unblackened circles. Here, each run had 200 families, $\alpha = .25$, and increasingly penetrant disease-susceptibility loci (up through complete penetrance). Standard error falls as the average LOD score increases, reflecting the increasing penetrance across groups. One completely recessive case, with no phenocopies and gene frequencies of .087 (locus A) and .15 (locus B), was run; this group is represented by a blackened triangle. A total of 200 runs were performed for these and all subsequent auxiliary simulation groups. The unblackened triangles represent a series of groups in which we varied α. These groups each had 200 families/run, the same penetrances used in the original simulations, and α values of .5, .75, and 1.0. As before, the standard error of the location estimate falls as the expected LOD score increases (reflecting the increasing values of α). The group represented by a blackened square was run with use of the same parameters as were used for the group with $\alpha = .5$, except that marker spacing was every 5 cM (instead of every 10 cM, as with all other groups). Four groups were run with use of the nonparametric LOD score to detect linkage, instead of the parametric LOD score (represented by unblackened squares). These groups had $\alpha = .25$, .50, .75, and 1.0; otherwise, parameters were the same as those for the group with 200 families/run. Finally, the curve given by $\text{SEL} = 7.0181 \times (\text{expected LOD})^{-0.5881}$ is shown, where SEL is the standard error of the location estimate and expected LOD is the expected LOD score at the disease locus. This curve was derived empirically and seems to fit the set of points well (97% of the variance in the data points is accounted for by this curve).

Although not useful for evaluating the results of published linkage studies, equation (1) still could be useful to investigators who wish to know the precision of location estimates for genes of a certain hypothesized effect. One could use the known gene effect to calculate an expected LOD score for a given sample; the approximate SE of the location estimate could then be derived by means of equation (1).

Acknowledgments

This work is supported by National Institutes of Health grant MH-41953. K.S.K. is the recipient of National Institute of Mental Health Research Scientist award MH-01277. The authors would like to thank Richard Straub for helpful comments.

Appendix A

Why Standard Errors from Method 1 Are Partly a Function of Chromosome Length

Inspection of figure 3 reveals that, for the groups with fewer numbers of families per run, a significant portion of the observations occur in the tails of the distributions, often very distant from the true locus. Many of these observations represent local maxima in the multipoint curve that happen to be higher than the peak due to signal from the disease locus. This kind of occurrence is especially common when there is little information for linkage analysis (because, say, relatively few families are available), resulting in a low signal-to-noise ratio. These local maxima occur randomly—that is, they are not related to the location of the disease-susceptibility locus.

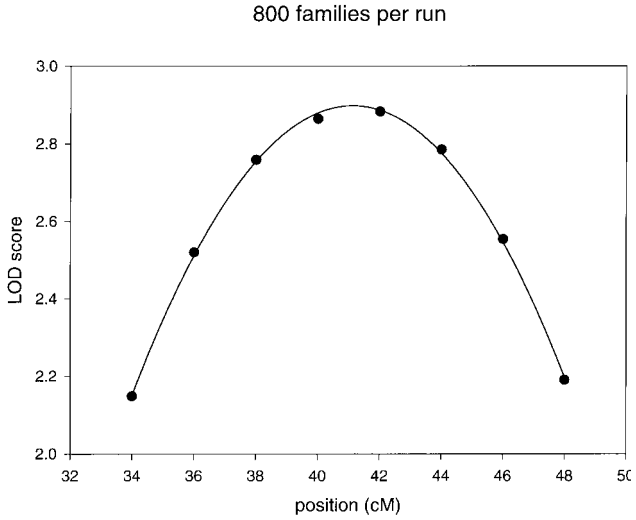

Figure A1 Plot showing the average or expected LOD score at several locations around the maximum for all 500 runs of the simulation group with 800 families/run. Also shown is the best-fitting quadratic function. The second derivative of this curve (the curvature) at its maximum is related to the standard error of the location estimate (for details, see the text). In this case, the best-fitting function was LOD = $-22.05 + (1.21 \text{ position}) - (0.015 \text{ position}^2)$, where LOD is expected LOD score and position is the position (in cM) along the simulated chromosome.

Extending the length of the chromosome would be equivalent to providing more chances for the multipoint curve to randomly achieve a local maximum that is greater than the peak due to the signal from the disease locus. For longer chromosomes, more runs that would have given location estimates related to the disease locus will give location estimates unrelated to the disease locus. Thus, the SE of the location estimate by this method is partly a function of the length of the chromosome.

Appendix B

Method 2 for Quantifying the Variation in Location Estimates

First, we divided the distributions of location estimates into 12 equal intervals (~10.1 cM each), covering the entire chromosome. This was done to eliminate the effects of information content due to marker position. Next, we fitted the following parametric form to the data points: $f = y_0 + \kappa \times \exp\{-0.5[(x - x_0)/b]^2\}$, where y_0 is the height of the rectangular distribution, κ is a nuisance parameter, and x_0 and b are, respectively, the mean and SE of the normal distribution. Parameter estimates were obtained by use of the Marquardt-Levenberg algorithm (Press et al. 1986). The resulting function explained 97% of the variance in the data points.

Appendix C

Method 3 for Quantifying the Variation in Location Estimates

We first calculated the average LOD score at each position along the chromosome, for all 500 runs in a simulation group. Eight of the resulting data points around the location of the disease locus are shown in figure A1, for the simulation group with 800 families/run. As one moves down the chromosome, the average LOD score rises to a maximum over the true disease locus and then declines on the other side. Next we fitted the following quadratic form to a set of points around the disease locus, for each simulation group in turn, by nonlinear regression: LOD = $y_0 + (a \times \text{position}) + (b \times \text{position}^2)$, where LOD is the LOD score, position is position along the simulated chromosome, and y_0, a, and b are parameters to be estimated. Because location estimates are maximum-likelihood estimates, the LOD score around the maximum is expected to have a quadratic form. Parameters for the curves were obtained by use of the Marquardt-Levenberg algorithm (Press et al. 1986). The data points and the resultant curve for the group with 800 families/run are shown in figure A1. The

curves fitted each set of points well (~99% of variance is explained) and seemed to faithfully capture the behavior of the points around the maximum LOD score. Then we took the second derivative of the resultant curve, with respect to position. Inserting the position of the maximum LOD score into the second derivative and multiplying by -1 one gives Fisher's expected information at the true disease-locus position (Kendall and Stuart 1979). Taking the square root of the reciprocal of Fisher's expected information gives the SE of the location estimate.

References

Arolt V, Lencer R, Nolte A, Muller-Myhsok B, Purmann S, Schurmann M, Leutelt J, et al (1996) Eye tracking dysfunction is a putative phenotypic susceptibility marker of schizophrenia and maps to a locus on chromosome 6p in families with multiple occurrence of the disease. Am J Med Genet 67:564–579

Brzustowicz LM, Honer WG, Chow EWC, Hogan J, Hodgkinson K, Bassett AS (1997) Use of a quantitative trait to map a locus associated with severity of positive symptoms in familial schizophrenia to chromosome 6p. Am J Hum Genet 61:1388–1396

Collins FS (1995) Positional cloning moves from perditional to traditional. Nat Genet 9:347–350

Darvasi A, Soller M (1997) A simple method to calculate resolving power and confidence interval of QTL map location. Behav Genet 27:125–132

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943–951

Durner M, Vieland VJ, Greenberg DA (1999) Further evidence for the increased power of LOD scores compared with nonparametric methods. Am J Hum Genet 64:281–289

Field LL, Tobias R, Thomson G, Plon S (1996) Susceptibility to insulin-dependent diabetes mellitus maps to a locus (IDDM11) on human chromosome 14q24.3-q31. Genomics 33:1–8

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

Hauser ER, Boehnke M (1997) Confirmation of linkage results in affected-sib-pair linkage analysis for complex genetic traits. Am J Hum Genet Suppl 61:A278

Hovatta I, Lichtermann D, Juvonen H, Suvisaari J, Terwilliger JD, Arajarvi R, Kokko-Sahin ML, et al (1998) Linkage analysis of putative schizophrenia gene candidate regions on chromosomes 3p, 5q, 6p, 8p, 20p and 22q in a population-based sampled Finnish family set. Mol Psychiatry 3:452–457

Kendall M, Stuart A (1979) The advanced theory of statistics. Macmillan Publishing, New York

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

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454

Kuokkanen S, Gschwend M, Rioux JD, Daly MJ, Terwilliger JD, Tienari PJ, Wikström J, et al (1997) Genomewide scan of multiple sclerosis in Finnish multiplex families. Am J Hum Genet 61:1379–1387

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

Levy DL, Holzman PS, Matthysse S, Mendell NR (1993) Eye tracking dysfunction and schizophrenia: a critical perspective. Schizophr Bull 19:461–536 (erratum: Schizophr Bull 19:685 [1993])

Matthews D, Fry L, Powles A, Weber J, McCarthy M, Fisher E, Davies K, et al (1996) Evidence that a locus for familial psoriasis maps to chromosome 4q. Nat Genet 14:231–233

Maziade M, Bissonnette L, Rouillard E, Martinez M, Turgeon M, Charron L, Pouliot V, et al (1997) 6p24-22 region and major psychoses in the eastern Quebec population: Le Groupe IREP. Am J Med Genet 74:311–318

Mein CA, Esposito L, Dunn MG, Johnson GC, Timms AE, Goy JV, Smith AN, et al (1998) A search for type 1 diabetes susceptibility genes in families from the United Kingdom. Nat Genet 19:297–300

Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F, Arolt V, et al (1995) An international two-stage genome-wide search for schizophrenia susceptibility genes. Nat Genet 11:321–324

Nair RP, Henseler T, Jenisch S, Stuart P, Bichakjian CK, Lenk W, Westphal E, et al (1997) Evidence for two psoriasis susceptibility loci (HLA and 17q) and two novel candidate regions (16q and 20p) by genome-wide scan. Hum Mol Genet 6:1349–1356

Ott J (1990) Cutting a Gordian knot in the linkage analysis of complex human traits. Am J Hum Genet 46:219–221

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical recipes: the art of scientific computing. Cambridge University Press, New York

Riley BP, Williamson R (1997) Non-parametric analysis of chromosome 6p24-22 marker data and schizophrenia in southern African Bantu-speaking families. Psychiatr Genet 7:131–132

Risch N (1990) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. Am J Hum Genet 46:242–253 (erratum: Am J Hum Genet 51:673–675 [1992])

SAS Institute (1990) SAS II: SAS/STAT user's guide, version 6, 4th ed. SAS Institute, Cary, NC

Sawcer S, Jones HB, Feakes R, Gray J, Smaldon N, Chataway J, Robertson N, et al (1996) A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. Nat Genet 13:464–468

Schwab SG, Albus M, Hallmayer J, Honig S, Borrmann M, Lichtermann D, Ebstein RP, et al (1995) Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. Nat Genet 11: 325–327

Smith CA (1963) Testing for heterogeneity of recombination fraction values in human genetics. Ann Hum Genet 27: 175–182

Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, et al (1995) A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. Nat Genet 11:287–293

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

Wilson AF, Bailey-Wilson JE, Pugh EW, Sorant AJM (1996) The genometric analysis simulation program (GASP): a software tool for testing and investigating methods in statistical genetics. Am J Hum Genet Suppl 59:A193