False Positive Rates in a Genomic Screen for Complex Quantitative Traits

William K. Scott, **Marcy C. Speer**, **Margaret A. Pericak-Vance**, **Carol S. Haynes**, **Suzanne M. Leal**, and **Linda M. Brzustowicz**

Section of Medical Genetics (W.K.S., M.C.S., C.S.H., M.A.P-V.), Duke University Medical Center, Durham, North Carolina; Laboratory of Statistical Genetics (S.M.L.), The Rockefeller University, New York, New York; Center for Molecular and Behavioral Neuroscience, Rutgers University, and Department of Psychiatry (L.M.B.), University of Medicine and Dentistry of New Jersey, Newark, New Jersey

Abstract

We conducted a genomic screen for genes associated with Q1, Q2, and Q3 in 239 nuclear pedigrees from replicate 115, Problem Set 2A. We compared false positive (FP) and true positive (TP) rates for three significance levels and two map densities. Using the 2 cM genetic map and $\alpha = 0.05$ produced the most FP but detected the greatest number of major genes. Following up only 31 plateaus (two or more adjacent markers with significant results) from the 2 cM screen eliminated some FP, but failed to detect MG3 for Q3. Multipoint analysis reduced the number of priority regions from 31 to seven; only two of these regions were TP. Replication of the two-point analysis of plateau markers in replicate 80 detected all of the genes associated with Q1 and Q2, but not Q3. Multipoint analysis in replicate 80 failed to replicate any genes associated with Q1, Q2, or Q3, but "replicated" two FP regions. While FP may be reduced by decreasing map density, considering only plateaus for follow up and decreasing significance levels, such adjustments may also fail to detect weak TP. Multipoint analysis and replication in independent data sets may not be reliable methods of distinguishing FP from TP.

Keywords

complex disease; linkage analysis; sib-pair analysis

INTRODUCTION

Genomic screening for genes underlying complex traits has quickly become the method of choice for human geneticists. The advent of microsatellite DNA markers and sophisticated computerized algorithms for detecting linkage has facilitated this development, and current capabilities make a complete screen of the genome feasible in a short period of time. Current genomic screens often use a hierarchical design, where a map of widely spaced markers (e.g., 10 cM) is used to identify regions suggestively linked ($p < 0.05$) to the disease trait.

These results are then followed up by genotyping more markers in the region and using multipoint analysis. A primary concern with genomic screens is the presentation and interpretation of results, since the large number of comparisons made in the same data set often leads to FP results.

Several approaches have been proposed for the interpretation of genomic screen results. Thomson [1994] suggested that in studies of complex traits, all markers with nominal p-values less than 0.05 should be followed up so weak linkages would not be missed. Linkage would then be suggested only when the original association was confirmed in multiple independent data sets or if very strong evidence (p < 0.001) existed in one data set. However, difficulties in the replication of linkages in complex traits have been noted [Suarez et al., 1994] and replication may not be reliable for confirming TP or eliminating FP results. Lander and Kruglyak [1995] agree that while markers with nominal p-values of 0.05 should be examined in more detail, publication of results should take place only when a genome-wide significance level of 7×10^{-4} is reached. This suggestion prompted concerns whether such corrections are appropriate in all cases and whether publication should rest on attainment of a particular significance level. Another method for controlling the number of FP followed up in a genomic screen is by only following up regions, called plateaus, that suggest linkage across an interval. These plateaus may be defined as a certain number of markers out of a set (e.g., 3 of 5) that have significant (<0.05) nominal p-values [Terwilliger JD, personal communication, Goldin and Chase, this volume]. The objective of this study was to conduct a genomic screen using a variety of significance levels, map densities, and methods of follow up to determine what combination of factors minimizes FP while retaining the power to detect true linkage.

METHODS

Two replicates (115, 80) of 239 nuclear pedigrees from Problem Set 2A were randomly chosen for analysis. All individuals from replicate 115 were used to determine if age and sex were associated with Q1, Q2 and Q3. Linear regression models were constructed with each quantitative trait as the outcome and age and sex as independent variables. The environmental factor (EF) was not included in the analysis due to errors in EF values in the data set. Sex was significantly associated with Q1, Q2 and Q3, and age was associated with Q1 (data not shown). Although age was associated only with Q1, the effect of both variables on each trait was controlled in the subsequent genomic screen.

The Haseman-Elston sib-pair method of linkage analysis, implemented in the S.A.G.E. SIBPAL program [S.A.G.E., 1994], was used for two-point analysis of linkage between microsatellite markers and the three traits while controlling for the covariates. All 367 markers were used, constituting an approximate 2 cM genomic screen. Additionally, a 10 cM screen was constructed from every fifth marker on each chromosome in the 2 cM screen. The number of markers with statistically significant evidence for linkage to the traits was determined at three significance levels: $\alpha = 0.05$, 0.01, and 0.001. Results were compared to the generating model and classified as TP, FP, true negative (TN), or false negative (FN). Markers located within 10 cM of a major gene underlying a trait (MG1, MG2, and MG3 for Q1; MG2 for Q2, MG3 for Q3) with statistically significant two-point results were

considered TP. Statistically significant results at markers located more than 10 cM from a major gene underlying the trait were considered FP. Results were classified FN if linkage was not detected at markers within 10 cM of the major genes associated with the trait. The FP rate was calculated by the formula FP/(FP+TN), and the FN rate was calculated by the formula FN/(FN+TP). Major genes were considered detected if at least one marker within 10 cM showed statistically significant evidence of linkage.

To minimize the number of positive results to follow up, we looked for plateaus in the results from the 2 cM genomic screen. Plateaus were defined as two or more adjacent markers with statistically significant results. These plateaus and one flanking marker on each side were then followed up in multipoint analysis using MAPMAKER/SIBS [Kruglyak and Lander, 1995]. Multiple linear regression was used to create values for Q1, Q2, and Q3 adjusted for age and sex to approximate trait values used in the two-point analysis. These adjusted values were used in the multipoint analysis. The Haldane mapping function and both the independent sib-pair and all sib-pair options were used for the multipoint analysis.

A final attempt to distinguish TP from FP was to repeat the analysis in replicate 80. All markers used in the replicate 115 multipoint analysis (plateau markers and flankers) were used in two-point analysis of replicate 80. Only markers with positive evidence for linkage in the replicate 115 multipoint analysis were used in the multipoint analysis of replicate 80.

RESULTS

In both the 2 cM and 10 cM genomic screens of replicate 115, the overall FP rate exceeded the significance levels due to multiple comparisons (Table I). While reducing the significance level from $\alpha = 0.05$ to $\alpha = 0.01$ cut the FP rate for the 2 cM screen in half, it also increased the FN rate and hampered the ability to detect linkage for Q2. Using a significance level as stringent as $\alpha = 0.001$ impaired the ability to detect linkage for all three traits; only MG2 is detected for Q1 and no genes are found for Q2 or Q3. Comparing the results of the 2 cM screen to the 10 cM screen, the 10 cM screen has fewer FP but fails to detect genes associated with Q2 (MG2) and Q3 (MG3). Only two genes (MG1 and MG2), both associated with Q1, are detected at $\alpha = 0.05$ in the 10 cM screen.

Although the FP rate is high, using the 2 cM screen and assessing statistical significance at $\alpha = 0.05$ has the most success in detecting major genes. Therefore, plateau analysis of the two-point results for the 2 cM screen at $\alpha = 0.05$ was used to reduce the number of positive results to follow up (Table II). Several FP markers for each trait were eliminated from further consideration using this method. In all, 31 priority regions were identified for follow up: 14 for Q1, seven for Q2, and 10 for Q3. Four of the 31 regions were TP: three priority regions for Q1 and one priority region for Q2 were associated with the major genes underlying the traits.

For replicate 115, multipoint analysis found that six of the 31 priority regions had maximum multipoint lod scores > 1 (p < 0.04). Four regions were associated with Q1 and two were associated with Q3. Only two of these regions, both associated with Q1, were TP (Table III). The maximum multipoint lod score of 1.73 (p = 0.0048) at D5G14 was the strongest

multipoint lod from any of the regions in the follow up analysis, and occurs 0.8 cM from the true location of MG1. The second TP was a region with a maximum multipoint lod score of 1.20 (p = 0.0188) at D4G20. This region was located 9.7 cM from MG3. However, the strength of these TP results did not meaningfully distinguish them from the other five FP results.

The two-point and multipoint analysis was also performed in replicate 80, using the markers identified as priority regions in replicate 115. Fifteen of 69 markers were replicated for Q1. Of those 15, seven were TP, and markers associated with all three major genes were replicated. For Q2, only four of 43 markers were replicated, only one of which was a TP. No markers out of 39 were replicated for Q3. Multipoint analysis of replicate 80 found two of the 31 priority regions had multipoint lod scores > 1 (D1G1-D1G9, p = 0.04; D2G38-D2G41, p = 0.04), both of which were FP.

DISCUSSION

The results of this study indicate that using a liberal initial significance level and a more tightly spaced map of markers gives the greatest power to detect major genes associated with Q1, Q2 and Q3. Despite the increase in FP, using a significance level of $\alpha = 0.05$ minimizes FN in both the 2 cM and 10 cM screens. When comparing the two map densities, the 10 cM screen has much less ability to detect major genes than the 2 cM screen. Therefore, a hierarchical design where priority regions are identified by a 10 cM screen and followed up would fail to detect most of the major genes associated with Q1, Q2, and Q3. This lack of power to detect effects would suggest that screening with a more dense genetic map (although perhaps not as dense as 2 cM) is a better approach than using a two-stage design with a 10 cM screen.

A reduction in the number of FP results to follow up is achieved by using plateau analysis, but at the expense of detecting the major gene associated with Q3. Length of the plateau is not an effective way of separating TP and FP; for each trait, the largest plateau is not associated with a major gene. As well, if the definition of plateau would have been more stringent, such as three or more adjacent markers with significant results, several regions containing genes would not have been followed up.

Multipoint analysis reduced the number of priority regions from 31 to six, but only two regions were TP. The maximum multipoint lod scores for the region were only 1.73 (p < 0.005) and 1.20 (p < 0.02). These results, while TP, do not meet the significance criteria of 7×10^{-4} suggested by Lander and Kruglyak [1995], or even the traditional significance level of a lod score of 3. However, the data set screened for linkage was not ascertained on the basis of the traits being studied; this may have limited the power of the data set to detect linkage. Therefore, while the result could be reported as suggestive of linkage, replication would be necessary to confirm that a genetic effect exists at these loci.

However, replication of results in genetic studies of complex traits is often difficult. As Suarez et al. [1994] noted, many results which are true effects may not be replicable, leading to erroneous conclusions that unreplicated results are false positives. Although several TP

two-point results were replicated in this study, the multipoint analysis failed to replicate the detection of MG1 for Q1. This failure to replicate the results from replicate 115 indicates that reliance on replication to confirm TP and eliminate FP results may not be a valid approach.

This study, based on only two data sets of 239 nuclear families, is a preliminary look at the tradeoff between FN and FP rates. In the future, comprehensive simulation studies based on larger numbers of replicates will be useful in clarifying the issues raised here. These results suggest that while false positives can be decreased by a combination of genotyping more widely spaced markers (reducing the number of comparisons made in the screen) and applying plateau analysis, using strict criteria for statistical significance risks failing to detect true linkages. Using fine-mapping tools and replicating the effects in independent data sets may reduce the number of FP while preserving the robustness of the genomic screen, but this approach may fail to separate TP from FP. Application of genome-wide significance levels or replication in independent data sets as criteria for publication risks failing to publicize TP linkage results.

ACKNOWLEDGMENTS

REFERENCES

Kruglyak L, Lander ES (1995): Complete multipoint sib-pair analysis of qualitative and quantitative traits.Am J Hum Genet 57:439–454. [PubMed: 7668271]

Lander E, Kruglyak L (1995): Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genet 11:241–247. [PubMed: 7581446]

S.A.G.E. (1994): Statistical Analysis for Genetic Epidemiology, Release 2.2 Computer program package available from the Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Suarez BK, Hampe CL, Van Eerdewegh P (1994): Problems of replicating linkage claims in psychiatry. In: Gershon ES, Cloninger CR (eds): "Genetic approaches to mental disorders," Washington, DC: American Psychiatric Press, pp 23–46.

Thomson G (1994): Identifying complex disease genes: progress and paradigms. Nature Genet 8:108–110. [PubMed: 7772120]

TABLE I.

Results from 2 cM and 10 cM Genomic Screens of Replicate 115

	2 cM Screen (367 markers)				10 cM Screen (80 markers)			
	N	MG	FP	FN	N	MG	FP	FN
Q1								
p < 0.05	63	3/3	54 (16%)	16 (64%)	11	2/3	9 (12%)	4 (67%)
p < 0.01	27	2/3	25 (7%)	23 (92%)	5	0/3	5 (7%)	6 (100%)
p < 0.001	6	1/3	5 (1%)	24 (96%)	1	0/3	1 (1%)	6 (100%)
Q2								
p < 0.05	47	1/1	46 (13%)	8 (89%)	11	0/1	11(14%)	2 (100%)
p < 0.01	21	0/1	21 (6%)	9 (100%)	6	0/1	6 (8%)	2 (100%)
p < 0.001	5	0/1	5 (1%)	9 (100%)	2	0/1	2 (3%)	2 (100%)
Q3								
p < 0.05	36	1/1	34 (9%)	5 (71%)	8	0/1	8 (10%)	2 (100%)
p < 0.01	13	1/1	12 (3%)	6 (86%)	2	0/1	2 (3%)	2 (100%)
p < 0.001	3	0/1	3 (<1%)	7 (100%)	0	0/1	0 (0%)	2 (100%)

N = number of significant markers

MG = number of major genes detected (at least one marker within 10 cM with statistically significant evidence of linkage).

FP = number (%) of false positives (significant results more than 10 cM from the true location of a gene).

FN = number (%) of markers 10 cM or less from a gene that did not have significant results.

TABLE II.

Plateau Analysis of Significant Markers, Replicate 115

Trait	Plateaus	Markers eliminated	Priority regions
Q1	15	15/64	D1G1–G8; D1G15–G18; D1G20–G22;D2G3–G4; D2G11–G12; D2G39–D2G40; D3G7–G9; D3G11–G14; D3G22–G24; D4G16–G17[*]; D4G21–G24; D4G28–G29; D5G12–G15[*]; D8G21–G22; D8G24–G26[*].
Q2	9	13/47	D3G1–G6; D6G1–G4; D6G16–17; D7G9–G16; D7G18–G20; D7G22–G24; D8G19–D8G20; D8G31–G32[*]; D8G34–G38.
Q3	9	15/36	D1G44–G45; D2G6–G8; D2G12–G15; D2G31–G32; D3G12–G13; D4G4–G5; D6G13–G14; D7G10–D7G11; D9G6–G7.

[*] = plateau associated with a major gene underlying the trait

TABLE III.

Multipoint Analysis of Replicate 115

	Major genes detected	Number of FP regions
Q1 (14 regions analyzed)		
$p < 0.05$	MG1, MG3	2
$p < 0.01$	MG1	1
$p < 0.001$	none	0
Q2 (7 regions analyzed)		
$p < 0.05$	none	0
$p < 0.01$	none	0
$p < 0.001$	none	0
Q3 (10 regions analyzed)		
$p < 0.05$	none	2
$p < 0.01$	none	0
$p < 0.001$	none	0