Genetic Epidemiology 14:953–958 (1997)

Bivariate Quantitative Trait Linkage Analysis: Pleiotropy Versus Co-incident Linkages

Laura Almasy, Thomas D. Dyer, and John Blangero

Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas

Power to detect linkage and localization of a major gene were compared in univariate and bivariate variance components linkage analysis of three related quantitative traits in general pedigrees. Although both methods demonstrated adequate power to detect loci of moderate effect, bivariate analysis improved both power and localization for correlated quantitative traits mapping to the same chromosomal region, regardless of whether co-localization was the result of pleiotropy. Additionally, a test of pleiotropy versus co-incident linkage was shown to have adequate power and a low error rate. © 1997 Wiley-Liss, Inc.

Key words: linkage analysis, pleiotropy, quantitative traits, statistical genetics

INTRODUCTION

Increasingly, linkage analyses are being performed with suites of related traits. Linkage between two correlated traits and a single chromosomal region is often taken as further support that a major gene controlling both traits exists in that region. While this is certainly a parsimonious and sensible explanation, that related traits mapping to the same chromosomal region represent pleiotropic effects of a single major gene in that region, there is an alternative explanation which must be explored.

Banks of related genes are sometimes found tightly clustered in a single region, presumably from duplication events involving a single ancestral gene. Examples of gene clusters in the human genome include the leukocyte antigens and the beta-hemoglobins. Thus, when two or more traits show linkage to the same region, even if they are obviously related, it is important to be able to differentiate between pleiotropic effects of a single locus influencing all the traits and separate tightly clustered loci each influencing a single trait.

Address reprint requests to Dr. Laura Almasy, Department of Genetics, Southwest Foundation for Biomedical Research, P.O. Box 760549, San Antonio, TX 78245-0549.

Toward this end, we present bivariate linkage analyses of the GAW 10 quantitative traits Q2, Q4, and Q5 with the chromosome 8 region containing major genes MG2 and MG4. We chose these traits and this region because Q2 and Q4, while highly correlated, share no major genes. The correlation between them is due to shared environmental influences and the co-incident linkage of each of these traits to the chromosome 8 region is a result of separate closely placed genes. (We use co-incident here in the sense of co-occurring, not with the meaning of chance or coincidental.) On the other hand, Q4 and Q5, also highly correlated, provide an example of true pleiotropy as both traits are influenced by a single gene in the region, MG4. Our aims were to demonstrate the pleiotropic effects of MG4 on Q4 and Q5, to exclude pleiotropic effects of a single chromosome 8 gene on Q2 and Q4, and to compare the power to detect linkage and the localization of a major gene or genes in bivariate and univariate multipoint variance component linkage analyses in general pedigrees.

METHODS

In univariate linkage analyses, we used a variance component method designed for use in pedigrees of arbitrary size and complexity [Blangero, 1995], which is an extension of the strategy developed by Amos [1994]. This approach is based on specifying the expected genetic covariances between relatives as a function of the identity-by-descent (IBD) relationships at a given locus. For a simple model in which a major locus and residual polygenes influence a trait, the covariance matrix for a pedigree is given by: $\Omega = \prod\sigma^2_m + 2\Phi\sigma^2_g + I\,\sigma^2_e$ where σ^2_m is the additive genetic variance due to the major locus, and $\prod$ is a matrix whose elements (π_{mij}) provide the probability that individuals *i* and *j* are IBD at a quantitative trait locus (QTL) linked to a marker locus; $\prod$ is a function of the estimated IBD matrix for the genetic marker itself and a matrix of the correlations between the proportions of genes IBD at the marker and at the QTL. In the above equation, σ^2_g is the genetic variance due to residual additive genetic factors, Φ is the kinship matrix, σ^2_e is the variance due to individual-specific environmental effects, and I is an identity matrix. By assuming multivariate normality as a working model within pedigrees, the likelihood of any pedigree can be easily written and numerical procedures can be used to estimate the variance component parameters. Although multivariate normality is assumed, the estimates of effect size are robust to deviations from normality [Amos, 1994; Beaty et al., 1985].

Use of the variance component linkage method requires an estimate of the IBD matrix for each pedigree. For ease of calculation, genotypes of 'dead' individuals were retained and exact IBD matrices were calculated at each marker using a recursive algorithm developed by Davis et al. [1996]. For multipoint analysis in these extended pedigrees, we used an extension of the sib-pair approach of Fulker et al. [1995], using relative pairs' IBD sharing at genotyped loci to estimate sharing at arbitrary points along the chromosome [Almasy et al.,1996]. This method, which has been implemented in the program SOLAR [Blangero and Almasy, 1996], generates estimates of the IBD probabilities at any point on a chromosome using a constrained linear function of observed IBD probabilities at genetic markers at known locations within the region. Using IBD probabilities for all 50 chromosome 8 markers simultaneously, IBD sharing was estimated every 1 cM along the chromosome for all relative pairs.

Using the variance component model, we tested the null hypothesis that σ^2_m equals zero (no linkage) by comparing the likelihood of this restricted model with that of a model in which σ^2_m is estimated. The difference between the two $\log_{10}$ likelihoods produces a lod score that is the equivalent of the classical lod score of linkage analysis. Multipoint univariate analyses were performed with quantitative traits Q2, Q4, and Q5 every 2 cM along chromosome 8 for 50 replicates of the data set. Within the 10 cM on either side of MG2 and MG4, which were located between 51 and 52 cM, multipoint linkage analysis was performed every 1 cM for all 200 replicates. Two-point univariate analyses were performed at the loci flanking MG2 and MG4, D8G26 and D8G27, for Q2, Q4, and Q5 in all 200 replicates. Covariates were included in the analyses on a trait-specific basis, given knowledge of the generating models. The environmental factor (EF) was used in all analyses of Q2, but not Q4 or Q5.

Extension to multivariate analysis is straightforward. The analogous covariance matrix for the pedigree is: $\Omega = M \otimes \prod + G \otimes 2\Phi + E \otimes I_n$, where $\otimes$ is the Kronecker product operator, M is the $t \times t$ additive genetic covariance matrix due to the QTL, G is the residual additive genetic covariance matrix, and E is the environmental covariance matrix. Testing for linkage involves the examination of the diagonal elements of M for deviation from zero. For the bivariate case considered here, we obtain trait-specific estimates of σ^2_m, σ^2_g, and σ^2_e as well as the three associated correlations ρ_m, ρ_g, and ρ_e. ρ_g and ρ_e represent the extent of shared residual additive genetic and environmental influences on the traits. ρ_m is a measure of shared major gene effects near the region for which linkage is being assessed. Using this model, we tested the null hypothesis that σ^2_m equals zero for both traits by comparing the likelihood of this restricted model to that of a model in which σ^2_m was estimated for the traits. The lod score produced by this method is <u>not</u> equivalent to the classical lod score of linkage analysis, as there are two degrees of freedom involved. Both univariate and bivariate linkage analysis routines have been implemented in the program SOLAR [Blangero and Almasy, 1996] using estimation procedures from FISHER [Lange et al., 1988]. Multipoint bivariate linkage analyses were performed on all 200 replicates for Q2 with Q4 and Q4 with Q5. Lod scores were calculated every 1 cM between 40 cM and 60 cM on chromosome 8.

To test pleiotropy versus co-incident linkage, likelihoods for the linkage model in which ρ_m was estimated were compared to models in which ρ_m was constrained to 0 (no shared major gene effects in the region, i.e.: co-incident linkage) and constrained to 1 (complete pleiotropy). This comparison was made at the location showing the maximum lod score for the replicate. In the case of ρ_m constrained to 0, the difference between these likelihoods is distributed as a chi-square with 1 degree of freedom. When ρ_m is constrained to 1, a boundary, the difference in likelihoods is distributed as a 1/2:1/2 mixture of χ^2_1 and a point mass at 0 [Self and Liang, 1987].

RESULTS

Power to detect MG2 and MG4 at a lod of 3.0 for any given replicate in the multipoint univariate analyses was excellent for Q4, mediocre for Q2, and poor for Q5 (Table I). At less stringent, suggestive significance levels (lod of 1.5 or 2.0), power in the multipoint univariate analyses was acceptable for Q2 but still low to moderate for Q5. Mean two-point lod scores for D8G26 and D8G27 over the 200 replicates were 60 to 70% of those seen in multipoint analysis and the percentage of Q2 and Q5 maximum lod scores

over 3.0 was approximately halved in the two-point analyses as compared to multipoint. Localization of QTLs in the multipoint univariate analyses was excellent throughout (Table II), with even Q5 giving a maximum lod score within 10 cM on either side of MG4 in nearly 90% of replicates. The range of locations of maximum lod scores was quite large for the Q2 and Q5 analyses. However, the outlying values of these ranges occurred in replicates with low maximum lod scores. In replicates with maximum lod scores of at least 3.0, the locations of these maxima ranged between 39 and 65 cM.

In assessing lod scores obtained in the bivariate and univariate analyses, it is important to remember that the bivariate analyses generally contain 2 degrees of freedom, while the univariate analyses have only 1. Thus lod scores are not directly comparable between the two analyses. However, we may associate each lod score with an appropriate p-value. A bivariate lod score of 3.5 is equivalent in p-value to a univariate lod score of 3.0. More generally, the p-value for the 2 degree of freedom lod can be obtained by transformation of the test statistic to 2*ln(10)*lod which is distributed approximately as a mixture of χ^2 distributions with the mixing frequencies being (1/2) for a one-degree of freedom χ^2 and (1/4) for a 2-degree of freedom χ^2 [Self and Liang, 1987].

Bivariate analyses (Table III) dramatically improved the power to detect linkage of Q5 to the MG4 region, with 85.5% of replicates having a lod score over 3.5 in the Q4Q5 analyses. Unexpectedly, detection of Q2 was helped by the inclusion of Q4 and vice versa. The maximum lod scores in the Q2Q4 bivariate analyses ranged from a low of 5.97 to a high of 21.87. Localization was also improved with 97.0% of replicates now showing a maximum lod score within 5 cM on either side of MG2 and MG4 for the Q2Q4 analysis, and 88.5% for the Q4Q5 analysis.

Tests of pleiotropy versus co-incident linkage were also highly successful. Likelihoods for the model in which ρ_m was estimated were compared to likelihoods for the models in which ρ_m was constrained to 0 or to 1. Power to reject the inappropriate model was excellent in both cases, with over 90% of replicates rejecting at the 0.05 level and approximately 70% of replicates still rejecting at the more stringent 0.01 level (Table IV). Conversely, the appropriate constrained model was rarely wrongly rejected. In the case of Q4 and Q5, there were no false negatives; the true model of complete pleiotropy was not rejected in any replicate. In the Q2 and Q4 analysis, there was only one incorrect rejection of the co-incident linkage model at a p-value of 0.05.

TABLE I. Power to Detect QTLs in Two-point and Multipoint Univariate Analyses in 200 Replicates of the Data Set

	Multipoint				D8G26		D8G27	
Trait	%≥1.5	%≥2.0	%≥3.0	Mean	Mean	%≥3.0	Mean	%≥3.0
Q2	83.5	70.0	43.5	2.92	1.85	20.5	1.84	18.0
Q4	98.0	96.5	87.0	5.57	3.24	49.5	3.93	68.5
Q5	56.0	40.5	15.5	1.82	0.97	3.0	1.22	7.0

TABLE II. Location of Maximum Lod Scores in Univariate Analyses, in cM from p-ter, and Proportion Showing Localization within 5 cM or 10 cM on Either Side of MG2 and MG4

	All 200 reps				Reps w/ max lod ≥ 3	
Trait	Mean	Range	± 5 cM	± 10 cM	± 5 cM	± 10 cM
Q2	51 cM	34-92	79.0%	94.5%	83.0%	97.5%
Q4	52 cM	41-64	90.0%	99.0%	93.0%	99.5%
Q5	51 cM	2-80	68.0%	88.0%	77.5%	93.5%

DISCUSSION

Multipoint analysis greatly improved power over two-point analysis for these traits, more than doubling the proportion of maximum lod scores over 3.0 for Q2 and Q5. Power to detect linkage of Q4 to MG4, which accounted for 28% of the phenotypic variance in this trait, was excellent, with over 85% of replicates showing a maximum lod score over 3.0. Power in the multipoint univariate analyses to detect linkage with the other traits was acceptable, given that MG2 accounted for only 21% of the variance in Q2 and MG4 only 14% of the variance in Q5. Although power to detect linkage for Q2 and Q5 on their own was poor, with fewer than 45% of replicates showing maximum lod scores over 3.0, Q2 and Q5 did show sufficiently high univariate lod scores to suggest their inclusion in a bivariate analysis. Localization was excellent in the univariate analyses, even when power was lacking, with over 85% of replicates showing a maximum lod score within 10 cM on either side of the major genes for all three traits.

Bivariate analysis improved localization and power, with over 85% of replicates showing lod scores of at least 3.5 for linkage of Q5 (analyzed with Q4) to the MG4 region and over 85% of maximum lod scores falling within 5 cM on either side of the major gene. This increase in power is promising as it would be very difficult to detect linkage to this region with Q5 alone. There was also a vast improvement in lod scores in the Q2Q4 bivariate analyses, which was somewhat unexpected, but is not inexplicable. When pleiotropy is present, the information provided by a trait with weak major gene influences in the region (Q5) is supplemented by the stronger signal of the second trait (Q4). However, the weaker trait provides no additional information over the stronger trait and thus the power of the bivariate analysis is no more than the power of univariate analysis in the stronger trait. When major locus pleiotropy is not present (Q2 and Q4), but the traits are correlated through another mechanism, each trait provides additional new information about the other, greatly increasing power over what was seen in either univariate analysis. Preliminary analysis of the first 50 replicates of the data set indicates that a similar effect is seen in bivariate analyses of Q2 and Q5. Averaging over the 50 replicates, the maximum bivariate lod score for Q2 and Q5 was 5.53 at 51 cM, while the maxima in the averaged univariate analyses were only 2.60 and 1.55 for Q2 and Q5 respectively.

Power to differentiate between pleiotropy and coincident linkage was excellent in all cases. The incorrect model was rejected a substantial portion of the time, depending

TABLE III. Power Over 200 Replicates to Detect MG2 and MG4 and Location of Maximum Lod Scores in Bivariate Analyses, with ρ_m Unconstrained

	Maximum lod scores		Localization	
Traits	Mean	% ≥ 3.5	Mean	± 5 cM
Q2 Q4	13.72	100	52 cM	97.0%
Q4 Q5	6.02	85.5	52 cM	88.5%

TABLE IV. Power to Reject the Incorrect Bivariate Model in 200 Replicates of the Data Set

p-value	Q2Q4 (rejecting ρ_m=1)	Q4Q5 (rejecting ρ_m=0)
0.05	93.0%	94.0%
0.02	83.0%	86.5%
0.01	69.5%	76.5%
0.001	38.5%	34.5%

on how stringent a p-value was employed. On the other hand, false negatives were rare and the correct model was rejected only once over both bivariate analyses combined, an error rate of 0.25%. The nature of the true model, pleiotropy or co-incident linkage, did not seem to affect either power to reject the incorrect model or false rejection of the true model. These analyses suggest that a chi-square p-value of 0.05 is a sufficiently conservative cut-off for the rejection of either pleiotropy or co-incident linkage.

There are, however, some situations which would complicate the interpretation of pleiotropy versus co-incident linkage in these bivariate analyses. Had there been linkage disequilibrium between MG2 and MG4, Q2 and Q4 would have shown some genetic correlation, in proportion to the strength of the disequilibrium, which would have reduced power to reject pleiotropy. Conversely, had there been epistatic or gene by environment interactions affecting the action of MG4 on one of Q4 or Q5, but not both, the genetic correlation between Q4 and Q5 would have appeared incomplete, reducing the power to reject co-incident linkage. Thus intermediate, indeterminate results are possible in these bivariate analyses and results indicating strong pleiotropy must still be scrutinized for confounding linkage disequilibrium within the region.

ACKNOWLEDGMENTS

The development of the statistical methods used in this work was supported by National Institutes of Health grants DK44297, HL45522, HL28972, and GM31575.

REFERENCES

Almasy L, Hixson JE, VandeBerg JL, MacCluer JW, Blangero J (1996): Multipoint quantitative trait linkage analysis in general pedigrees. Am J Hum Genet 59:A173.

Amos CI (1994): Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 54:535-543.

Beaty TH, Self SG, Liang KY, Connolly MA, Chase GA, Kwiterovich PO (1985): Use of robust variance components models to analyse triglyceride data in families. Ann Hum Genet 49:315-328.

Blangero J (1995): Multivariate oligogenic linkage analysis of quantitative traits in general pedigrees. Am J Hum Genet 57:A11.

Blangero J, Almasy LA (1996): SOLAR: Sequential Oligogenic Linkage Analysis Routines. Technical Notes No. 6. Population Genetics Laboratory, Southwest Foundation for Biomedical Research, San Antonio, TX.

Davis S, Schroeder M, Goldin LR, Weeks DE (1996): Nonparametric simulation-based statistics for detecting linkage in general pedigrees. Am J Hum Genet 58:867-880.

Fulker DW, Cherny SS, Cardon LR (1995): Multipoint interval mapping of quantitative trait loci, using sib pairs. Am J Hum Genet 56:1224-1233.

Lange K, Weeks D, Boehnke M (1988): Programs for pedigree analysis: Mendel, Fisher, and dGene. Genet Epidemiol 5:471-472.

Self SG, Liang K-Y (1987): Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605-610.