

Bivariate linkage analysis of cholesterol and triglyceride levels in the Framingham Heart Study

Xuyang Zhang¹ and Kai Wang*²

Address: ¹Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, USA and ²Department of Biostatistics, Division of Statistical Genetics, The University of Iowa, Iowa City, Iowa, USA

Email: Xuyang Zhang - xuyang-zhang@uiowa.edu; Kai Wang* - kai-wang@uiowa.edu

* Corresponding author

from Genetic Analysis Workshop 13: Analysis of Longitudinal Family Data for Complex Diseases and Related Risk Factors
New Orleans Marriott Hotel, New Orleans, LA, USA, November 11–14, 2002

Published: 31 December 2003

BMC Genetics 2003, **4**(Suppl 1):S62

This article is available from: <http://www.biomedcentral.com/1471-2156/4/s1/S62>

Abstract

We performed a bivariate analysis on cholesterol and triglyceride levels on data from the Framingham Heart Study using a new score statistic developed for the detection of potential pleiotropic, or cluster, genes. Univariate score statistics were also computed for each trait. At a significance level 0.001, linkage signals were found at markers GATA48B01 on chromosome 1, GATA21C12 on chromosome 8, and ATA55A11 on chromosome 16 using the bivariate analysis. At the same significance level, linkage signals were found at markers 036yb8 on chromosome 3 and GATA3F02 on chromosome 12 using the univariate analysis. A strong linkage signal was also found at marker GATA112F07 by both the bivariate analysis and the univariate analysis, a marker for which evidence for linkage had been reported previously in a related study.

Background

Elevated triglyceride and cholesterol levels are two risk factors for cardiovascular diseases. These risk factors are often correlated with each other. In order to map the possible pleiotropic/clustered genes underlying the inheritance of these two traits, we performed a bivariate linkage analysis using a score statistic developed by Wang [1]. This score statistic is asymptotically equivalent to the likelihood ratio statistic and is straightforward to compute. We apply this statistic to data from Cohort 1 and Cohort 2 of the Framingham Heart Study.

Methods

Data

Participants in Cohort 1 had up to 16 reported cholesterol levels, and up to 3 reported triglyceride levels. For participants in Cohort 2, cholesterol and triglyceride levels were reported up to 5 times. These two cohorts together provided 22,040 measurements on the cholesterol level and

9,155 measurements on the triglyceride level (including all repeated measurements on all individuals). Individuals who lacked any measurements of cholesterol level or triglyceride level were excluded. A single linear regression of cholesterol on age was fit across different individuals and different measurements. The residuals from the regression fit were averaged for each individual. This average was used as the age-adjusted cholesterol level for that individual. The same method was used to obtain age-adjusted triglyceride level for each individual. Sib pairs from the same nuclear family or from different nuclear families that belonged to the same pedigree were regarded as biologically unrelated. For the case of univariate traits, there are reports showing that treating dependent sib pairs as independent ones does not increase the type I error rate of the test [2].

All sib pairs in all the pedigrees in Cohort 1 and Cohort 2 were generated, but not all of these sib pairs were used at

Table 1: Number of markers available and analyzed on each chromosome

Chr	Markers Provided	Markers Analyzed	Sib Pairs ^A
1	32	30	965
2	30	28	981
3	27	27	380
4	21	21	1215
5	26	24	719
6	23	19	1187
7	22	20	1245
8	19	19	1298
9	19	17	401
10	22	20	970
11	17	15	788
12	19	18	1178
13	12	11	1524
14	17	14	1292
15	13	12	1318
16	14	14	1218
17	16	15	1284
18	15	14	1291
19	11	10	1426
20	11	10	1515
21	6	6	1798
22	7	6	1843

^AThe sample size is the same for all markers on the same chromosome.

the same time due to missing marker data. Genetic Analysis Workshop 13 (GAW13) provided identity-by-descent (IBD) sharing probabilities for some relative pairs (including sib pairs) at each of the scanned markers. The IBD sharing probabilities for a sib pair were available only for some markers. To simplify the programming, we excluded those markers at which there were less than 1000 sib pairs whose IBD sharing probabilities were available. Then, for each chromosome, we used only those sib pairs whose IBD sharing probabilities were available for all the remaining markers on that chromosome. See Table 1 for a summary of the number of markers excluded and the number of sib pairs used for each chromosome.

Analysis

The bivariate score statistic is computed based on the observed phenotypic data on sib pairs. The phenotypic data of a sib pair can be denoted by a vector of four (adjusted) measurements – cholesterol levels on sib 1 and sib 2, and triglyceride levels on sib 1 and sib 2. Let x_i be the phenotypic data on the i^{th} sib pair and Σ_0 be the sample variance-covariance of x_i . As an average of the residuals of a regression, the sample mean of cholesterol levels on sib 1 and sib 2 is 0, so is the sample mean of triglyceride level. Let Σ_0 be a 4×4 symmetric matrix whose (i,j) element is denoted by a_{ij} . Note that a_{11} and a_{33} are the variances of the cholesterol and triglyceride levels, respectively, of the first sib in the pairs. Similarly, a_{22} and a_{44} are

the variances of the cholesterol and triglyceride levels of the second sib in the sib pairs. The off-diagonal terms represent covariances: $a_{13} = a_{31}$ is the covariance between cholesterol and triglycerides for the first sib in the sib pairs, and $a_{24} = a_{42}$ is the covariance for the second sib in the sib pairs. Since the sib-sib relationship in a sib pair is symmetric, we expect that $a_{11} \approx a_{22}$, $a_{33} \approx a_{44}$ and $a_{13} \approx a_{24}$ when the sample size is large. Alternatively, we can also use the (adjusted) measurements on cholesterol and triglycerides on all sibs (do not distinguish sib 1 from sib 2) in calculating the entries of Σ_0 . Then there would be $a_{11} = a_{22}$, $a_{33} = a_{44}$, and $a_{13} = a_{24}$. Since the sample size is fairly large, we expect both methods give similar Σ_0 .

Define

$$w_i = (w_{i1}, w_{i2}, w_{i3}, w_{i4})^t = \Sigma_0^{-1} x_i$$

and

$$z_i = w_{i1}w_{i2}a_{11} + (w_{i1}w_{i3} + w_{i2}w_{i4})a_{13} + w_{i3}w_{i4}a_{33}.$$

Denote the proportion of alleles that are shared IBD by the i^{th} sib pair by π_i . Let $\bar{\pi}$ and \bar{z} be the sample means of $\{\pi_i\}$ and $\{z_i\}$, respectively. Define

$$b = \sum_{i=1}^N (\pi_i - \bar{\pi})(z_i - \bar{z}),$$

where N is the total number of sib pairs. When the putative locus is not linked to any trait locus, the expectation of b is 0 and its variance is $\text{Var}(b) = N s_{\pi}^2 s_z^2$, where s_{π}^2 and s_z^2 are the sample variances of $\{\pi_i\}$ and $\{z_i\}$, respectively. The score statistic S for the bivariate phenotypes is defined by $S = b^2/\text{Var}(b)$ if $b > 0$; $S = 0$ otherwise. When the putative locus is not linked to any quantitative trait loci (QTL), the asymptotic distribution of this one-sided tests statistic, S , is $0.5 \chi^2_0 + 0.5 \chi^2_1$ [1]. The score statistic S is a special case described by Wang [1] – the locus specific variances and covariance for the two traits are assumed to be proportional to their total variances and covariance.

Results

The score statistic S was calculated for every screened marker. In addition, the univariate score statistic of Wang and Huang [3] was also calculated for cholesterol level and triglyceride level separately. For sib-pair data, the type of data used in our analyses, this univariate score statistic is equivalent to other methods [4,5]. The p -values of these three score statistics (one for the bivariate phenotypes, one for each of the two univariate phenotypes) at each marker location are plotted in Figure 1. Markers with p -values less than the significance level of $\alpha = 0.005$ are shown in Table 2.

At the significance level 0.005, 10 markers were identified by the bivariate score statistic: 2 each from chromosome 1 (at 212 cM and 233 cM) and 7 (at 128 cM and 155 cM), and 1 each from chromosome 3 (at 112 cM), 4 (at 105 cM), 5 (at 19 cM), 6 (at 166 cM), 8 (at 140 cM), and 16 (at 64 cM). Five out of the 10 markers were also identified by the univariate score statistic for the adjusted triglyceride level. They were the two on chromosome 1, one on chromosome 7 (at 155 cM), one on chromosome 8, and one on chromosome 16. None of the 10 markers were identified by the univariate score statistic for the age-adjusted cholesterol level. The results seem to suggest that there were large overlaps of linkage signals between the bivariate score statistic and the univariate score statistic for the age-adjusted triglyceride level. There were no overlaps of linkage signals between the bivariate score statistic and the univariate score statistic for the age-adjusted cholesterol level. There were 5 markers that were identified by the bivariate score statistic, but not identified by any of the univariate score statistics. There were 3 markers whose p -values were below 0.001: one on chromosome 1 at 212 cM, one on chromosome 8 at 140 cM, and the other on chromosome 16 at 64 cM. The regions suggested by these 3 markers may be investigated in future genotyping and analysis.

Discussion

We performed a bivariate analysis of cholesterol and triglyceride levels on sib-pair data from the Framingham Heart Study using a method recently developed by Wang [1]. This method is asymptotically equivalent to the likelihood ratio statistic, but is straightforward to calculate. We also calculated the univariate score statistics for cholesterol and triglyceride levels separately. Five markers were identified by both the bivariate score statistic and the univariate score statistic for the adjusted triglyceride level, while the results of the bivariate score statistics had no overlap with the univariate score statistic for the age adjusted cholesterol levels.

The method in Wang [1] is general enough to handle general pedigrees, but we only applied it to sib pairs that were extracted from general pedigrees. This is because the programming for sib pairs is relatively easy and was feasible given the time constraint for GAW13. Some linkage information may have lost due to the fact that dependent sib pairs were treated as independent sib pairs, but the type I error rate of the test statistic is expected to be valid.

In a related study, Shearman et al. [6] used the ratio of triglyceride level to high-density lipoprotein cholesterol level as the phenotype of interest. Linkage evidence was reported at marker GATA112F07 (155 cM on chromosome 7), a marker that resulted in a p -value 0.0020 for the bivariate score statistic used in the current report. These authors reported a LOD score 1.5 at 70 cM on chromosome 16 with multipoint mapping. We used single-point IBD sharing probabilities with the bivariate score statistic and obtained a significant linkage signal ($p = 0.0001$) for marker ATA55A11 (64 cM on chromosome 16), 6 cM away from the locus they identified. Other markers in Table 2 that have small p -values for the bivariate or univariate score statistics but that did not show evidence for linkage in Shearman et al. [6] include GATA48B01, 036yb8, GATA21C12, and GATA3F02.

One caveat about bivariate analyses is that they are not always more powerful than univariate analyses. Theoretical [7] and simulation studies [1,8,9] demonstrate that when the polygenic correlation is in the same direction as the major gene correlation, a bivariate analysis may have lower power than a univariate analysis.

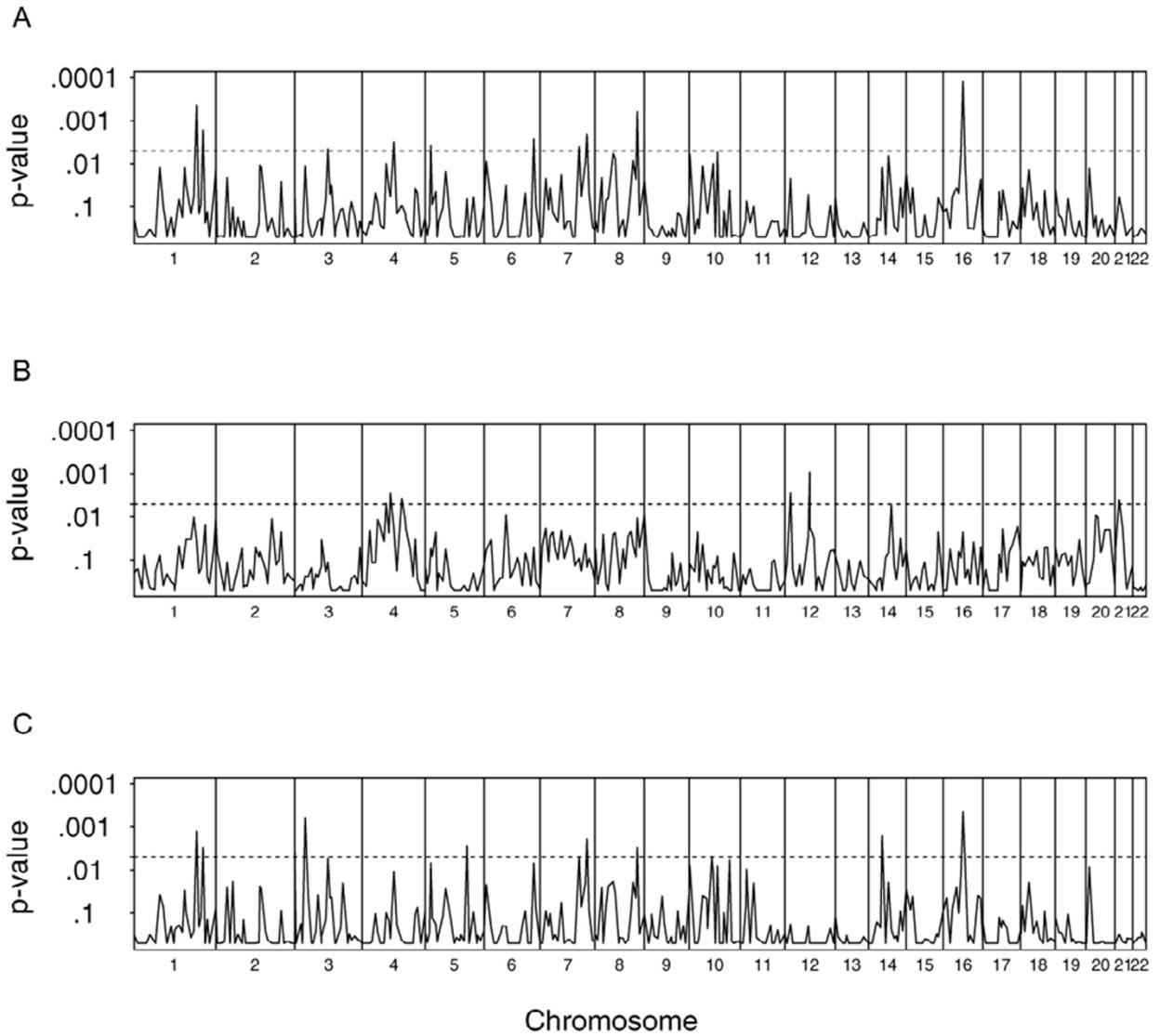


Figure 1
p-values for the three statistics throughout the genome Bivariate score statistic (A), univariate score statistic for adjusted cholesterol level (B), and univariate score statistic for adjusted triglyceride level (C).

Table 2: Summary of markers that are significant at significance level 0.005

Chr.	Marker	p-telomere (cM)	Bivariate Score statistic	Univariate score statistic	
				Cholesterol	Triglyceride
1	Gata48B01	212	11.029(.0004)		9.141(.0013)
1	GATA87F04	233	8.618(.0017)		7.602(.0029)
3	3PTEL25	1			8.358(.0019)
3	036yb8	37			10.445(.0006)
3	GATA128C02	112	6.799(.0046)		
4	GATA24H01	78		6.670(.0049)	
4	ATA2A03	93		7.665(.0028)	
4	GATA2F11	105	7.534(.0030)		
4	ATA26B08	130		7.159(.0037)	
5	GATA2A04	19	7.119(.0038)		
5	GATA2H09	139			7.682(.0028)
6	242zg5	166	7.776(.0026)		
7	GGAA6D03	128	7.036(.0040)		
7	GATA112F07	155	8.274(.0020)		8.410(.0019)
8	GATA21C12	140	10.474(.0006)		7.603(.0029)
12	GATA49D12	18		7.695(.0028)	
12	GATA3F02	81		9.674(.0009)	
14	GATA4B04	44			8.648(.0016)
16	ATA55A11	64	13.489(.0001)		11.049(.0004)
21	GGAA3C07	13		7.043(.0040)	

Acknowledgments

We thank three anonymous reviewers for their constructive comments. This work is supported in part by a clinical research center grant P0-DC-02748 from the National Institute on Deafness and Other Communication Disorders (PI: Dr. J. Bruce Tomblin; Investigators include XZ) and National Institute of Mental Health grant R01-52841 (PI: Dr. Veronica Vieland; Investigators include KW).

References

1. Wang K: **Mapping quantitative trait loci using multiple phenotypes in general pedigrees.** *Hum Hered* 2003 in press.
2. Amos CI, Elston RC, Wilson RF, Bailey-Wilson JE: **A more powerful robust sib-pair test of linkage for quantitative traits.** *Genet Epidemiol* 1989, **6**:435-449.
3. Wang K, Huang J: **A score-statistic approach for mapping quantitative-trait loci with sibships of arbitrary size.** *Am J Hum Genet* 2002, **70**:412-424.
4. Sham PC, Purcell S: **Equivalence between Haseman-Elston and variance-components linkage analyses for sib pairs.** *Am J Hum Genet* 2001, **68**:1527-1532.
5. Putter H, Sandkuijl LA, van Houwelingen JC: **Score test for detecting linkage to quantitative traits.** *Genet Epidemiol* 2002, **22**:345-355.
6. Shearman AM, Ordovas JM, Cupples LA, Schaefer EJ, Harmon MD, Shao Y, Keen JD, DeStefano AL, Joost O, Wilson PVF, Housman DE, Myers RH: **Evidence for a gene influencing the TG/HDL-C ratio on chromosome 7q32.3-qter: a genome-wide scan in the Framingham Study.** *Hum Mol Genet* 2000, **9**:1315-1320.
7. Jiang C, Zeng ZB: **Multiple trait analysis of genetic mapping for quantitative trait loci.** *Genetics* 1995, **140**:1111-1127.
8. Allison DB, Thiel B, Jean PS, Elston RC, Intante MC, Schork NJ: **Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages.** *Am J Hum Genet* 1998, **63**:1190-1201.
9. Amos CI, de Andrade M, Zhu DK: **Comparison of multivariate tests for genetic linkage.** *Hum Hered* 2001, **51**:133-144.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

