Am. J. Hum. Genet. 71:1175–1182, 2002

Report

Evaluating the Results of Genomewide Linkage Scans of Complex Traits by Locus Counting

Steven Wiltshire,[1,2] Lon R. Cardon,[2] and Mark I. McCarthy[1,2]

[1]Imperial College Genetics and Genomics Research Institute, Imperial College, London, and [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

The evaluation of results from primary genomewide linkage scans of complex human traits remains an area of importance and considerable debate. Apart from the usual assessment of statistical significance by use of asymptotic and empirical calculations, an additional means of evaluation—based on counting the number of distinct regions showing evidence of linkage—is possible. We have explored the characteristics of such a locus-counting method over a range of experimental conditions typically encountered during genomewide scans for complex trait loci. Under the null hypothesis, factors that have an impact on the informativeness of the data—such as map density, availability of parental data, and completeness of genotyping—are seen to markedly influence the number of regions of excess allele sharing and the empirically derived genomewide significance of the associated LOD score thresholds. In some circumstances, the expected number of regions is less than one-quarter of that predicted under the assumption of a dense map and complete extraction of inheritance information. We have applied this method to a previously analyzed data set—the Warren 2 genome scan for type 2–diabetes susceptibility—and demonstrate that more regions showing evidence for linkage were observed in the primary genome scan than would be expected by chance, across the whole range of LOD scores, even though no single linkage result achieved empirical genomewide statistical significance. Locus counting may be useful in assessing the results from genome scans for complex traits in general, especially because relatively few scans generate evidence for linkage reaching genomewide significance by dense-map criteria. By taking account of the effects of reduced data informativeness on the expected number of regions showing evidence for linkage, a more meaningful, and less conservative, evaluation of the results from such linkage studies is possible.

Genomewide linkage scans have become widely used tools in research efforts to unravel the genetics of complex traits in humans. Since the first published scan, of type 1 diabetes (Davies et al. 1994), >100 genome scans of multifactorial diseases in humans have been published (Altmuller et al. 2001). However, less than one-third of these have identified regions achieving genomewide significance (at $P = .05$) when interpreted according to guidelines derived on the assumption of dense-marker genotyping and complete (or near complete) extraction of inheritance information (Lander and Kruglyak 1995).

This failure to detect significant linkage in a large proportion of scans reflects a number of processes, ranging from methodological considerations—such as inadequate sample size, extensive clinical heterogeneity, and latent genotyping error—to etiological factors—such as the phenotype-genotype relationship and the genetic architecture of the trait (Lernmark and Ott 1998; Weiss and Terwilliger 2000; Altmuller et al. 2001; Nicolae and Cox 2002). In addition, experimental factors that influence the proportion of information extracted from the data set—such as pedigree structure, the completeness and accuracy of genotype data, and marker density—are expected to have substantial effects, both on the power of a study to detect true linkage and on the significance of any linkage finding. For this reason, it is widely accepted that accurate estimates of the genomewide statistical significance of linkage results are best obtained empirically (Sawcer et al. 1997; Ott 1999; Douglas et al. 2000; Gordon et al. 2000; Hirschhorn et al. 2001).

An alternative method of evaluating the results of such

Received May 22, 2002; accepted for publication July 16, 2002; electronically published September 25, 2002.

Address for correspondence and reprints: Dr. Steven Wiltshire, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, United Kingdom. E-mail: steven.wiltshire@well.ox.ac.uk

0002-9297/2002/7105-0016$15.00

Table 1

Expected Numbers of IRLs per Genome Scan under the Null Hypothesis

Map Density and IRL LOD Score Threshold[a]	Expected Number of IRLs per Genome Scan			
	Parental Data Available		No Parental Data Available	
	No Missing Genotypes	15% Missing Genotypes	No Missing Genotypes	15%Missing Genotypes
10 cM:				
⩾ .59 (P ⩽.05)	10.611	9.180	9.002	8.577
⩾ 1.18 (P ⩽.01)	2.820	2.341	2.326	2.125
⩾ 3.00 (P ⩽.0001)	.051	.038	.044	.030
⩾ 2.20	.290	.228	.251	.198
⩾ 3.63	.016	.007	.014	.007
⩾ 5.30	.001	.002	0	.001
5 cM:				
⩾ .59 (P ⩽.05)	12.609	11.299	10.501	9.720
⩾ 1.18 (P ⩽.01)	3.525	3.099	2.764	2.590
⩾ 3.00 (P ⩽.0001)	.070	.061	.044	.033
⩾ 2.20	.375	.314	.263	.244
⩾ 3.63	.022	.013	.008	.010
⩾ 5.30	.002	0	0	0
1 cM:				
⩾ .59 (P ⩽.05)	16.680	15.496	14.413	13.581
⩾ 1.18 (P ⩽.01)	5.182	4.624	4.197	3.996
⩾ 3.00 (P ⩽.0001)	.107	.093	.091	.068
⩾ 2.20	.597	.537	.449	.449
⩾ 3.63	.031	.027	.023	.022
⩾ 5.30	.002	0	.001	.001

Note.—These results are from analyses of 1,000 replicates of 500 two-sib families analyzed with a linear model using GENEHUNTER PLUS.

[a] P values in parentheses after each IRL LOD score threshold represent nominal pointwise significance levels. LOD score ⩾2.20 indicates suggestive linkage, LOD score ⩾3.63 indicates significant linkage, and LOD score ⩾5.30 indicates highly significant linkage, according to dense-map significance guidelines (Lander and Kruglyak 1995).

genome scans, complementary to the use of genomewide significance levels, is to compare the actual number of loci showing evidence for linkage observed from the genome scan with the number expected by chance, determined empirically under the experimental and data conditions prevailing in the study. For an evaluation based on locus counting to be meaningful, however, only independent (that is, unlinked) regions should be considered in the comparison. This locus-counting approach is appealing, because complex traits are under the control of multiple genes and because evidence for the influence of at least some of them is anticipated from a genomewide linkage scan. Furthermore, the ability to determine whether a genome scan has detected more regions showing evidence for linkage than expected by chance when no single region has achieved genomewide statistical significance will be of particular benefit.

The expected number of loci under the null hypothesis, determined from the theory of large deviations, formed the basis of widely used guidelines for significance level (Lander and Kruglyak 1995): the threshold for designating a region as showing "suggestive" evidence for linkage (LOD = 2.20) was defined as that LOD score expected once by chance per genome scan, given a dense map and near-complete information extraction. However, despite recent consideration of this measure (Sawcer et al. 1997; Hsieh et al. 2001; Lindholm et al. 2001), there has been no formal examination of locus counting as a means of evaluating results from genetic studies. In particular, the behavior of such a measure has not been established under the experimental circumstances when it is most likely to be used, namely, the evaluation of primary genome scan data from which complete extraction of inheritance information has not been possible. To address this, we have examined, by simulation, the genomewide null distribution of the widely used allele-sharing LOD score for dichotomous traits (using both linear and exponential models [Kong and Cox 1997]) over a range of experimental conditions, from which we have derived the numbers of independent loci over a range of LOD score thresholds.

For our simulation analyses, we generated complete autosomal genomes (total length 34.6 M [Kosambi {K}]), under the null hypothesis of no linkage, for sets of fam-

Table 2

Empirical Genomewide Significance Levels of IRL LOD Score Thresholds

Map Density and IRL LOD Score Threshold[a]	Empirical Genomewide Significance			
	Parental Data Available		No Parental Data Available	
	No Missing Genotypes	15% Missing Genotypes	No Missing Genotypes	15% Missing Genotypes
10 cM:				
3.00 (P = .0001)	.050	.038	.043	.030
2.20	.259	.205	.222	.181
3.63	.016	.007	.014	.007
5.30	.001	.002	<.001	.001
5 cM:				
3.00 (P = .0001)	.068	.057	.043	.033
2.20	.315	.262	.231	.216
3.63	.022	.013	.008	.010
5.30	.002	<.001	<.001	<.001
1 cM:				
3.00 (P = .0001)	.103	.087	.086	.067
2.20	.457	.401	.351	.361
3.63	.031	.026	.023	.022
5.30	.002	<.001	.001	.001

Note.—These results are from analyses of 1,000 replicates of 500 two-sib families analyzed with a linear model using GENEHUNTER PLUS.

[a] P values in parentheses after each IRL LOD score threshold represent nominal pointwise significance levels. LOD score ⩾2.20 indicates suggestive linkage, LOD score ⩾3.63 indicates significant linkage, and LOD score ⩾5.30 indicates highly significant linkage, according to dense-map significance guidelines (Lander and Kruglyak 1995).

ilies that each contain only two affected siblings. These simulations were based on the microsatellite marker maps from CEPH (Dib et al. 1996; see the Généthon Web site) and the Marshfield Foundation (Broman et al. 1998; see the Center for Medical Genetics, Marshfield Medical Research Foundation, Web site), and both assumed markers with four equally frequent alleles and no undetected genotyping error. To generate data sets differing in the degree of inheritance information available for extraction, several parameters were varied, including marker density (10 cM and 5 cM, to cover the range of marker densities encountered in primary genome scans, and 1 cM, to approximate genomewide dense mapping); presence and absence of parental genotypes (to model the situations typically encountered when studying early-onset and late-onset diseases, respectively); and genotyping completeness (0% and 15% missing genotypes). The effects of varying sample size (100, 200, and 500 families) were also considered. For each combination of parameters, we generated 1,000 replicates and analyzed these with both linear and exponential models implemented in GENEHUNTER PLUS and ASM (Kong and Cox 1997). We recorded the maximum LOD score of every independent region showing evidence for linkage (IRL), treating such regions as unlinked (and therefore independent) if their maxima were ⩾40 cM (K) apart, following the reasoning presented by Ott (1999). From these, we determined the number of IRLs observed at various defined LOD score thresholds, together with the empirical genomewide significance levels of these thresholds.

The results from these simulations, available in full from the IRL Study, Wellcome Trust Centre for Human Genetics, Web site, demonstrate that neither sample size nor analytical model (i.e., linear or exponential) exerts a strong influence on the null IRL distribution. In the presentation of our findings here, we therefore concentrate on simulations performed under the assumption of 500 sib-pair families, analyzed with a linear model. Tables 1 and 2 display the number of IRLs expected under the null hypothesis at several frequently used LOD score thresholds, together with associated empirical genomewide significance levels; the complete distributions are shown in figures 1 and 2.

Our simulation results demonstrate the strong impact of experimental factors that influence the extraction of inheritance information from genetic data—map density, availability of parental genotypes, and completeness of genotype data—on the null IRL distribution and the associated empirical significances. As expected, the number of IRLs at any given LOD score threshold rises when map density increases, parental genotypes become available, and/or missing genotype rate falls (table 1; fig. 1). In parallel, the same three parameters influence consid-

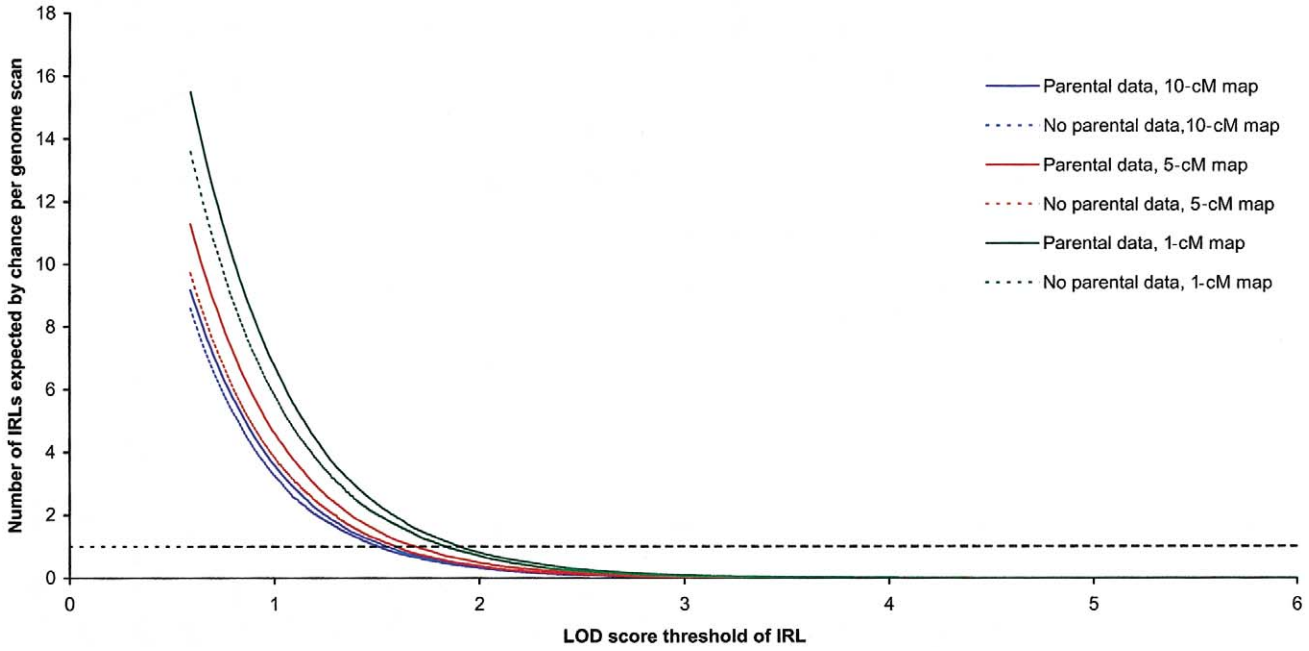

Figure 1 Null IRL distributions for 500 two-sib families, simulated with a missing genotype rate of 15% and analyzed with a linear model using GENEHUNTER PLUS. The figure shows the effects of three different marker-map densities and the availability of parental genotypes on the expected number of IRLs per genome scan, over the LOD score threshold range 0.59–6. The position of one IRL expected by chance per genome scan is indicated by the horizontal broken line.

erably the empirical genomewide significance of the LOD score thresholds, and the *P* value associated with any given LOD score becomes smaller as map density increases, parental data becomes available, and genotyping becomes more complete (table 2; fig. 2).

These findings emphasize the need to account for reduced data informativeness when evaluating primary linkage scans by locus counting, as well as by measures of statistical significance. Our simulations suggest that, under the experimental and data conditions encountered during a typical primary autosomal genome scan (a 10-cM marker map and ~15% missing genotypes), an IRL with a LOD score of 1.51–1.55 is expected to occur once by chance per autosomal genome scan and that a LOD score of 2.80–2.88 is associated with a genomewide significance of .05, depending on the availability of parental data (table 3). These thresholds contrast with those of LOD = 2.20 and LOD = 3.63, respectively, predicted under assumptions of complete information extraction from dense-marker-map scenarios. However, these results are in close agreement with thresholds predicted from single-point analyses, using a 10-cM marker map (Lander and Kruglyak 1995; Ott 1999), which have indicated values ~20% lower than the dense-map thresholds. However, most striking is the decrease in the frequency of an IRL with LOD = 2.20 from once per genome scan—according to the dense-map criteria—to only ~0.2 times per scan, under conditions of a typical primary genome scan (table 1). These findings reveal the extent to which dense-map criteria may be conservative when applied to results from primary genome scans (Witte et al. 1996; Sawcer et al. 1997; Nicolae and Cox 2002).

Our analyses with 1-cM marker maps also demonstrate, under situations in which near complete extraction of inheritance information is possible, that the null IRL distribution shows evidence of convergence to the threshold criteria predicted by large-deviation theory and 0.1-cM maps (Lander and Kruglyak 1995). From our simulations using the most informative model (the presence of parental data and no missing genotypes, with a mean entropy-based information content of 97%), we expect ~17 IRLs with LOD $\geq$ 0.59, compared with ~20 predicted for an autosomal genome on the basis of asymptotic theory (see equation 1 in Lander and Kruglyak 1995). Furthermore, we expect an IRL with LOD = 1.96 to occur once by chance per genome scan, and we expect an empirical genomewide significance level of .05 to be associated with a LOD score of 3.30 (table 3). These values compare reasonably well with the published dense-map LOD score thresholds of 2.20 and 3.63, respectively (Lander and Kruglyak 1995; Ott 1999).

To illustrate the application of this locus-counting approach to the evaluation of results from actual genetic

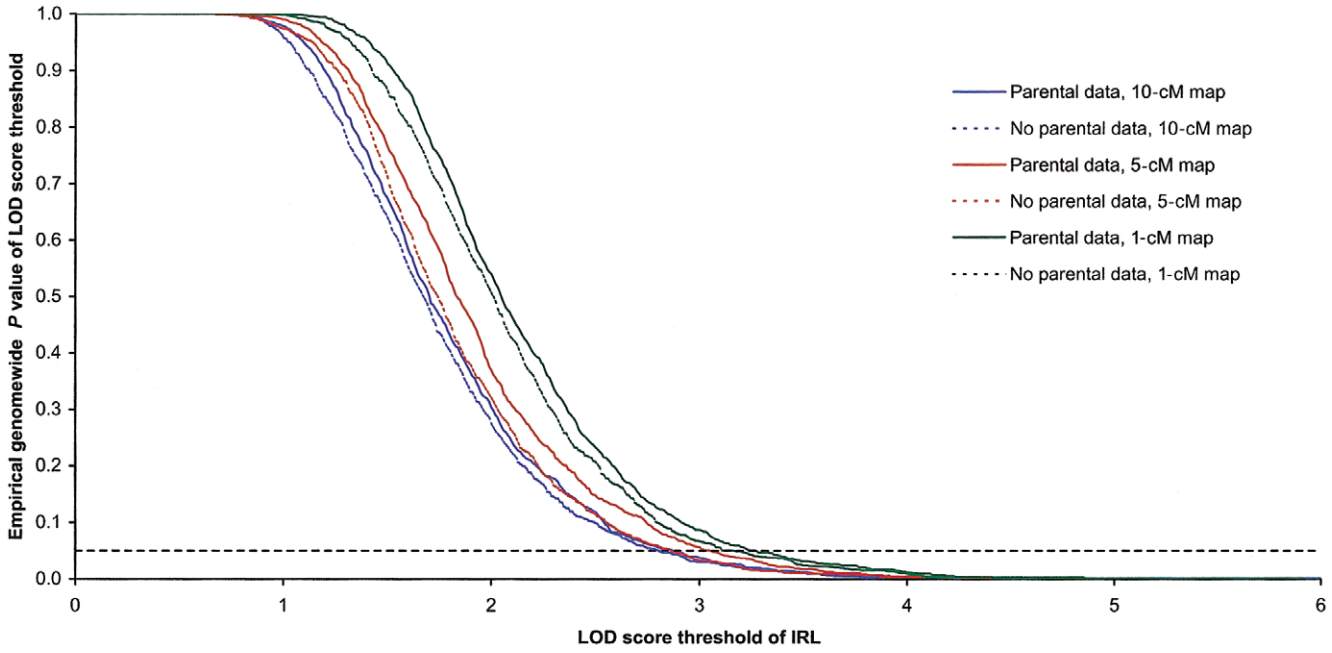

Figure 2 Empirical genomewide significance levels for 500 two-sib families, simulated with a missing genotype rate of 15% and analyzed with a linear model using GENEHUNTER PLUS. The figure shows the effects of three different marker-map densities and the availability of parental genotypes on the empirical genomewide significance levels of the full range of IRL LOD score thresholds. The position of an empirical genomewide significance level of .05 is indicated by the horizontal broken line.

studies, we have reexamined the Diabetes UK Warren 2 genome scan for type 2–diabetes susceptibility (Wiltshire et al. 2001). This genome scan examined 573 families multiplex for type 2 diabetes. Most of the families were small sibships (the majority with only two affected sibs), with very few parental genotypes. The primary scan was conducted using an autosomal microsatellite marker map of 418 markers, with a mean marker spacing of 9.26 cM (Haldane [H]), and the overall missing genotype rate in typed individuals was 14% (see the Warren 2 Project Information on the Wellcome Trust Centre for Human Genetics Web site). To estimate the null distribution of IRL LOD scores for the Warren 2 study, 1,000 replicates of the autosomal genome were simulated under the null hypothesis of no linkage, using the precise marker map, allele frequencies, and missing genotype pattern observed in the experimental data through use of SIMULATE (see the Laboratory of Sta-

Table 3

Selected IRL LOD Score Thresholds

Map Density and LOD Score	Empirically Observed LOD Score			
	Parental Data Available		No Parental Data Available	
	No Missing Genotypes	15% Missing Genotypes	No Missing Genotypes	15% Missing Genotypes
10 cM:				
Expected once by chance per genome scan	1.627	1.548	1.556	1.510
Associated with an empirical genomewide $P = .05$	2.992	2.877	2.956	2.803
5 cM:				
Expected once by chance per genome scan	1.773	1.693	1.644	1.594
Associated with an empirical genomewide $P = .05$	3.166	3.058	2.928	2.867
1 cM:				
Expected once by chance per genome scan	1.962	1.906	1.855	1.839
Associated with an empirical genomewide $P = .05$	3.304	3.271	3.308	3.161

Note.—These results are from analyses of 1,000 replicates of 500 two-sib families analyzed with a linear model using GENEHUNTER PLUS.

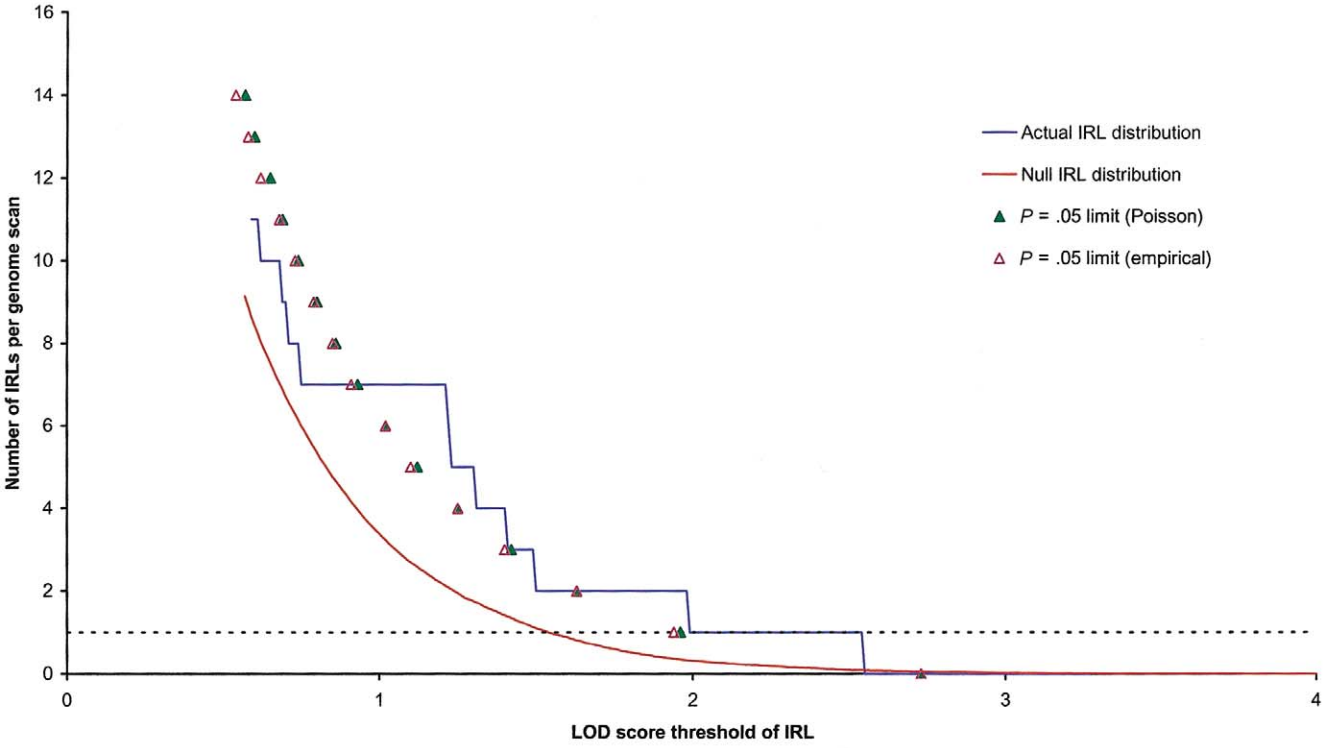

Figure 3 Null IRL distribution (i.e., the mean number of IRLs at each LOD score threshold under the null hypothesis) generated from the Warren 2 genome scan data, and the actual IRL distribution observed from the Warren 2 genome scan, analyzed with an exponential model using ALLEGRO. The $P = .05$ significance limits for the number of IRLs per genome scan, estimated empirically and from a Poisson distribution, are shown in order to indicate the significance of the actual IRL distribution. The position of one IRL expected by chance per genome scan is indicated by the horizontal broken line.

tistical Genetics, Rockefeller University, Web site). These were then analyzed for linkage, using ALLEGRO (Gudbjartsson et al. 2000). To ensure consistency with the original linkage analyses, which were conducted using a Haldane map function, regions of linkage were considered to be independent if their respective maxima were separated by ⩾55 cM (H), which is equivalent to 40 cM (K).

The actual Warren 2 linkage scan detected 11 IRLs, with LOD scores ⩾0.59, 7 with LOD scores ⩾1.18, and 1 with a LOD score ⩾2.20 (8p21-22, LOD = 2.55, genomewide $P = .098$); no region showed evidence for linkage reaching genomewide significance ($P \leqslant .05$) according to empirical calculations (Wiltshire et al. 2001). In contrast, the null IRL distribution generated from the Warren 2 sample suggests that we expect ~9 IRLs at LOD ⩾0.59, ~2 IRLs at LOD ⩾1.18, ~0.2 at LOD ⩾2.20, and ~0.03 at LOD ⩾ 3.63 to occur by chance. Furthermore, we expect to see an IRL with LOD = 1.54 once by chance, and we expect a LOD score of 2.84 to be associated with an empirical genomewide significance level of .05. These results agree closely with those from the simulations shown in tables 1 and 2, despite the differences in pedigree structure and marker characteristics between the actual Warren 2 data and the simulated pedigree data sets. We have compared the actual distribution of linkage results from the Warren 2 genome scan with the null IRL distribution (that is, the mean number of IRLs per genome scan at each LOD score threshold) determined from the Warren 2 sample (fig. 3). At any given LOD score threshold, the frequency distribution of regions of linkage approximates a Poisson distribution (Lander and Kruglyak 1995). To obtain an estimate of the significance of our findings, we have determined the number of IRLs with cumulative probability of .95 (i.e., representing a one-tailed significance limit of .05) for LOD score thresholds across the range, both empirically from our simulations, and from a cumulative Poisson distribution (Ross 1988) with means taken from the null IRL distribution (fig. 3). We observe a shift of the actual IRL distribution from the null IRL distribution toward, and at several points exceeding, the $P = .05$ limits calculated from the null IRL distribution. These findings suggest that the Warren 2 genome scan detected more regions showing evidence for linkage across the whole LOD score range than would be ex-

pected by chance, at a significance level of ~.05, even though no individual linkage result achieved statistical significance at the genomewide level.

In conclusion, we have examined the use of locus counting as an additional method for evaluating complex-trait genome-scan results that is complementary to measurements (empirical or asymptotic) of statistical significance. We have shown by simulation that experimental factors influencing the informativeness of the data—such as marker density and the availability of complete genotypes—have a marked effect on the number of IRLs at given LOD score thresholds expected under the null hypothesis of no linkage. These factors must therefore be taken into consideration when undertaking such an evaluation. However, the analytical model appears to have little impact on null IRL distributions generated from the simple pedigrees examined here; however, this may not be true for more complex pedigree structures, and it merits further investigation. Consequently, for a typical primary (~10-cM) genome scan of a complex trait, we expect 8–11 IRLs with LOD = 0.59 and 2–3 with LOD $\geq$ 1.18 by chance per scan, depending on the precise characteristics of the data; IRLs with LOD scores $\geq$3 occurring by chance are very rare events in such scans. Furthermore, we expect to see by chance one region per genome scan showing evidence for linkage with a LOD score of between 1.5 and 1.7.

We recommend performing and presenting the results from such an analysis over the full range of LOD score thresholds for the data in question and have demonstrated this approach by application to the Warren 2 genome scan data. When this is done, the shift of the entire actual IRL distribution in the Warren 2 data, relative to the null IRL distribution, becomes apparent. The overall significance of any such shift in actual IRL distributions can be estimated, as we have shown, although the development of a formal measure of significance merits attention. Although this method appears promising as a means of evaluating genome scan results, several caveats should be emphasized. First, we caution against performing or reporting such an analysis at a single LOD score, because this may be unrepresentative and may invite researchers to present only the most extreme result. Second, it is important to note that a shift in the actual IRL distribution is not, itself, a measure of the genomewide significance of the associated LOD score thresholds. Third, locus counting does not allow determination of which regions detected during the actual study represent true positives and which are false positives. Despite these caveats, we believe that locus counting will be a useful additional tool for the evaluation of primary genome scans for complex trait loci. The use of empirical methods of evaluation that are tailored to the precise characteristics of the data in question is likely to provide more realistic and optimistic assessments of the evidence for complex trait linkage.

Acknowledgments

We thank Dr. Stacey Cherny (Wellcome Trust Centre for Human Genetics, University of Oxford) and Dr. Jeff Lessem (Institute of Behavioral Genetics, University of Colorado), for the development and maintenance of the Linux cluster computing facility at the Wellcome Trust Centre for Human Genetics, and Dr. Goncalo Abecasis (Center for Statistical Genetics, University of Michigan), for C++ allele-dropping code. We thank the Diabetes UK Warren 2 Consortium (Mark McCarthy, Andrew Hattersley, Graham Hitman, Jonathan Levy, Mark Walker, Stephen O'Rahilly, Stephan Menzel, John Bell, and Mark Lathrop) for access to the Warren 2 Project genotype data. We are grateful to Dr. Richard Mott (Wellcome Trust Centre for Human Genetics, University of Oxford), for useful discussion, and to the anonymous reviewers, for their helpful comments. This study was funded by the Wellcome Trust and by Diabetes UK.

Electronic-Database Information

URLs for data presented herein are as follows:

Center for Medical Genetics, Marshfield Medical Research Foundation, http://research.marshfieldclinic.org/genetics/ (for genetic linkage maps)
Généthon, http://www.genethon.fr/ (for genetic linkage maps)
IRL Study, Wellcome Trust Centre for Human Genetics, http://www.well.ox.ac.uk/IRL (for additional results from the present study)
Laboratory of Statistical Genetics, Rockefeller University, http://linkage.rockefeller.edu/ (for SIMULATE)
Warren 2 Project Information, Wellcome Trust Centre for Human Genetics, http://www.well.ox.ac.uk/warren2/ (for Warren 2 genome-scan marker information)

References

Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M (2001) Genomewide scans of complex human disease: true linkage is hard to find. Am J Hum Genet 69:936–950

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861–869

Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SCL, Jenkins SC, Palmer SM, Balfour KM, Rowe BR, Farrall M, Barnett AH, Bain SC, Todd JA (1994) A genome-wide search for human type 1 diabetes susceptibility genes. Nature 371:130–136

Dib C, Fauré S, Fizames C, Samson D, Druout N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Douglas JA, Boehnke M, Lange K (2000) A multipoint method

for detecting genotyping errors and mutations in sibling-pair linkage data. Am J Hum Genet 66:1287–1297

Gordon D, Leal SM, Heath SC, Ott J (2000) An analytic solution to single nucleotide polymorphism error-detection rates in nuclear families: implications for study design. Pac Symp Biocomput 663–674

Gudbjartsson DF, Jonasson K, Frigge ML, Kong CA (2000) Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13

Hirschhorn JN, Lindgren CM, Daly MJ, Kirby A, Schaffner SF, Burtt NP, Altshuler D, Parker A, Rioux JD, Platko J, Gaudet D, Hudson TJ, Groop LC, Lander ES (2001) Genomewide linkage analysis of stature in multiple populations reveals several regions with evidence of linkage to adult height. Am J Hum Genet 69:106–116

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y et al. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 26:163–175

Hsieh C-L, Oakley-Girvan I, Balise RR, Halpern J, Gallagher RP, Wu AH, Kolonel LN, O'Brien LE, Lin IG, Van Den Berg DJ, Teh C-Z, West DW, Whittemore AS (2001) A genome screen of families with multiple cases of prostate cancer: evidence of genetic heterogeneity. Am J Hum Genet 69:148–158

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

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

Lernmark A, Ott J (1998) Sometimes it's hot, sometimes it's not. Nat Genet 19:213–214

Lindholm E, Ekholm B, Shaw S, Jalonen P, Johansson G, Pettersson U, Sherrington R, Adolfsson R, Jazin E (2001) A schizophrenia-susceptibility locus at 6q25, in one of the world's largest reported pedigrees. Am J Hum Genet 69:96–105

Nicolae DL, Cox NJ (2002) MERLIN … and the geneticist's stone? Nat Genet 30:3–4

Ott J (1999) Analysis of human genetic linkage. The Johns Hopkins University Press, Baltimore

Ross S (1988) A first course in probability. Macmillan, New York

Sawcer S, Jones HB, Judge D, Visser F, Compston A, Goodfellow PN, Clayton D (1997) Empirical genome-wide significance levels established by whole genome simulations. Genet Epidemiol 14:223–229

Weiss KM, Terwilliger JD (2000) How many diseases does it take to map a gene with SNPs? Nat Genet 26:151–157

Wiltshire S, Hattersley AT, Hitman GA, Walker M, Levy JC, Sampson M, O'Rahilly S, Frayling TM, Bell JI, Lathrop GM, Bennett A, Dhillon R, Fletcher G, Groves CJ, Jones E, Prestwich P, Simecek N, Subba Rao PV, Wishart M, Foxon R, Howell S, Smedley S, Cardon LR, Menzel S, McCarthy MI (2001) A genomewide scan for loci predisposing to type 2 diabetes in a UK population (The Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. Am J Hum Genet 69:553–569

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