BMC Genetics

Proceedings

Genome-wide linkage analysis of blood pressure under locus heterogeneity

Xinqun Yang[1], Kai Wang*[1], Jian Huang[1,2] and Veronica J Vieland[1,3]

Address: [1]Department of Biostatistics, Division of Statistical Genetics, The University of Iowa, Iowa City, USA, [2]Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, USA and [3]Department of Psychiatry, The University of Iowa, Iowa City, USA

Email: Xinqun Yang - xinqun-yang@uiowa.edu; Kai Wang* - kai-wang@uiowa.edu; Jian Huang - jian@stat.uiowa.edu; Veronica J Vieland - vvieland@mail.public-health.uiowa.edu

* Corresponding author

Abstract

We describe a method for mapping quantitative trait loci that allows for locus heterogeneity. A genome-wide linkage analysis of blood pressure was performed using sib-pair data from the Framingham Heart Study. Evidence of linkage was found on four markers (GATA89G08, GATA23D06, GATA14E09, and 049xd2) at a significance level of 0.01. Two of them (GATA14E09 and 049xd2) seem to overlap with linkage signals reported previously, while the other two are not linked to any known signals.

Background

High blood pressure (BP) is an important risk factor for cardiovascular disease and is a leading cause of mortality in industrialized countries [1]. Being a complex trait, BP may be influenced by genes at different locations on the genome in different families (inter-family locus heterogeneity). Presumably, allowing for locus heterogeneity in the analysis will increase the power to detect linkage.

We propose a model that allows for locus heterogeneity for quantitative trait loci (QTLs). This model generalizes the locus heterogeneity model for qualitative traits [2,3] to the case of continuous traits. A genome-wide linkage analysis was conducted on the Framingham Heart Study data using the likelihood-ratio statistic based on this model.

Methods

Data

The Framingham Heart Study consists of two cohorts with a total of 10,333 subjects. The first cohort includes 5209 individuals who were recruited initially in 1948 when the study began. The second cohort includes 5124 subjects who are offspring, or spouses of offspring, of the initial participants. Longitudinal data were recorded over follow-up years in both cohorts. In the total 330 pedigrees provided by the Genetic Analysis Workshop 13 (GAW13), 1702 subjects were genotyped at at least one marker and 2885 subjects were phenotyped for BP at least once.

To remove the effect of hypertension treatment, we focused only on those subjects who never received any antihypertensive medication. There are 1909 such subjects. Their characteristics are summarized in Table 1. The systolic blood pressures were adjusted for age, sex, and body mass index (BMI) in the following manner. The averages of BP, age, and BMI were computed for each subject, and then the average of BP was regressed over those of age and BMI as well as sex, separately for each of the two cohorts. The residuals of the simple linear regression were used as the (adjusted) phenotypic values. The two-point identity-by-descent (IBD) sharing probabilities on

Table 1: Summary statistics for subjects in Cohort 1 and Cohort 2 (standard errors are in parentheses)

Cohort	Number by Gender			Age (yr)	BMI (kg/m^2)	SBP (mm Hg)
	Male	Female	Total			
1	304	269	573	56.0 (7.4)	25.7 (3.7)	130.1 (16.1)
2	659	677	1336	40.9 (10.1)	25.5 (4.3)	118.5 (12.3)

relative pairs provided by GAW13 were used. Only sib pairs were used in the analysis. Sib pairs from the same sibship or from different sibships in the same pedigree were treated as if they were biologically unrelated.

A total of 400 markers were used in the genome scan. Not all sib pairs were genotyped at every marker. So the number of sib pairs used in the analysis varied from marker to marker, roughly from 400 to 1500.

Significant departure from normality was detected on the (adjusted) BP measurement using the Shapiro-Wilk test (*p*-value < 0.0001). To reduce the impact of the non-normality of the data on the performance of the proposed statistic, the BP measurements were transformed based on the normal copula model [4] so the transformed trait values follow a standard normal distribution, with a mean and variance of 0 and 1, respectively.

Statistical methods

Consider a sib pair whose trait values are denoted by y_1 and y_2, respectively. The trait values are standardized such that they have mean 0 and variance 1. It is assumed that the candidate locus has an additive effect but no dominance effect on the trait. Let ρ_0, ρ_1, and ρ_2 be the correlation coefficients of y_1 and y_2 when the number of alleles that are shared IBD by the sib pair is 0, 1, and 2, respectively. Because there is no additive effect, we have $\rho_1 = (\rho_0 + \rho_2)/2$ [4]. When there is no linkage, ρ_1 can be estimated by taking the value of correlation coefficient between y_1 and y_2 [5,6]. So ρ_1 can be treated as known.

Conditional on the IBD sharing status, y_1 and y_2 are assumed to have a bivariate normal distribution, which is completely characterized by the correlation coefficients ρ_0, ρ_1, or ρ_2. Let π_0, π_1, and π_2 be the probabilities that the sib pair shares 0, 1, or 2 alleles IBD, respectively, and then the likelihood function for y_1 and y_2 is

$$L(\rho_2; y_1, y_2) = \pi_0 \phi (y_1, y_2; 2\rho_1 - \rho_2) + \pi_1 \phi (y_1, y_2; \rho_1) + \pi_2 \phi (y_1, y_2; \rho_2),$$

where $\phi(\cdot, \cdot; \cdot)$ denotes the density function of a bivariate normal distribution. When the candidate locus is not

linked to any QTL, $\rho_1 = \rho_2$ and the above likelihood reduces to $L(\rho_1; y_1, y_2) = \phi(y_1, y_2; \rho_1)$. Otherwise, $\rho_2 > \rho_1$.

Let α be the probability that a QTL is linked to the candidate locus for the sib pair. Then the likelihood function for the sib pair would be

$$\alpha L(\rho_2; y_1, y_2) + (1 - \alpha)L(\rho_1; y_1, y_2).$$

Let *i* index the *i*th sib pair in a collection of sib pairs, then the log-likelihood function for these sib pairs is

$$l(\alpha, \rho_2) = \Sigma_i \log[\alpha L(\rho_2; y_{i1}, y_{i2}) + (1 - \alpha)L(\rho_1; y_{i1}, y_{i2})],$$

where y_{i1} and y_{i2} are the trait values for the *i*th sib pair. The hypotheses of interest are

$$H_0: \rho_2 = \rho_1, \alpha \in [0,1] \text{ vs. } H_1: \rho_2 > \rho_1, \alpha \in [0,1]$$

This model is an extension of the heterogeneity model of Smith [2] for dichotomous traits to continuous traits. The likelihood ratio statistic was used as the test statistic. To obtain the maximum likelihood estimates (MLEs) of parameters, we used a two-dimensional grid search over α and ρ_2, with the grid size of 0.01 for both parameters.

The asymptotic distribution of the likelihood ratio statistic is not known analytically. So a simulation study with 10,000 replicates was carried out. The trait locus is assumed to have two equally frequent alleles and its heritability is set to 0.3. The unlinked marker is assumed to be fully polymorphic. The 90th, 95th, 99th, and 99.9th percentiles of the distribution are found to be 2.72, 3.91, 6.61, and 11.07, respectively.

Results

A genome scan is conducted using three statistics: the test statistic of Haseman-Elston method (H-E) [7], the likelihood ratio statistic (HET-LRT) for the heterogeneity model described above, and the likelihood ratio statistic (HOM-LRT) for the homogeneity model, which is a special case of the heterogeneity model with α fixed at 1. The markers that are significant for at least one statistic at significance level 0.01 are listed in Table 2.

Table 2: Significant markers at significance level 0.01 (*p*-values are in parentheses)

Marker	Chr	cM	H-E	HOM-LRT	HET-LRT
GATA6F06	3	79	5.40 (0.009)	6.02 (0.007)	
GATA28F03	4	73	7.42 (0.003)		
GATA89G08	5	98	10.06 (0.001)	6.68 (0.005)	6.68 (0.009)
GATA23D06	8	26	6.99 (0.004)	6.68 (0.005)	6.68 (0.009)
GGAA20C10	8	60	6.45 (0.005)		
GATA14E09	8	94	11.36 (0.0003)	6.76 (0.004)	6.76 (0.009)
GATA21C12	8	140	6.15 (0.006)		
GATA62F03	9	14	10.88 (0.0004)		
GGAA2F11	10	117	8.99 (0.001)	5.42 (0.010)	
ATA29A06	12	161	10.79 (0.0004)		
GATA13D05	12	166	6.50 (0.005)		
GATA30A03	14	92	9.38 (0.001)		
ATA3A07	16	23	7.64 (0.002)	5.39 (0.010)	
049xd2	16	44	8.92 (0.001)	7.23 (0.003)	7.23 (0.007)
ATC6A06	17	67	5.50 (0.009)		
ATA7D07	18	89	6.81 (0.004)		
GATA188F04	21	40	6.05 (0.006)	5.88 (0.008)	

At significance level 0.01, four significant markers are identified by statistic HET-LRT. Among the four markers, GATA14E09 on chromosome 8 is very close to 8q21.11, a region that showed linkage to BP [8]. 049xd2 on chromosome 16 is between 16p12 and 16p13.1, a region in which linkage to BP was found by two independent studies [8,9]. To the best of our knowledge, the other two markers, i.e., GATA89G08 on chromosome 5 and GATA23D06 on chromosome 8, do not link to any previous findings.

Discussion

We introduce a heterogeneity model for mapping QTLs using sib-pair data. A genome scan was performed to map QTLs affecting BP variation on a population-based data using this new method. At significance level 0.01, evidence of linkage is found in four marker regions on chromosomes 5, 8, and 16. Two of the markers seem to overlap with linkage signals in previous studies, while the other two are not linked to any previous findings.

Of the three statistics used in the genome scan, H-E provides the most linkage signals, followed by HOM-LRT, and then by HET-LRT. At the four markers at which HET-LRT is significant (and at many nonsignificant markers that are not shown), the statistic value of HET-LRT is the same as that of HOM-LRT. Finer grid size may change the values of HOM-LRT and HET-LRT, but the changes are expected to be small. Experiments on the four significant markers with grid size 0.001 strongly support this claim. In these experiments, HET-LRT is still the same as HOM-LRT for all four markers.

The proposed statistic HET-LRT is intended for a population sample. Its performance under selected samples is unknown. Although the normal copula model can be used on selected samples to recover normality, its effectiveness has yet to be investigated. It is worthwhile to point out that our analysis has excluded individuals on antihypertensive medication. Therefore, careful attention should be paid when generalizing our results to general population. For a recent review on the issues dealing with antihypertensive treatments, see Palmer [10].

We have been using the posterior probability of linkage (PPL) to assess linkage signals across heterogeneous data sets [11-14]. Since the proposed heterogeneity model for quantitative traits is parallel to that for qualitative traits, it is reasonable to adapt our previous work on qualitative traits to the current setting. The details are being worked out.

The proposed model is for sib-pair data only. It is expected that the use of nonindependent sib-pairs, like what we did in this analysis, does not affect the type I error rate asymptotically. It is of interest to see how this model can be generalized to general pedigrees and how the generalized model performs.

Acknowledgments

References

1. Guidelines Subcommittee: **World Health Organization – International Society of Hypertension Guidelines for the Management of Hypertension.** *J Hypertens* 1999, **17:**151-183.
2. Smith CAB: **Testing for heterogeneity of recombination fraction values in human genetics.** *Ann Hum Genet* 1963, **27:**175-182.
3. Terwilliger JD: **A likelihood-based extended admixture model of oligogenic inheritance in 'model-based' and 'model-free' analysis.** *Eur J Hum Genet* 2000, **8:**399-406.
4. 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.
5. Tang H, Siegmund D: **Mapping quantitative trait loci in oligogenic models.** *Biostatistics* 2001, **2:**147-162.
6. Wang K: **Efficient score statistics for mapping quantitative trait loci with extended pedigrees.** *Hum Hered* 2002, **54:**57-68.
7. Haseman JK, Elston RC: **The investigation of linkage between a quantitative trait and a marker locus.** *Behav Genet* 1972, **2:**3-19.
8. Rice T, Rankinen T, Province MA, Province MA, Chagnon YC, Perusse L, Borecki IB, Bouchard C, Rao DC: **Genome-wide linkage analysis of systolic and diastolic blood pressure: The Quebec Family Study.** *Circulation* 2000, **102:**1956-1963.
9. Harrap SB, Wong ZY, Stebbing M, Lamantia A, Bahlo M: **Blood pressure QTLs identified by genome-wide linkage analysis and dependence on associated phenotypes.** *Physiol Genomics* 2002, **8:**99-105.
10. Palmer LJ: **Loosening the cuff: important new advances in modelling antihypertensive treatment effects in genetic studies of hypertension.** *Hypertension* 2003, **41:**197-198.
11. Vieland VJ: **Bayesian linkage analysis, or: how I learned to stop worrying and love the posterior probability of linkage.** *Am J Hum Genet* 1998, **63:**947-954.
12. Wang W, Vieland VJ, Huang J: **A Bayesian approach to replication of linkage findings.** *Genet Epidemiol* 1999, **17(suppl 1):**S749-S754.
13. Vieland VJ, Wang K, Huang J: **Power to detect linkage based on multiple sets of data in the presence of locus heterogeneity: comparative evaluation of model-based linkage methods for ASP data.** *Hum Hered* 2001, **51:**199-208.
14. Wang K, Huang J, Logue M, Vieland VJ: **Combined multipoint analysis of multiple asthma data sets based on the posterior probability of linkage.** *Genet Epidemiol* 2001, **21(suppl 1):**S73-S78.