

Use of Weighted p-Values in Regional Inference Procedures

Xiaohong Yang, Alisa M. Goldstein, Gary A. Chase, Joseph L. Gastwirth, and Lynn R. Goldin

Genetic Epidemiology Branch (X.Y., A.M.G., L.R.G.), National Cancer Institute, NIH, Rockville, Maryland; Department of Biostatistics and Research Epidemiology (G.A.C.), Henry Ford Health System, Detroit, Michigan; Department of Biostatistics (J.L.G.), George Washington University, Washington, DC

Our previous studies have demonstrated that the power to detect linkage was improved by calculating a moving average of consecutive p-values in a small region as compared with testing all single p-values. The goal of this study was to test whether the power can be improved further with an alternative method whereby the middle p-values in the sequence were given more weight than the others. We also wanted to compare the moving average tests with multipoint linkage tests. The simulated extended pedigree data from the general population was analyzed to identify two major genes (MG1 and MG5) underlying two quantitative traits (Q1 and Q5). We used the variance components method implemented in the GENEHUNTER program to test for linkage of 14-marker regions each on chromosome 19 and chromosome 1 to the adjusted quantitative traits Q1 and Q5, respectively, in all 50 replicates. As before, we found that the moving average test was more powerful than a test based on single p-values. In some cases, the weighting procedure increased the power further and was similar to that of multipoint analysis, but this was not consistently found. In addition, all methods had low power and it is not possible to make a general conclusion that some weighting schemes are better than others. © 2001 Wiley-Liss, Inc.

Key words: linkage, multipoint analysis, power, quantitative trait, regional linkage methods

INTRODUCTION

Linkage analysis for complex diseases often has low power due to the small to moderate effects of the susceptibility genes and to the limited sample size of families

Address reprint requests to Dr. Xiaohong (Rose) Yang, Genetic Epidemiology Branch, DCEG/NCI/NIH, Bldg. EPS/Rm. 7005, 6120 Executive Blvd., Rockville, MD 20852.

© 2001 Wiley-Liss, Inc.

that are ty
analysis of
proposed t
keep the p
p-value ne
low. Subse
wider than
two or mo
linkage tha
1999]. A s

In th
two quanti
Genetic A
statistic ac
greater we
to increase
consecuti
power of t

METHOD

The
the quanti
accounted
gene, MG
regressio
untransfo
and other
independ
the first
predict Q
of these c
adjustme

The resi
quantitat

All
Genotyp
major ge
primary
consider
cM as th

Regional

Joseph L. Gastwirth,

National Cancer Institute,
Research Epidemiology
Department of
Washington, DC

to detect linkage was
p-values in a small
of this study was to
an alternative method
more weight than the
gests with multipoint
the general population
MG5) underlying two
components method
linkage of 14-marker
adjusted quantitative
fore, we found that the
on single p-values. In
urther and was similar
found. In addition, all
general conclusion that
ey-Liss, Inc.

regional linkage methods

power due to the small to
ted sample size of families

ology Branch, DCEG/NCI/NIH,

that are typically available. The threshold appropriate for significant linkage in the analysis of complex diseases is currently under debate. Lander and Kruglyak [1995] have proposed that critical values for significant linkage for a whole genome scan be set to keep the posterior false positive rate no greater than 5%. Because of the highly stringent p-value needed for a significant test at a single locus, the power of linkage tests can be low. Subsequently, Terwilliger et al. [1997] demonstrated that true linkage peaks are wider than false peaks. We previously demonstrated that calculating a moving average of two or more consecutive p-values in a small region was more powerful for detecting linkage than that of testing all single p-values [Goldin et al., 1999; Goldin and Chase, 1999]. A similar procedure has been suggested by Hoh and Ott [2000].

In this study, we utilized a variance components method to test for linkage between two quantitative traits (Q1 and Q5) and two major genes (MG1 and MG5) using the Genetic Analysis Workshop (GAW) 12 simulated data. Given the shape of a linkage statistic across a chromosome that contains a true trait locus, it is possible that giving greater weights to the middle markers could improve the power. We have thus attempted to increase the power of the regional test even more by applying unequal weights to the consecutive p-values instead of using a simple average. We also wanted to compare the power of the regional method with that of a complete multipoint analysis.

METHODS

The quantitative trait Q1 was first chosen as the focus for our study. Subsequently, the quantitative trait Q5 was also selected for the analysis since the major gene effect accounted for a higher proportion of the trait variance. Our goal was to map the major gene, MG1, for Q1 and the major gene, MG5, for Q5. We first performed multiple linear regression analyses to adjust separately for the effects of covariates on Q1 and Q5. The untransformed values of Q1 and Q5 were used as the dependent variables, respectively, and other quantitative traits (Q2, Q3, Q4), age, sex, household, EF1 and EF2 as the independent variables. The regression coefficients were computed and compared using the first 10 replicates. Age and sex were the only consistently significant variables to predict Q1. All variables except household were significant to predict Q5. Mean values of these coefficients from these 10 replicates were taken. The regression models used for adjustment of Q1 and Q5 were:

$$\begin{aligned} Q1 &= 10.28 + 1.935*Sex + 0.1*Age; \\ Q5 &= 14.413 + 2.468*Sex + 0.106*Age + 0.011*EF1 + 0.125*EF2 \\ &\quad + 0.134*Q1 + 0.035*Q2 + 0.151*Q3 + 0.114*Q4. \end{aligned}$$

The residual values of Q1 and Q5 corrected for these covariates were used as the quantitative values in the following analyses.

All 50 replicates from the general population were included in the linkage analyses. Genotype information on deceased individuals was also included in all analyses. Since the major genes MG1 on chromosome 19 and MG5 on chromosome 1 were known to be the primary determinants of genetic variances for the quantitative traits Q1 and Q5, we considered 14 markers each on chromosome 19 and chromosome 1, respectively, with 3 cM as the average interval between adjacent markers. The trait locus was located at the

center of the interval with about 20 cM spanning on each side of this 14-marker region. We used the variance components method implemented in the program GENEHUNTER (version 2.0) for both two-point and multipoint linkage analyses. This method was chosen since it has been shown to have greater power than traditional sib-pair analysis [Pratt et al., 2000]. Because the GENEHUNTER program is best suited for pedigrees of moderate size, we broke the 23 extended pedigrees into 42 smaller families for the linkage analyses. Initially, 1,257 individuals from each replicate were included in the analysis; 420 additional people were dropped by the program. Thus 837 individuals (including living and deceased) were included in the final analyses. Since our goal was to compare the power of alternative methods, we did not attempt to maximize the number of informative individuals in the analysis. Two-point and multipoint lod scores were computed under the assumption of additive variance. p-Values at each locus from both two-point and multipoint linkage analyses were calculated after transforming the lod scores to chi-square statistics ($\text{chi-square} = 4.6 \times \text{lod}$) using the SAS program [SAS, 1999]. Because of the boundary constraint on the quantitative trait locus (QTL) variance, we assumed that the distribution of the lod score was a 50:50 mixture of a chi-square statistic with one degree of freedom and a point mass at zero [Self and Liang, 1987].

We have previously shown by simulation that power to detect linkage was improved by calculating a moving average of two or more consecutive two-point p-values as compared with testing all single p-values [Goldin et al., 1999]. We have extended this method to test whether the power can be improved further by applying unequal weights to the consecutive p-values. We had previously determined the thresholds for significant (defined as a posterior false positive rate of 5% for a whole genome scan) and suggestive (defined as no more than one false positive finding per genome scan) linkage by simulating 1,000 genome scans under the null hypothesis of no linkage of a trait locus to the map. Using the same simulations, we re-computed these thresholds for different weighting schemes and then calculated the power to detect linkage (using the Haseman-Elston test) for the data sets simulated with linkage. Table I shows the thresholds and power for the case of a trait with 90% heritability located in the middle of a map of 11 markers spaced 3 cM apart (see Goldin et al. [1999] for details of the simulations). The power is modestly increased when more weight is placed on the two inner markers.

We applied this method to all 50 replicates of the GAW12 simulated data and compared the power from these different weighting methods with that of standard two-point and multipoint analysis for both significant and suggestive linkage thresholds. For the Q5 analysis, two-point lod scores could not be obtained for one replicate (#32) and it was dropped from the power analyses. For two-point analysis, we applied these different weighting rules to every group of four consecutive markers shifting by one marker each time. For multipoint analysis, the significance of the test was solely based on p-values at each tested location. For the single p-value test in both two-point and multipoint

TABLE I. Comparison of Power of Linkage Detection for Different Weighting Schemes (3-cM Map, 90% Trait Heritability)

Weights for 4 markers				Threshold		Power	
M1	M2	M3	M4	Significant	Suggestive	Significant	Suggestive
0.25	0.25	0.25	0.25	0.0028	0.01755	51.6	86.3
0.20	0.30	0.30	0.20	0.002713	0.017162	52.2	86.4
0.10	0.40	0.40	0.10	0.0023915	0.014486	53.5	87.0
0.05	0.45	0.45	0.05	0.0018941	0.011608	54.2	86.6

TABLE II. Power to Detect Linkage of Q1 Based on Different Rules

Criterion	Power (%) for suggestive linkage	Power (%) for significant linkage
Two-point linkage analysis		
Single p-value	22.0	2.0
Average of 4 p-values	26.0	6.0
Unequal weight 1 (0.2, 0.3, 0.3, 0.2)	26.0	6.0
Unequal weight 2 (0.1, 0.4, 0.4, 0.1)	22.0	6.0
Unequal weight 3 (0.05, 0.45, 0.45, 0.05)	22.0	8.0
Multipoint linkage analysis		
	28.0	6.0

analyses, the thresholds were 9.5×10^{-5} and 117.5×10^{-5} , respectively for significant and suggestive linkage based on our previously simulated data for a 3-cM marker map [Goldin et al., 1999]. Detection of linkage within 20 cM of the trait locus was considered to be a true positive. Multiple detections within the 14-marker region were counted only once. Since we applied the variance components linkage test, it is possible that the test thresholds computed from the Haseman-Elston test were not exactly correct. However, in this study we were interested in comparing the relative power of the different methods rather than in the absolute power to detect linkage.

RESULTS

Tables II and III show the associated power for all of the inference rules based on all 50 replicates for Q1 and 49 replicates for Q5, applying the thresholds that are shown in Table I.

For Q1, there was a small improvement in power for suggestive linkage detection using some of the regional inference methods compared with a single, two-point p-value criterion, but for Q5, the differences among all methods were too small to draw conclusions. However, there was a consistent slight increase in power for significant linkage detection by regional inference for both Q1 and Q5. The best improvement was seen when the markers in the middle of the region were given more weight (0.1, 0.4, 0.4, 0.1 and 0.05, 0.45, 0.45, 0.05). The results for multipoint analysis were close to those obtained from regional two-point analysis. However, it should be noted that all methods had relatively low power and it is not possible to definitely conclude that some weighting schemes are better than others. The results tend to show a narrow range of power values (especially in the case of Q1). Simulations under a broader range of models are needed to better define the improvement of power using weighted p-values.

DISCUSSION

In this paper, using the simulated data of GAW12, we have demonstrated that regional inference with averaged or weighted p-values increased the power to detect significant linkage in a two-point analysis. Giving substantially more weight to the center p-values (0.05, 0.45, 0.45, 0.05) resulted in the best improvement of the power for detecting the significant linkage of both Q1 and Q5. Some of these regional rules (0.1, 0.4, 0.4, 0.1 and 0.05, 0.45, 0.45, 0.05) were comparable to or even slightly more powerful than the standard multipoint analysis, although the overall power values were

de of this 14-marker region. The program GENEHUNTER was used. This method was chosen for a sib-pair analysis [Pratt et al., 1999] as it is best suited for pedigrees of smaller families for the analysis. Families were included in the analysis. Thus 837 individuals were analyzed. Since our goal was to maximize the number of multipoint lod scores were calculated at each locus from both sides after transforming the lod scores using the SAS program [SAS, 1999]. To account for locus (QTL) variance, we used a chi-square statistic [Liang, 1987].

Linkage was improved by using two-point p-values as a criterion. We have extended this by applying unequal weights to the thresholds for significant linkage (genome scan) and suggestive linkage (genome scan) linkage by applying a linkage of a trait locus to the thresholds for different linkage (using the Haseman-Elston test). Table I shows the thresholds and the results in the middle of a map of 11 markers (the two inner markers).

GAW12 simulated data and compared with that of standard two-point linkage thresholds. For one replicate (#32) and it was applied these different weighting by one marker each solely based on p-values at two-point and multipoint

TABLE III. Power to Detect Linkage of Q1 Based on Different Weighting Schemes (3-cM)

Replicate	Power	
	Significant	Suggestive
1	51.6	86.3
2	52.2	86.4
3	53.5	87.0
4	54.2	86.6

TABLE III. Power to Detect Linkage of Q5 Based on Different Rules

Criterion	Power (%) for suggestive linkage	Power (%) for significant linkage
Two-point linkage analysis		
Single p-value	38.8	10.2
Average of 4 p-values	40.8	12.2
Unequal weight 1 (0.2, 0.3, 0.3, 0.2)	38.8	14.3
Unequal weight 2 (0.1, 0.4, 0.4, 0.1)	38.8	16.3
Unequal weight 3 (0.05, 0.45, 0.45, 0.05)	38.8	16.3
Multipoint linkage analysis	40.8	14.3

low. However, these results were not consistent across the simulations examined so more studies are needed to determine the overall usefulness of unequal weights. Even though multipoint linkage analysis is also considered a smoothing technique, p-values are examined at single points along the genome. In addition, a multipoint analysis requires more computational time as compared with a two-point analysis when dealing with extended families and dense markers, and can be affected by errors in marker order and distances. Thus, our method provides a reasonable alternative to conducting multipoint analyses, especially in initial linkage scans.

Our conclusions are limited by the fact that the power of finding significant linkage was low and only two traits were examined. In a real study, genes that accounted for 21% of the trait variance (as did MG1) or 37% (as did MG5) would be considered to be relatively large effects so the low power of the method we used is of some concern. Clearly the absolute power to detect a trait locus depends on the underlying genetic model, sample size, and test statistic applied so in a real study, some of these variables can be altered. For example, in this particular simulation, the method of variance components analysis as implemented in SOLAR [Almasy and Blangero, 1998] would have probably been a better method of analysis than the GENEHUNTER approach. SOLAR can analyze the entire extended pedigree structures and would have allowed the calculations to be based on a larger sample size resulting in higher power of linkage detection for these traits. Still, the regional methods proposed here are applicable to any linkage statistic.

One limitation of our method is the need to determine the threshold for linkage detection by some empirical method. In this study, we used thresholds computed from null hypothesis simulations using a slightly different data structure and analytic method, so it is possible that those thresholds were not exactly correct. However, the relative ordering of the thresholds should be correct so that the relative powers of the different methods should be the same. Ideally, one should be able to compute empirical thresholds for each specific study by simulating many replicate genome scans under the null hypothesis and computing the same statistics, similar to the approach applied by Gordon et al. [2001]. However, some computational issues need to be resolved in order to conduct such simulations efficiently.

It could also be argued that applying multiple weighting methods will increase the overall type I error of the linkage test. One may be able to develop a compromise weighting from the model-optimal ones that will have relatively high efficiency over the space of plausible genetic models for a trait.

REFER

Almasy L
Hu
Goldin L
P30
Goldin L
to
Gordon D
cor
Hoh J, O
97
Lander E
rep
Pratt SC,
va
SAS Inst
Self SG,
tes
Terwillig
sc

REFERENCES

Almasy L, Blangero J. 1998. Multipoint quantitative trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198-211.

Goldin LR, Chase GA. 1999. Comparison of two linkage inference procedures for genes related to the P300 component of the event related potential. *Genet Epidemiol* 17(Suppl. 1):S163-7.

Goldin LR, Chase GA, Wilson AF. 1999. Regional inference with averaged p-values increases the power to detect linkage. *Genet Epidemiol* 17:157-64.

Gordon D, Hoh J, Finch SJ et al. 2001. Two approaches to consolidating results from genome scans of complex traits: Selection of methods and scan statistics. *Genet Epidemiol*, this volume.

Hoh J, Ott J. 2000. Scan statistics to scan markers for disease susceptibility. *Proc Natl Acad Sci USA* 97:9615-7.

Lander E, Kruglyak L. 1995. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241-7.

Pratt SC, Daly MJ, Kruglyak L. 2000. Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am J Hum Genet* 66:1153-7.

SAS Institute, Inc. 1999. SAS User's Guide, version 8. Cary, NC: SAS Institute, Inc.

Self SG, Liang KY. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 82:605-10.

Terwilliger JD, Shannon WD, Lathrop GM, et al. 1997. True and false positive peaks in genome-wide scans: Applications of length-biased sampling to linkage mapping. *Am J Hum Genet* 61:430-8.