Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results

Eric Lander[1,2] & Leonid Kruglyak[1]

Genetic studies are under way for many complex traits, spurred by the recent feasibility of whole genome scans. Clear guidelines for the interpretation of linkage results are needed to avoid a flood of false positive claims. At the same time, an overly cautious approach runs the risk of causing true hints of linkage to be missed. We address this problem by proposing specific standards designed to maintain rigor while also promoting communication.

[1]*Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge Massachusetts 02142, USA*
[2]*Department of Biology, Massachusetts Institute of Technology, Cambridge Massachusetts 02139, USA*

Genetic dissection of complex traits is becoming central to mammalian genetic analysis. In the fifteen years since it was recognized that genetic inheritance can be traced with naturally occurring DNA sequence variation [1], the identification of genes responsible for simple mendelian traits has become a straightforward, if still demanding, task. Over 500 such genes have been mapped to specific chromosomal regions in the human and more than 60 have been cloned based on their position. These breakthroughs are steadily reshaping biological and medical thinking. Yet, many of the most important medical conditions-including heart disease, hypertension, diabetes, asthma, schizophrenia, and manic depression —show much murkier inheritance patterns. The geneticists' challenge is now to tease apart the multifactorial causes of these diseases.

In principle, the solution is clear. Genetic mapping of any trait —simple or complex —boils down to finding those chromosomal regions that tend to be shared among affected relatives and tend to differ between affecteds and unaffecteds. Conceptually, this amounts to a three-step recipe: scan the entire genome with a dense collection of genetic markers; calculate an appropriate linkage statistic S(x) at each position x along the genome; and identify the regions in which the statistic S shows a significant deviation from what would be expected under independent assortment.

Yet, these deceptively simple instructions conceal a thorny question: since the statistic S(x) fluctuates substantially just by chance across an entire genome scan, what constitutes a 'significant' deviation? What standard should be required for declaring linkage?

Although biologists often greet statistical issues with glazed-eyed indifference, we believe that the resolution of this particular question has important consequences for the future of our field. To reach our goal, geneticists must chart a prudent course between Scylla and Charybdis.

Adopting too lax a standard guarantees a burgeoning literature of false positive linkage claims, each with its own gene symbol *(ASTH56, ASTH57,* ...).Scientific disciplines erode their credibility when a substantial proportion of claims cannot be replicated —even more so when the claims reach not only the professional journals but also the evening news. Psychiatric genetics provides a cautionary tale, in which a spate of non-replicable fmdings in the mid-1980s undermined support for such studies[2–7]. It is thus essential that there be a sufficiently stringent standard that linkage is claimed only when there is a high likelihood that the assertion will stand the test of time.

On the other hand, adopting too high a hurdle for reporting results runs the risk that the nascent field will be stillborn. Initial genetic analyses may fall short of the strict threshold for statistical significance, but may nonetheless point to important regions deserving intensive investigation. Without channels by which investigators can report such tentative hints of linkage, the discovery of disease genes may be delayed in an overzealous attempt to avoid all error.

Striking the right balance requires both a mathematical understanding of how often positive results will occur just by chance and a value judgment about the relative costs of false positives and false negatives. Our goal here is to provide an accessible treatment of the first subject and to offer a concrete proposal regarding the second.

Statisticalsignificance in genome-wide scans

In searching for disease genes, it is important to distinguish between pointwise significance levels and genome-wide significance levels. The pointwise (also called nominal) significance level is the probability that one would encounter such an extreme deviation *at a specific locus* just by chance. The genome-wide significance level is the probability that one would encounter a deviation *somewhere* in a whole genome scan. The former concerns a single test of the null hypothesis of no linkage; the latter involves fishing over a large number of tests to find the most significant result.

Consider the following idealized sib pair study. An investigator collects *n* pairs of affected sibs, genotypes them using a perfect genetic map that is fully informative at every point in the genome, and calculates the average proportion 1t(x) of alleles shared identical-by-descent at each location x in the genome. Geneticists traditionally report the results at each location in one of three essentially equivalent ways: a Z-score, a lod score, or a *P* value. The Z-score is the number of stan-

Box 1 Going to extremes

How often will a linkage statistic S(x) exceed a specified threshold T by chance In a whole genome scan? The mathematical theory of large deviations provides an answer, which is applicable to a wide variety of experimental designs and statistics. The answer is given by a simple formula involving four quantities: the pointwise significance level of T; the size of the genome; the rate of fluctuation of the statistic; and the threshold T itself.

The result is simplest to state for a normally distributed statistic, a Z-score. The number of regions in which the statistic Z exceeds a relatively large level T in a whole genome scan has a Poisson distribution with mean:

$$I.I(T) = [C + 2pGT2)\, a(T) \qquad (1)$$

and the genome-wide significance level of exceeding T is a'(T) = 1--i!"'I'(I) (which is = I.I(T), when this quantity is small)8.9.35. The quanti1ies in the equation ara defined as follows: (i) the expression a(T) denotes the pointwise significance level of exceeding level T; 0i Ilthe constants C and G denote, respectively, the number of chromosomes and the genome length measured in Morgans (C = 23 and G = 33 for the human); (iii) the constant p measures how rapidly the statistic S(x) fluctuates, which reflects the total crossing over rate between the genotypes being comparad.+ The threshold for suggestive linkage is found by solving I.I(T) = 1 and for significant linkage by solving I.I(T) = 0.05. It is worth noting that the factor [C + 2pGT2)-the equivalent of the standard Bonferroni correction for multiple testing — measures the effective number of tests carried out in seerching the entire genome.

The result can also be applied to lod scores, since X= (21og 10) lod follows a chi-squared distribution (with the number of degrees of freedom depending on the number of additional parameters maximized in rhe hypothesis to be tested as opposed to in the null hypothesis)* . The pointwise significance level is thus determined from the appropriate chi-squared table, and the factor [C + 2pGF) is replaced by [C + 2pGX).

Table 1 shows the resulting thresholds for a variety of situations, including the following:

• Allele-sharing methods in affected relative pairs. For straightforward affected relative pair analysis, the appropriate value of p is given in Table 1 for studies involving a fixed type of relative pair36. For studies involving a mixture of relative types, one should use the weighted average of p for the different relative types (or simply note that the thresholds for typical relative pairs are all roughly in the range of 1o-'3-5 x 1Q-4 for suggestive linkages and 5 x 1o-s–1!.J"i for significant linkages).

In principle, the same thresholds apply to the more complex APM analysis, which similarly involves a normally distributed statistic based on pairwise comparisons among a mixture of affected relatives. In practice, some APM analyses involve a relatively small number of families and so caution is required in applying thresholds based on asymptotic assumptions of large sample size.

• QTL mapping in experimental crosses. The lod score involves 1 d.f. in the case of a backcross or an intercross in which a single parameter is esti-

mated (purely recessive, dominant or additive model) and 2 d.f. in the case of an intercross in which two parameters are estimated (both additive and dominance components). The appropriate values of p are given in Table 1.

* In mathematiCal terms, p Is related to the autocorrelation functoon of S(x): p = -C'(0)/2, wheta C'(0) is the derivative at 0 (taken as the limfrom above) of the autocorrelation functoon C(x) = E[S(O)S(x)VE[S^2(0)}.

'In the case of human linkage analysis X Is asymptotlcally distributed as a 1/2:112 mixture of a chl-squated and a point masa at zetO. This Is due to the one-sided nature of the test, the!ls, the slgnificance level determined from a chl-squate<l table has to be dlvkled by 2. In QTL mapping In experimental croee., the teat Is usually two-sided since loci from either strain can affect the phenotype.

dard deviations by which 7t(x) exceeds its null expectation of 0.50; it follows a normal distribution when *n* is large. The lod score (or MLS-Maximum Lod Score, the lod score maximized over a set of parameters) is the log-likelihood ratio of the data under the hypothesis that the allele sharing proportion has the observed value 7t(x) as compared to the hypothesis that there is no excess sharing; the distribution of this statistic is related to a chi-squared distribution when *n* is large. The *P* value reflects the pointwise chance of observing a deviation as high as 7t(x) under independent assortment.

Suppose, for example, that a study of 100 sib pairs reveals an allele sharing proportion of 61% somewhere in the genome. This result corresponds to a Z-score of 3.1, a lod score of2.1, and a nominal Pvalue ofO.OOl. Should one be impressed by this finding? It clearly depends on how often such deviations would arise by chance in a whole genome search.

The mathematical theory of large deviations holds the answer, as was pointed out a few years ago[8,9]. The expected number of chromosomal regions in which a linkage statistic exceeds a threshold Tis given by a simple formula I.I.(T), explained in Box 1. In fact, the number of such regions approximately follows a Poisson distribution with mean I.I.(T), and the chance of finding at least one such region is thus 1-e-ll(TJ = I.I.(T) when I.I.(T) is small. The approximation becomes asymptotically exact when the number of sib pairs is large and the threshold T is high. In fact, it is accurate enough for practical purposes provided that *n* is at least 50. (It is worth noting that, while lod score analysis of a small number oflarge families can be quite sensitive to changes in a few key data points, non-parametric statistics based on a large number of small families tend quickly to normal distributions and tend to be robust.)

Fig. 1 shows the results in graphical form. Focusing on *P* values, we expect regions significant at $P = 0.05$ to occur about two dozen times by chance (that is, at least once on most chromosomes); $P = 0.01$ about 7-8 times; $P = 0.001$ slightly more than once; $P = 0.0001$ about 0.2 times; and $P = 0.00002$ about 0.05 times. In other words, there is a 5% chance of randomly finding a region with a *P* value as extreme as 2×10^{-5}. To keep the chance of encountering a false positive at no more than 5%, one must therefore impose a threshold of Z 4.1, lod3.6 or $P = 2 \times 10^{-5}$. With any less stringent

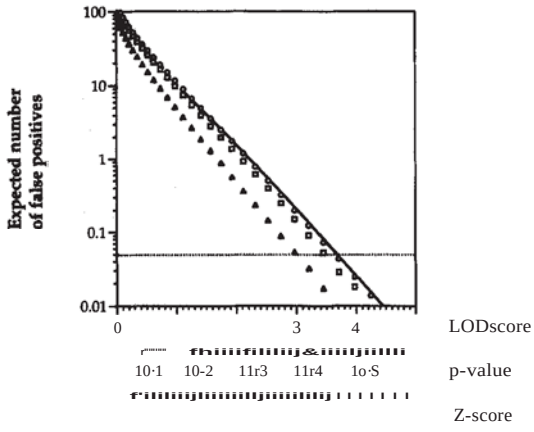

Fig. 1 Number of false positives expected in a whole genome scan for a given threshold of lod score, Z score or pointwise *P* value. Solid line represents asymptotic expectation for a perfect genetic map, based on the theory described in the Box 1. Symbols represent results for 100 sib pairs obtained from 100,000 simulations using genetic maps. with markers spaced every 0.1 eM (circles), every 1 eM (squares), and every 10 eM (triangles). The genome is assumed to consist of 23 chromosomes, with total length 3450 eM. Note the close correspondence between the asymptotic theory ana the 0.1 eM simulation. The dotted line Indicates the 5% genome-wide significance level.

Box 2 A simulated genome scan

To illustrate the random fluctuations expected in a whole-genome scan, we generated simulated genotypes assuming independent assortment throughout the genome — that is, that there are no trait-causing loci. All positive scores in such data necessarily represent random fluctuations, *not* true linkages. Fig. 2 shows the result of a typical (unselected) simulated genome scan, In which 100 sib pairs and parents were 'genotyped' for markers having heterozygosity of 0.8 and average spacing of 3 eM. A total of 22 regions reached the nominal significance of 0.05, while 6 reached nominal significance of 0.005; these numbers can be compared with dense-map expectations of 31 and 7, respectively. A single region on chromosome 14 reached the status of suggestive linkage, as expected, while no region showed significant linkage. If these results had occul1'8d In a real dataset, an investigator would likely call attention to the possibility of linked genes on chromosome 14 and to the p..-ce of a peak near the HLA region of chromosome 6p. The example thus illustrates that false positives can cluster, occur in candidate regions, and otherwise mimic true loci.

Some investigators have advocated pursuing all regions with nominal significance of $P = 0.05$, by attempting to 'replicate' the significance level in a second data set. To illustrate the perils of this approach, we applied it by generating a second set of random, simulated data. Of the 22 regions with $P = 0.05$ in the first set, four regions produced similar significance levels in the second data set as well. Fortuitously, these included the small peak near HLA and the larger peak on chromosome 14. The example illustrates that when low significance thresholds are used, many regions will be followed up and spurious 'replication' can occur by chance.

Applying the standards proposed here avoids these problems. No significant linkages are found in the first, the second, or the combined dataset.

threshold, there is a substantial chance (> 5%) of reporting false linkages. To illustrate the point, we describe a simulated whole-genome scan in Box 2.

The standard may seem harsh at first glance, but it accords well with historical practice. The traditional threshold oflod 3 for classical two-point linkage studies of simple mendelian traits corresponds to an asymptotic pointwise significance level of $P = 10\text{-}4$ (ref. 10), for a genome-wide significance level of about 9%. In fact, the lod score threshold need only to be raised to 3.3, corresponding to $P = 5 \times W\text{-}^{5}$, to achieve the recommended genome-wide significance level of 5%.

It is worth noting that there is a widespread misconception in human genetics that a lod score of 3 is equivalent to a significance level of only $P = 10\text{-}^{3}$. This error is root-edina confusion about the meaning of lod scores and P values. Lod scores concern the *ratio* of two probabilities, while P values refer to a single absolute probability. Specifically, lod = 3 means that the observed data is 10^{3}-fold more likely to arise under a specified hypothesis of linkage than under the null hypothesis of independent assortment. By contrast, $P = W\text{-}^{3}$ means that the probability of encountering as large a lod score as observed is IQ-3 under the null hypothesis. One can convert a lod score to a chi-squared statistic by multiplying by 2(log,10) = 4.6 and then use standard statistical tables, taking into account the one-sided nature of the test, to confirm that a lod score of 2.1 corresponds to $P = 10\text{-}^{3}$, while a lod score of 3.0 corresponds to the more extreme $P = 10\text{-}4$.

Are whole-genome thresholds overty stringent?

Some geneticists might object to imposing such a stringent standard for declaring linkage. Certain arguments have been advanced in the hopes of gaining special dispensation. It is worth considering them in turn.

• *"My study only looked at a few markers (or a few chromosomal regions), so it's not fair to impose a threshold based on a whole genome search."* The extreme example of this argument would be a geneticist who finds a weakly positive score with the first marker and seeks to employ the pointwise significance level-asserting that only- a single hypothesis has actually been tested. The fallacy is that the investigator would not have abandoned the search if the first marker had been negative, but would have persevered until a positive result was obtained or the entire genome was examined. Having assembled a large patient collection, the geneticist is committed to a whole genome search. It makes no sense to employ a different threshold depending on whether the inevitable false positive fluctuations happen to occur earlier rather than later in the search.

• *"My study only involved a genome scan with markers every 10 eM, so it's not fair to impose a threshold based on an infinitely dense genetic map."* Again, the analysis does not stop with the sparse map. Initial hints oflinkage with a single marker are immediately pursued by using multipoint methods and by peppering the region with a dense collection of markers. In any region that matters, geneticists rapidly extract the complete inheritance information — with the explicit hope of increasing the linkage score.

A hierarchical search — in which one performs a genome scan with a sparse map and then follows up 'interesting' regions with a denser map-is an efficient study design[11, 12], but the resulting false posi-

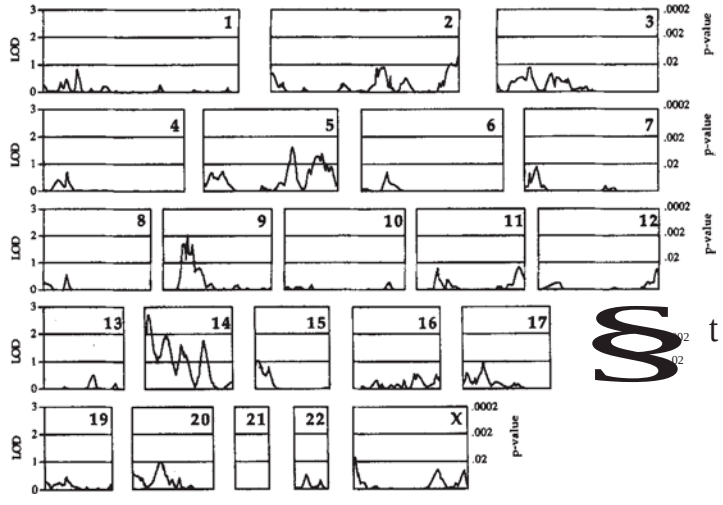

Fig. 2 Simulated genome scan with no trait loci segregating. Chromosomal size is proportional to genetic length, taken from ref. 33. Multipoint lod scores were computed as described34.

tive rate is essentially the same as if a dense map had been used throughout the genome (D. Siegmund, personal communication). This is because the false positives are almost invariably included among the regions chosen for follow-up.

The dense-map threshold turns out not to be that draconian. If one performed only single point analysis with an evenly spaced map, the thresholds for a genome-wide significance level of 5% would not be dramatically different: the lod score thresholds would decrease by ~20% for a 10 cM map; ~15% for a 5 cM map; ~10% for a 2 cM map; and ~7% for a 1 cM map (Fig. 1 shows first and last cases). Moreover, these thresholds would be appropriate only if one did not use multipoint analysis or a denser map to obtain more information. To our thinking, it is better to extract the full inheritance information and find the best *P* value.

In the modern world, it is fair to assume that highly motivated investigators squeeze as much information as possible from the available family material and their results should thus be measured against the corresponding threshold for a dense genome scan. (Some backsliding might be countenanced if strong prior evidence exists to restrict the search to a region; possible cases include a true single-point test of a highly relevant candidate gene, a test of the HLA region for an autoimmune disease, and an X-chromosome scan for a trait with convincing prior evidence of sex linkage.)

Notwithstanding our desire to avoid spurious linkages, we must always remember that regions that fall short of statistical significance may nonetheless be correct. Unfortunately, there is no way to distinguish between small peaks that represent weak true positives and peaks of the same height arising from random fluctuations, assuming that all inheritance information has been extracted. It would be irresponsible to consign such potentially valuable hints to the dustbin of laboratory history. What then is to be done?

Proposed standards

Back in the days when linkage studies of even the simplest trait required heroic efforts and good fortune, the human genetics community adopted standards to promote both rigor and communication. Lod scores of 3 were required to declare linkage in official chromosome committee reports, but weaker evidence could still be shared in more informal vehicles such as the McKusick *Newsletter*, a predecessor to the modern *Mendelian Inheritance in Man*[13].

Clear thinking about complex traits would be served by reviving such an approach. Specifically, we propose the following classification based on the number of times that one would expect to see a result at random in a dense, complete genome scan:

- *Suggestive linkage*-statistical evidence that would be expected to occur one time at random in a genome scan.
- *Significant linkage* —statistical evidence expected to occur 0.05 times in a genome scan (that is, with probability 5%).
- *Highly significant linkage*-statistical evidence expected to occur 0.001 times in a genome scan.
- *Confirmed linkage*-significant linkage from one or a combination of initial studies that has subsequently been confirmed in a further sample, preferably by an independent group of investigators. For confirmation, a nominal *P* value of 0.01 should be required (see below);

In the case of sib pair studies, the first three categories would correspond to pointwise significance levels of 7×10^{-4}, 2×10^{-5}, and 3×10^{-7} and lod scores of 2.2, 3.6, and 5.4. The corresponding *P* values for other study designs differ somewhat (Box 1, Table 1).

Suggestive linkage results will often be wrong, but they are worth reporting —if accompanied by an appropriate warning label about their tenuous nature. Investigators concerned about coming up empty-handed in a genome scan can take comfort from the fact that they can expect, by definition, to find about one suggestive linkage for every trait studied. On the other hand, journal editors must weigh how much attention to accord such results. At the least, specialty journals should actively support the reporting of suggestive linkages in some format. Indeed, it is worth reporting all regions with a nominal *P* value of $P = 0.05$ encountered in a *complete* genome scan, but without any claims of linkage.

Because suggestive linkages are so speculative, they should not be assigned gene names lest medical genetics be overrun with illusory loci. Geneticists should enter into a non-proliferation pact, under which genes symbols are reserved for significant linkages. Indeed, traditional usage has been to assign gene names only to confirmed linkages. The appropriate nomenclature committees should take up this issue and develop specific guidelines.

It is worth pointing out that even significant linkages will turn out to be false positives 5% of the time, that is, once in 20 genome scans. Because of the bias that only positive results tend to get reported, the observed false positive rate will be

Table 1 Thresholds for mapping loci underlying complex traits

Mapping method	crossover rate p	suggestive linkage P value (lod)	significant linkage *P* value (lod)
lod score analysis in human	1	1.7×10^{-3} (1.9)	4.9×10^{-5} (3.3)
Allele-sharing methods in human			
sibs and half-sibs	2	7.4×10^{-4} (2.2)	2.2×10^{-5} (3.6)
grandparent-grandchild	1	1.7×10^{-3} (1.9)	4.9×10^{-5} (3.3)
uncle-nephew	5/2	5.6×10^{-4} (2.3)	1.8×10^{-5} (3.7)
first cousin	8/3	5.2×10^{-4} (2.3)	1.6×10^{-5} (3.7)
first cousin, once removed	20/7	4.8×10^{-4} (2.4)	1.5×10^{-5} (3.8)
second cousin	16/5	4.2×10^{-4} (2.4)	1.3×10^{-5} (3.8)
QTL mapping in mouse or rat			
Backcross (1 d.f.)	1	3.4×10^{-3} (1.9)	1.0×10^{-4} (3.3)
Intercross (1 d.f., additive)	1	3.4×10^{-3} (1.9)	1.0×10^{-4} (3.3)
Intercross (1 d.f., recessive)	4/3	2.4×10^{-3} (2.0)	7.2×10^{-4} (3.4)
Intercross (1 d.f., dominant)	4/3	2.4×10^{-3} (2.0)	7.2×10^{-4} (3.4)
Intercross (2 d.f.)	1.5	1.6×10^{-3} (2.8)	5.2×10^{-5} (4.3)

Sib pair analysis involves no dominance component, and thus each sib pair is equivalent to two half-sib pairs. Lod score thresholds for the possible triangle method for sib pairs[37] may be computed by similar methods (D. Siegmund, personal communication); these thresholds are 2.6 for suggestive and 4.0 for significant linkage. Genome size Is assumed to be 3300 cM for the human and 1600 cM for the mouse and the rat. A typographical error appeared in the table of ref.38, which listed the significant *P* value for half-sib and sib pairs as 3×10^{-5}. The correct *P* value Is 2.2×10^{-5}, as shown above.

higher in the published literature. While individual investigators cannot do anything about this problem, it offers an additional rationale for conservative standards.

Thomson [14] recently proposed criteria for *putative* linkage that turn out to be essentially equivalent to our standard for *suggestive* linkage. Unfortunately, these criteria have been widely misinterpreted as implying genome-wide significance, despite Thomson's dear statement to the contrary. In ,fact, Thomson (pers. comm.) endorses the standards proposed above.

Replications and extensions

Linkage results must be replicated to be credible. We suggest that the term "replication study" should be reserved for situations in which *significant* linkage has already been obtained in an initial study (or combination of studies). Weaker findings do not merit the same standing as prior hypotheses. For example, there will be many regions with a nominal *P* value of 0.05 and some will appear to be 'replicated' in a second study just by chance (Box 2). We prefer the term "extension study" for the process of testing of additional families in the hope of first reaching the genome-wide significance level. Once significant linkage is found, it is appropriate to speak of 'replicating' the result.

Because replication involves testing an established prior hypothesis, the multiple testing problem associated with genome-wide search does not apply. Nonetheless, some caution is still required. The initial localization for a linkage is typically spread over a broad region of about 20 eM. Because one is searching over an interval, there is a multiple testing problem writ small: the chance of finding a *P* value of 0.05 *somewhere* within a 20 eM interval is greater than 5%. It turns out that a pointwise *P* value of'" 0.01 is needed for an interval-wide significance level of 5%. Accordingly, *P* = 0.01 should be required to declare confirmation at the 5% level. Note that this correction is equivalent to a multiple testing (or Bonferroni) correction for 5 markers.

Failure to replicate does not necessarily disprove a hypothesis. Linkages will often involve weak effects, which may turn out to be weaker in a second study. Indeed, there is a subtle but systematic reason for this: positive linkage results are somewhat biased because they include those weak effects that random fluctuations helped push above threshold, but exclude slightly stronger effects that random fluctuations happened to push below threshold. Initial positive reports will thus tend to overestimate effects, while subsequent studies will regress to the true value (see also ref. 15). Replication studies should always state their power to detect the ptoposed effect with the given sample size. Negative results are meaningful only if the power is high. Regrettably, many reports neglect this issue entirely.

When several replication studies are carried out, the resul$ may conflict — with some studies replicating the otiginal findings and others failing to do so. This may reflect population heterogeneity, diagnostic differen<:es, or simply statistical fluctuation. Careful meta-analysis of *all* studies may be useful to assess whether the overall evidence for linkage is convincing.

Suggestive linkages should be pursued in extension studies, in which old and new datasets are combined to see whether a significance evidence of linkage can be found. To combine results among studies, it is always best to pool the raw data and re-analyze the entire dataset. Lod scores can be added across studies, but only when they are computed by the same method, with the same set of markers, and at the same map position. Other meta-analysis techniques exist[16].

Statistical aficionados may recognize that extension studies involve a subtle multiple-testing problem of their own, because a significant result in *any* of the individual datasets or the combined dataset is often taken as evidence oflinkage. A modest multiple-testing correction to the genome-wide significance level should therefore be used in extension studies; the appropriate correction depends on study design. Of course, any combined analysis should include *all* studies — both positive and negative — to avoid biasing the results. If the combined analysis yields significant linkage, it is then appropriate to undertake a replication study.

Pursuing hunches

Formal procedures are useful for standardizing the general acceptance of linkage claims. Still, gene hunters should not be inhibited from pursuing all hints and hunches, including: following up all regions with pointwise *P* values of 0.05 (even though many will prove to be illusory); being encouraged if they find substantially more suggestive linkages than the one expected by chance (even though real loci cannot be distinguished from false positives); and using epidemiological arguments to infer the existence of loci with small effects (even though such inferences are highly model-dependent). It will, however, be worth having rigorous evidence in hand before undertaking positional cloning to avoid the unpleasant prospect of chasing a phantom locus.

Hints of linkage are usually followed by testing for linkage in larger datasets. Some true susceptibility loci, however, may never show significant linkage because they confer a very small increased risk and have common alleles. The proof that such loci are involved in disease aetiology must come from other data. Linkage disequilibrium can offer a powerful complement to traditional linkage studies. For loci having small effects but relatively few alleles in a population, tests of linkage disequilibrium can be much more sensitive than tests of linkage. A good example is *IDDM2,* the insulin gene, for which strong evidence of linkage disequilibrium is obtained in many datasets that fail to show linkage[17,18].

Linkage disequilibrium can be used in an exploratory fashion to pursue suggestive (or weaker) linkages (for example, ref. 19). Appropriate correction for multiple testing is essential in such applications — because multiple regions and many different haplotypes are tested for disease association. This topic has not received adequate attention and is an important area for future statistical research. Finally, we note that linkage disequilibrium studies should use family-based controls whenever possible to avoid false positive findings due to population stratification[20-22].

Other models and difficulties

The basic principles above apply to any analysis of

complex traits —whether by linkage analysis, allele-sharing methods, or quantitative trait mapping in experimental crosses (Box 1). The pointwise *P* values vary somewhat according to the method, but they are typically in the range of 10-3-10-4 for suggestive and 10-4-10-5 for significant linkage.

Nettlesome problems remain, however. Investigators often try out multiple diagnostic schemes for defining affectation status, as well as multiple models of inheritance for linkage analysis. Similarly, studies of quantitative traits may examine a large number of phenotypes. Datasets are frequently stratified using additional criteria, for example HLA genotype. What statistical price should be exacted for such fishing over multiple models? If the models are statistically independent, the observed *P* values should be multiplied by the number of models (which is known as the Bonferroni correction). This prescription is too conservative in the case of closely related models (such as correlated phenotypes), but there is no general guidance for how to proceed other than simulation. Even simulation poses a challenge, in that millions of simulations are needed to accurately estimate *P* values in the range of 10-5. Techniques such as importance sampling can make simulations much more feasible[23], and they should be performed whenever possible.

An additional difficulty is that false positive rates can be much higher than estimated if model parameters (such as gene frequency) are misspecified, if sample size is small, and if other assumptions of statistical independence are violated. A careful consideration of all these factors is beyond the scope of this commentary, but they offer an additional reason for caution in interpreting linkage results.

Examples of complex trait analyses

IDDM. Recent genome scans for insulin dependent diabetes mellitus (IDDM) illustrate the issues well. Davies and colleagues[24] used markers at an average spacing of 10 eM to survey the genome in 96 sib pairs, and then followed up some regions with lod 1 (*P* = 0.05) in two further collections with 102 and 84 sib pairs. Sib pairs were analysed together, and also divided according to HLA sharing. In the initial screen, only HLA met the standard for significant linkage, with lod = 8. Two further regions, on chromosomes Sq and X, showed suggestive linkage. A total of 20 regions had lod 1, which is not significantly greater than would be expected by chance.

TWo regions that fell somewhat short of the criterion for suggestive linkage were chosen for followup. A region on chromosome 11q (named *IDDM4* near *FGF)* had a *P* value of 0.01 in the combined dataset, but showed suggestive linkage in sib pairs sharing 1 or 0 alleles at HLA. In fact, an independent study found a nearly significant linkage in this region, but only in the subset of sibs in which both carried HLA-DR3[25]. Although the two studies were not jointly analyzed, it is likely the combined data would reveal significant linkage —indicating that this locus is probably real. The second region on chromosome 6q (named *IDDMS)* fell short of suggestive linkage in the combined dataset; it remains unclear whether there is in fact a susceptibility locus in this region. Interestingly, the same group subsequently reported that a region on chromosome 2 that fell far short of suggestive linkage (*P* = 0.01) showed evidence of linkage disequilibrium in some populations[19]. If widely confirmed (see, for example, ref. 26), this would underscore the value of linkage disequilibrium studies for identification of weak susceptibility loci.

A major contribution of these studies is that they demonstrate that there are no other loci with major effects comparable to HLA. The authors recognized this fact, but provided a valuable spur to further investigation by identifying the most promising regions for further study.

The IDDM story remains a work in progress[27,28]. It will probably require joint analyses of multiple datasets to sort out which of the hints of linkage are real. In general, it would be valuable if data from published genome scans were routinely deposited in an accessible form to facilitate such work.

Schizophrenia. Evidence for a susceptibility locus on chromosome 6p in a large collection of pedigrees from Ireland was reported by Wang *et a/.*[29] in a recent issue and Straub *et* a/.[30] in the current issue of this journal. The lod scores in these papers came extremely close to the standard for signficant linkage (corresponding to genome-wide significance levels in the range of 0.05-0.10). The authors carefully stress the need for replication.

Happily, two independent datasets reported in this issue by Moises *et* a/.[31] and Schwab *et al.*[32] appear to provide such evidence —each showing suggestive (and nearly significant) linkages in roughly the same region. Although a joint analysis has not yet been undertaken, it is dear that chromosome 6p meets the standard for significant-and, probably, confirmed -linkage.

Conclusion

The study of complex traits promises to be among the most important and challenging areas of mammalian genetics. As with any endeavor, the field will be shaped by the standards adopted by its practitioners. The traditional threshold of lod 3 has provided a rigorous standard that must be met to declare linkage for a simple Mendelian trait; it corresponds to a genome-wide false positive rate in the neighborhood of 5%. The proposed standard for significant linkage simply extends this same logic to the situation of complex traits; the required *P* values accord well with the practice in human genetics over the past three decades. At the same time, the category of suggestive linkage should facilitate reporting of tantalizing but unproven findings. By adopting clear rules for communication, human geneticists will be well prepared for the avalanche of information about to descend.

Acknowledgments

We thank M. Boehnke, A.Chakravarti, R. Elston, W. Ewens, S. Ghosh, f. MacCluer, M. Mahtani, M. McCarthy, f. Nolan, f. Ott, N. Schork, D. Siegmund, R. Spielman, f. Terwilliger, G. Thomson, f. Todd, and D. Weeks for helpful comments on the manuscript. This work was supported in part by grants from the National Center for Human Genome Research (to E.S.L. and L.K.).

1. Botstein, D., White, D.L., Skolnick, M. & Davis, R.W. Construction of a genetic Unkage map in man using restriction fragment length polymorphisms. Am. *J. hum. Genet.* 32, 314-{331 (1980).
2. Baron, M. eta/. Genetic linkage between X-chromosome mari<ers and bipolar affective illness. *Nature* 326, 289-292 (1987).
3. Egeland, J.A. et *at.* Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature* 325, 783-787 (1987).
4. Kelsoe, J.R. eta/. Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. *Nature* 342, 238-243 (1989).
5. Sherrington, R. *et at.* Localization of a susceptibility locus for schizophrenia on chromosome 5. *Nature* 336, 164-167 (1988).
6. Kennedy, J.L. et *at.* Evidence against Unkage of schizophrenia to markers on chromosome 5 in a northern Swedish pedigree. *Nature* 336, 167-170 (1988).
7. Baron, M. *et at.* Diminished support for Unkage between manic depressive illness and X-chromosome markers in three Israelipedigrees. *Nature Genet.* 3,49-55 (1993).
8. Lander, E.S. & Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185-199 (1989).
9. Feingold, E., Brown, P.O. & Siegmund, D. Gaussian models for genetic Unkage analysis using complete high-resolution maps of identity by descent. *Am. J. hum. Genet.* 53, 234-251 (1993).
10. Chotai, J. On the lod score method in linkage analysis. *Ann. hum.* Genet. 48, 35!H78 (1984).
11. Elston, R.C. Designs for the global search of the human genome by Unkage analysis. In Proceedings *of* the *16th international Biometrics* conterence39-51 (Hamitton, New Zealand, 1992).
12. Brown, D.L., Gorin, M.S. & Weeks, D.E. Efficient strategies for genomic searching using the affected-pedigree-member method of linkage analysis. *Am. J. hum.* Genet. 54,544-¢52 (1994).
13. McKusick, V.A. & Edwards, J.H. Unassigned syntenic groups and theoretical considerations. Second international workshop on human gene mapping. Cytogenet. Cei/Genet. 14,196-198(1975).
14. Thomson, G. Identifying complex disease genes: progress and paradigms. *Nature Genet.* 8,108-110 (1994).
15. Suarez, B.K., Hampa, C.L. & Van Eerdeweigh, P. Problems of replicating Unkage claims in psychiatry. In *Genetic approaches* to *mental disorders* (eds. Gershon, E.S. & Cloninger, C.R.) 23-46 (American Psychiatric Association, Washington, DC, 1994).
16. Cox, D.R. & Hinkley, D.V. *Theoretical* Statistics (Chapman & Hall, 1974).
17. Spielman, R.S., McGinnis, R.E. & Ewens, W.J. Transmission test for Unkage disequilibrium: the insulin gene region and insulin-dependant diabetes mellttus (lOOM). *Am. J. hum.* Genet. 52, 506-516 (1993).
18. Bennett, S.T. *et at.* Susceptibility to human type 1 diabetes at *IDDM2* is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nature Genet.* 9,284-92 (1995).
19. Copeman, J.B. eta/. Unkage disequilibrium mapping of a type 1 diabetes susceptibility gene *(IDDM7)* to chromosome 2q31-q33. *Nature Genet.* 9, 8()-85 (1995).
20. Ewens, W.J. & Spielman, R.S. The transmission/disequilibrium test: history, subdivision, and admixture. *Am. J. hum.* Genet. 57, 455-464 (1995).
21. Gough, S.C.L. *et at.* Mutation of the glucagon receptor gene and diabetes mellitus in the UK: association or founder effect? *Hum.* mol. *Genet.* 4, 1609-1612 (1995).
22. Thomson, G. Mapping disease genes: family-based association studies. *Am. J. hum.* Genet. 57, 487-498 (1995).
23. Terwilliger, J.D. & Ott, J. A multi-sample bootstrap approach to the estimation of maximized-over-models lod score distributions. Cytogenet. *Cell* Genet. 59, 142-144(1992).
24. Davies, J.L. *et at.* A genome-wide search for human typ1 diabetes susceptibiltygenes. Nature 371, 130-136 (1994).
25. Hashimoto, L. et *at.* Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. *Nature* 371, 161-164 (1994).
26. Luo, D., Maclaren, N.K., Huang, H., Muir, A. & She, J. Intrafamilial and case-control association analyses of D2S152 in insulin-dependent diabetes mellitus. *Autoimmunity* (1995, in press).
27. Todd, J.A. Genetic analysis of type 1 diabetes using whole genome approaches. *Proc. natn. Acad.* Sci. U.S.A. 92, 856Cl-8565 (1995).
28. Luo, D. *et at.* Affected-sib-pair mapping of a novelsusceptibility gene to insulin-dependent diabetes mellttus (IDDM8) on chromosome 6q25-q27. Am. *J. hum.* Genet. 57, 911-919(1995).
29. Wang, S. *et at.* Evidence for a susceptibility locus for schizophrenia on chromosome 6pter-p22. *Nature* Genet. 10, 41-46 (1995).
30. Straub, R.E. eta/. A potential vulnerabiltty locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. *Nature* Genet. 11, 287-293 (1995).
31. Moises, H.W. eta/. An international two-stage genome-wide search for schizophrenia susceptlbiltty genes. *Nature* Genet. 11,321-324 (1995).
32. Schwab, S.G. *et at.* Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nature* Genet. 11, 325-327 (1995).
33. Gyapay, G. *et at.* The 1993-94 Genethon human genetic linkage map. *Nature* Genet. 7,246-339 (1994).
34. Kruglyak, L. & Lander, E.S. Complete multipoint sib pair analysis of qualitative and quantitative traits. *Am. J. hum.* Genet. 57, 439-454 (1995).
35. Kruglyak, L. & Lander, E.S. A nonparametric approach for mapping quantitative trait loci. Genetics 139,1421-1428 (1995).
36. Feingold, E. Markov processes for modeling and analyzing a new genetic mapping method. *J. appl. Prob.* 30,766-779 (1993).
37. Holmans, P. Asymptotic properties of affected-sib-pair linkage analysis. *Am. J. hum.* Genet. 52, 362-374 (1993).
38. Lander, E.S. & Schork, N.J. Genetic dissection of complex tratts. *Science* 285,2037-2048 (1994).