

Detection of Major Genes Underlying Several Quantitative Traits Associated with a Common Disease Using Different Ascertainment Schemes

Sudha Iyengar, Francesc Calafell, and Kenneth K. Kidd

Department of Genetics, Yale University School of Medicine, New Haven, Connecticut

Using the Problem 2A data sets of GAW10, we assessed the power of four ascertainment schemes to localize major genes underlying a disease trait; the schemes were based on disease or quantitative trait status of nuclear families. MAPMAKER/SIBS was used to perform sib-pair analysis for all four data sets using marker data from three chromosomes, 4, 5 and 8. Each scheme varied in power to identify major genes underlying the quantitative traits depending on the genetic architecture of the data set. Three different methods, Haseman-Elston quantitative trait locus (QTL) regression analysis, maximum likelihood variance estimation and a non-parametric method, were used to assess the strength of linkage in all four data sets. False positive mappings localizing to the same region of the genome, verifiable across all three methods did not occur. Two major genes, MG1 and MG2, were successfully assigned to chromosomes 5 and 8, respectively, by at least one of the ascertainment schemes. MG1 was localized under two schemes, selection of families with exactly two affected sibs and selection of families with two sibs who had extremely discordant values for Q1. Additional weak evidence of the location of MG1 was also obtained under the other two ascertainment schemes. MG2 could not be detected by analyzing data sets ascertained either by affected sib pairs or by sib pairs with extremely discordant values for Q1. In the data set ascertained by a third strategy, selection of families with sib pairs extremely discordant for Q2, MG2 could be mapped to chromosome 8. A random ascertainment scheme yielded a data set in which we could find weak evidence for MG1 and no evidence for MG2. Thus our ability to detect major genes underlying the QTL depended on several factors which included the ascertainment scheme, the population allele frequencies, linkage and epistasis.
© 1997 Wiley-Liss, Inc.

Address reprint requests to Dr. Sudha Iyengar, I-347 SHM, Department of Genetics, Yale University School of Medicine, 333 Cedar St., New Haven, CT 06520-8005.

© 1997 Wiley-Liss, Inc.

Key words: complex disease, extremely discordant sib pairs, quantitative trait loci

INTRODUCTION

For most common diseases currently under investigation, data are available on several quantitative traits associated with the disease. Affection status, in at least a few such diseases, is defined by setting a threshold value for only one of the measured quantitative traits. Families with two or more affected individuals have been sampled from the population to localize major genes underlying all the quantitative traits associated with the disease. In the case of an oligogenic disease, the ascertainment strategy outlined above enriches the sample for affecteds and hence increases the frequency of the allele(s) which is strongly associated with both disease and high values of the diagnostic quantitative trait in the sample. The effect of such ascertainment strategies on other genes that have a less immediate effect, but are still important in the etiology of the disease, has not been evaluated. Indeed, it may not always be possible to identify all genes segregating for and involved with a particular disease in a given sample. Real problems are encountered when there is epistatic interaction between loci or when alternative ascertainment schemes identify different loci for quantitative traits associated with the disease. The analyses presented in this report address several questions related to the effects of various ascertainment strategies on quantitative traits associated with a common disease. A caveat of this analysis is that its results cannot be readily extrapolated to all populations, since differences in gene frequencies of the major genes underlying the quantitative traits are expected between populations and data were provided only on a single population. Furthermore, the epistatic interactions between loci make the relative effects of loci frequency dependent.

METHODS

Description of Ascertainment Schemes. Four sets of 500 nuclear families, each set obtained through one of four different ascertainment schemes, were selected for analysis from the GAW10 Problem 2A data set. Selection criteria were based on either affection status, or extreme discordance for Q1 or Q2 among sib pairs [Eaves and Meyer, 1994; Risch and Zhang, 1995], or random selection. Under the affected-pair ascertainment scheme families with exactly two affected sibs were chosen. Under the extremely discordant sib-pair scheme families were selected in which at least one sib was in the lower decile of the quantitative trait and at least one other sib was in the upper decile of the same quantitative trait; the QTL were corrected for age, gender and the environmental factor (EF) before the selection criteria were applied. This scheme was applied to two traits, either Q1 or Q2. In the last ascertainment scheme, families were selected solely on sibship size and regardless of affection status, Q1 values or Q2 values. Sibship size in the randomly ascertained set was chosen to match the distribution of sibship sizes in the families selected for extremely discordant values of Q1. In all selection schemes families were sampled across replicates, starting at nuclear family one, until 500 families meeting the selection criteria were obtained. Average sibship size ranged from 3.748 to 3.782 across ascertainment schemes.

Determination of the Effects of Covariates on the Quantitative Traits. Measures of normality were evaluated for Q1-Q5 to determine if data transformations

were necessary. Correlations between Q1-Q5, age and EF were examined by estimating Pearson correlation coefficients between all the variables in each sampled data set. A regression analysis was performed to determine the effect of covariates, age, gender, and EF on Q1-Q5 in each sampled data set. Interaction terms were not considered. Data on Q1-Q5 were adjusted for significant covariates and the residuals were tested for deviations from normality. All analyses were performed using the statistical package SPSS. (SPSS Inc., Chicago, Ill.)

Multipoint Sib-pair Analysis of the Quantitative Traits. Two parametric methods and one nonparametric method for linkage analysis of QTLs in sib pairs, implemented in the program MAPMAKER/SIBS [Kruglyak and Lander, 1995], were utilized to scan three chromosomes, 5, 8 and 4. These chromosomes were chosen knowing beforehand that the three major genes that affect Q1 map on them. Prior knowledge of the generating model enabled us to assess the relative power of the different ascertainment schemes. Linkage analysis was performed on traits Q1 to Q5. Some of the chromosomes analyzed did not contain major genes affecting some of the traits. Such cases (e.g., chromosome 5 and Q5) served as controls. Before sib-pair analysis was attempted, the effects of significant covariates on quantitative traits were removed by replacing the original values of quantitative traits with their residuals in a regression. No correction was made for ascertainment bias.

RESULTS

The data for 500 nuclear families under each ascertainment scheme are summarized (Table I). Variability in total sample size was small, ranging from 1,874 to 1,891 individuals across selection schemes. However, the number of affected individuals sampled varied greatly (1,000-115). An excess of affected females was observed in the data sets selected for two affected individuals and for sibs with extreme values for Q1. Values in parentheses in Table I represent the percentage of affected females or males among all females or males, respectively, and indicate the magnitude of this excess. Selecting for two affected sibs increased the probability of obtaining individuals with high

TABLE I. Description of 500 Families Analyzed Using Four Different Ascertainment Schemes

	Ascertainment schemes			
	Two sibs affected	Extremely disc. for Q1	Extremely disc. for Q2	Random
No. of individuals	1,891	1,879	1,874	1879
No. affected	1,000	322	134	115
No. affected per sibship	2	0.64	0.27	0.23
Mean age (sd)	53.49 (14.42)	43.16 (15.08)	43.71 (14.89)	44.85 (14.99)
Percent male	33.7	44.7	47.9	46.2
No. of affected males	196 (30.7%)	55 (6.5%)	22 (2.4%)	18 (2.1%)
No. of affected females	804 (64.2%)	267 (25.7%)	112 (11.5%)	97 (11.5%)
No. families with 0 aff	-	219	397	414
No. families with 1 aff	-	246	76	64
No. families with 2 aff	500	31	23	16
No. families with 3 aff	-	3	4	5
No. families with 4 aff	-	1	-	1

values for Q1, Q2 and Q3, which is reflected as a nonsignificant increase in mean values of these traits (data not shown). As expected, sampling for sibs with extreme phenotypic differences for either Q1 or Q2 increased the variance but not the mean values for these traits.

The data needed no transformation prior to correction for covariates as preliminary analysis of the quantitative trait data revealed that normality assumptions were not violated. Regression analyses were performed on the untransformed data set to determine significant covariates. Three covariates, age, gender and EF, were identified in each data set, and all further analyses were performed on the residuals of the regression among the quantitative traits and the covariates. Since three of the data sets were not randomly sampled from the population and a sampling bias is expected, Pearson correlation coefficients were computed comparing residuals from the population data to those obtained from selected family data to determine the extent of the bias. Correlation coefficients ranging from 0.989 to 0.999 were observed, indicating that sampling bias, if present, is minimal.

Multipoint sib-pair linkage analyses were performed on all five quantitative traits, using MAPMAKER/SIBS. This program uses a nonparametric method for linkage analysis which determines the degree of identity by descent (IBD) sharing among sibs at each point in the genome. Lod score or Z-score statistics from three analyses (Haseman-Elston QTL regression analysis, maximum likelihood QTL variance estimation, and a nonparametric method) incorporated into MAPMAKER/SIBS were then used to evaluate the strength of linkage. Regardless of the selection scheme, care was taken to incorporate the phenotype and genotype information from all pairs of sibs, using the "all possible pairs option" in MAPMAKER/SIBS. Over 80% of the families analyzed had at least three sibs. It was necessary to use data on all sibs to ensure that use of a particular selection scheme did not violate the underlying distributional assumption, leading to false positive linkage [Risch and Zhang, 1995]. Although we examined only chromosomes 4, 5 and 8 under the selected ascertainment schemes, the sib-pair analyses were performed for all quantitative traits, Q1- Q5. The results regarding the major genes associated with Q1 and Q2 are described in Table II; Q3 to Q5 served as controls, and save for Q4 (see below), no significant linkage was detected.

False positives were expected given the large number of tests performed. Several regions with lod scores or Z-scores ranging from 0.2 to 1 contributing to the "background noise" were observed. Major genes were considered "detected" when significantly positive lod scores of 4 or a Z-score of 4.1 were observed. Results were verified by analyzing the data simultaneously by Haseman-Elston regression analysis, Maximum Likelihood variance estimation and nonparametric QTL analysis. In all cases, if a gene was detected by one method, large positive lod or Z-scores were also obtained by the other

TABLE II. Detection of Major Genes Associated with Q1 and Q2 Using MAPMAKER/SIBS. In Parentheses, Pointwise p-Values Computed from Lod Scores According to Lander and Kruglyak [1995]

Ascertainment scheme	Chromosomes and traits examined		
	Chr. 5 (MG1), Q1	Chr. 8 (MG2), Q2	Chr. 4 (MG3), Q3
Two affected sibs	detected ($5 \cdot 10^{-12}$)	not detected	not detected
Extremely disc. Q1	detected (10^{-24})	not detected	marginal ($9 \cdot 10^{-4}$)
Extremely disc. Q2	marginal ($6 \cdot 10^{-3}$)	detected ($1 \cdot 10^{-8}$)	not detected
Random	marginal ($3 \cdot 10^{-3}$)	not detected	not detected

two methods. In four cases positive lod or Z-scores below 4 or 4.1 but above background were observed. These were suggestive of linkage and were labeled as “marginal” if they could be verified by all three analyses. We observed a few peaks greater than baseline (a lod or Z-score above 1) in some tests. However, false positives verified in all three analyses were not observed.

DISCUSSION

Using four sets of 500 families sampled from the population under various schemes we were able to detect the effects of MG1 (chromosome 5, D5G14-D5G15; Fig. 1) and MG2 (chromosome 8, D8G26-D8G27) on quantitative traits Q1 and Q2. The magnitude of lod or Z-scores was largest using sib pairs selected for extreme phenotypic values indicating that this was the best method for localizing major genes underlying QTLs. For example, using the Haseman-Elston regression analysis and a sample of 500 families having sibs with extreme values for Q1, a lod score of over 22 was obtained as evidence for MG1 being on chromosome 5, between markers D5G14 and D5G15. We found that such families were approximately as frequent in the data set (approx. 1 in 20) as those having two affected sibs and that a sample of 100 such families would have been sufficient to obtain evidence for linkage. A limitation of this sampling strategy is that it overestimates the variance contribution of the major gene to the QTL and requires that other methods be used in conjunction to determine the true genetic variance [Risch and Zhang, 1995].

Fig 1. Lod-score curve, estimated by the Haseman-Elston regression method, for Q1 on chromosome 5. Five hundred families selected for sibs with extremely discordant Q1 values were used in the analysis. MG1 maps between loci D5G15 and D5G16.

Given that the direct contribution of MG2 to the variance of Q1 (0.48%) was minimal we could not detect its effect on Q1 [Eaves, 1994]. Combined segregation and linkage analyses were not performed to test the hypothesis that the interaction between MG1 and MG2 explained an additional 13.4% of the genetic variance in Q1. We were unable to observe the effects of MG2 on Q2 in any data set but one selected for extreme phenotypic differences in Q2 between a pair of sibs. Evidence for segregation of MG3 was obtained from the “marginal” lod or Z-scores on chromosome 4 for Q3 using the data set with extreme phenotypic values for Q1 (Table II). Examination of the most common genotype sampled under the various selection schemes provides an explanation for these findings. Sampling using two affected sibs or sibs with extreme phenotypic values for Q1 leads to an excess of individuals with a high value for Q1 in these data sets. The major proportion of the individuals sampled for high values for Q1 will be those who have the most common allele at MG1 associated with disease and those with an MG3 genotype which directly influences disease susceptibility. In the data sets selected for two affected sibs or extremely discordant values of Q1 most of the selected individuals will be females, as major genes MG1 and MG2 contribute greater genetic variance to both Q1 and Q3 in females than in males. These findings explain the preponderance of affected females in these two data sets.

While testing for linkage between Q2 and MG2 we obtained provisional evidence for linkage of markers on chromosome 8 with Q4 in three data sets. Given the generating model these results are not surprising as MG2 and MG4 are in proximity to each other. The strongest evidence for linkage between a gene on chromosome 8 and MG4 was obtained from the data set where two sibs had extreme phenotypic differences for Q1, indicating that MG4 segregated in at least some of the families in this data set.

These analyses show that selection of samples utilizing various selection schemes skews segregation of QTLs not directly sampled for. The bias is dependent on factors such as population gene frequencies, the contribution of the major gene to the genetic variance of the QTL, and linkage.

ACKNOWLEDGMENTS

We express our gratitude to Dr. Hong Yu Zhao for helpful discussions. FC was supported by a post-doctoral fellowship awarded by CIRIT (Catalonia, Spain). This work was supported by a USPHS grant MH39239 to KKK.

REFERENCES

- Eaves L (1994): Effect of genetic architecture on the power of human linkage studies to resolve the contribution of quantitative trait loci. *Heredity* 72:175-192.
- Eaves L, Meyer J (1994): Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. *Behav Genet* 24:443-455.
- Kruglyak L, Lander ES (1995): Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439-454.
- Lander E, Kruglyak L (1995): Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241-247.
- Risch N, Zhang H (1995). Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 268:1584-1589.