BMC Genetics

Proceedings

Genome-wide linkage scan for genes affecting longitudinal trends in systolic blood pressure

Kevin B Jacobs[1,2], Courtney Gray-McGuire[1], Kevin C Cartier[1] and Robert C Elston*[1]

Address: [1]Department of Epidemiology and Biostatistics, Division of Genetic and Molecular Epidemiology, Case Western Reserve University, Cleveland, Ohio, USA and [2]Division of Statistical Genetics and Bioinformatics, The OPAL Group, Cleveland, Ohio USA

Email: Kevin B Jacobs - jacobs@theopalgroup.com; Courtney Gray-McGuire - mcguire@darwin.cwru.edu; Kevin C Cartier - kcartier@darwin.cwru.edu; Robert C Elston* - rce@darwin.cwru.edu

* Corresponding author

Abstract

Only one genome scan to date has attempted to make use of the longitudinal data available in the Framingham Heart Study, and this attempt yielded evidence of linkage to a gene for mean systolic blood pressure. We show how the additional information available in these longitudinal data can be utilized to examine linkages for not only mean systolic blood pressure (SBP), but also for its trend with age and its variability. Prior to linkage analysis, individuals treated for hypertension were adjusted to account for right-censoring of SBP. Regressions on age were fitted to obtain orthogonal measures of slope, curvature, and residual variance of SBP that were then used as dependent variables in the model-free linkage program SIBPAL. We included mean age, gender, and cohort as covariates in the analysis. To improve power, sibling pairs were weighted for informativity using weights derived from both the marker and trait. The most significant results from our analyses were found on chromosomes 12, 15, and 17 for mean SBP, and chromosome 20 for both SBP slope and curvature.

Background

Hypertension is a complex disorder that involves both environmental and genetic components [1]. Unfortunately, little is understood about the genetics of the overall variance of blood pressure or the changes in blood pressure over time. Levy et al. [2] conducted a genome scan of the Framingham Heart Study data using longitudinal blood pressure measures in an effort to describe the underlying variation of blood pressure in the community. This report presents a similar analysis of longitudinal blood pressure measures from the Framingham Heart Study data. However, in addition to examining the age-adjusted mean, as did Levy et al. [2], we evaluated longitudinal measures including the slope, the curvature, and the log residual variance of systolic blood pressure (SBP) over time.

Methods

Sample and phenotype definition

Selection criteria, study design, and data collection methods for the Framingham Heart Study have been detailed previously [3,4]. The sample available for study consists of 330 families with 2803 sibling pairs, a small subset of the total Framingham Heart Study data that were suitable for linkage studies. Phenotype information for all individuals with exam data between the ages of 25 and 75 (2413 individuals) was used to obtain longitudinal phenotypes as follows: let t_{ij} be the SBP measured at age x_{ij}, on the ith indi-

vidual, $j = 1, 2, ..., n_i$. To adjust for the effects of SBP due to treatment for hypertension, we chose to employ a simple method that imputes an "untreated" SBP by taking the maximum of: 1) SBP measured at time points when the individual was being treated to lower blood pressure plus a constant (chosen to be 15 mm Hg, based on results of an independent study of the effect of treatment on SBP after being off treatment for 2 weeks (M. Bochud, personal communication)), 2) 140 mm Hg, the clinical threshold for the diagnosis of hypertension for the Framingham study, and 3) the last SBP measurement before treatment commenced.

Define γ_{ij} as

$$
\gamma_{ij} = \begin{cases} t_{ij} & \text{if individual } i \text{ was not treated at time } j \\ \max(t_{ij} + 15, 140, t_{ik}) & \text{otherwise} \end{cases},
$$

where t_{ik} is the most recent untreated measure of SBP for individual i at some time-point $k < j$. Unlike other methods, we do not attempt to infer information about the tail of the uncensored SBP distribution from those individuals who, in spite of being clinically hypertensive, do not receive treatment for their condition. We judged that there was insufficient information on such individuals in the available Framingham data set to attempt that style of imputation.

Orthogonal polynomial summarization of longitudinal SBP

For each individual, we then transformed these n_i values of γ_{ij} into four new variables: the mean (b_{i1}), the slope (b_{i2}), the curvature (b_{i3}), and log residual variance ($\log s_i^2$), by constructing orthogonal polynomials for unequally spaced variables [5]. Let $\mathbf{z}_i$ be an $n_i \times 3$ matrix,

$$
\mathbf{z}_i = \begin{bmatrix} 1 & x_{i1} - \overline{x} & x_{i1}^2 - \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}^2 - \left(x_{i1} - \overline{x}_i\right)\sum_{j=1}^{n_i} x_{ij}^2 \left(x_{ij} - \overline{x}_i\right) / \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}_i\right)^2 \\ 1 & x_{i2} - \overline{x} & x_{i2}^2 - \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}^2 - \left(x_{i2} - \overline{x}_i\right)\sum_{j=1}^{n_i} x_{ij}^2 \left(x_{ij} - \overline{x}_i\right) / \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}_i\right)^2 \\ \vdots & \vdots & \vdots \end{bmatrix}.
$$

The first three new variables for the ith individual are the elements of

$$
\mathbf{b}_i = (\mathbf{z}_i' \mathbf{z}_i)^{-1} \mathbf{z}_i' \mathbf{y}_i,
$$

where $\mathbf{y}_i$ is the $n_i \times 1$ vector of SBP values for the ith individual, and the fourth new variable is the logarithm of the residual variance of the model,

$$
s_i^2 = \frac{\mathbf{y}_i'(\mathbf{y}_i - \mathbf{z}_i \mathbf{b}_i)}{n_i - 3}.
$$

Linkage analysis

Multipoint allele sharing identical by descent (IBD) was estimated at approximately 2-cM intervals using the GENIBD computer program from the S.A.G.E. package [6]. After removing uninformative individuals and partitioning pedigrees into independent sections (with no loss of linkage information), 14 pedigrees were still too large to be analyzed efficiently with GENIBD. Those pedigrees were split into nuclear families that were analyzed as though independent, once genotype inference and elimination had been performed to minimize information loss.

Linkage analysis was then performed on approximately 1077 sibling pairs (varying by genetic marker informativity) by regressing the weighted squared sums and differences of each of the new variables (denoted below as γ_a and γ_b for the pair of sibs a and b) on the mean proportion of alleles shared IBD and covariates for cohort membership, gender, mean age, and a conditional mean effect:

$$
E\left[\frac{1}{\sigma_s^2}(\gamma_a + \gamma_b)^2 + \frac{1}{\sigma_d^2}(\gamma_a - \gamma_b)^2 \right] = \beta_1 + \beta_2 \hat{\pi}_{ab} + \beta_3 c_a c_b + \beta_4 g_a g_b + \beta_5 a_a a_b + \beta_6 (\gamma_a + \gamma_b),
$$

where σ_s^2 and σ_d^2 are the residual variances of similar regressions on the mean-corrected squared trait sums and differences [6-9], $\hat{\pi}_{ab}$ is the estimated multipoint mean marker allele sharing IBD between the full sibling pair a and b, c_a and c_b are indicator variables for membership in the offspring cohort, g_a and g_b are mean-corrected indicator variables for gender (male = 0, female = 1), and a_a and a_b are the mean ages of a and b over all exams. Including the sum of the trait values as a covariate allows for the simultaneous estimation of the mean trait value conditional on marker allele sharing and all covariates, which in turn increases the power to detect linkage.

The above model was fitted using generalized least squares to account for the trait correlations among pairs of sibling pairs that occur in sibships of size greater than two [7,8]. Additionally, regression weights based on trait and marker informativity were incorporated to increase the power of our linkage test and account for the differential information content (especially in the amount of longitudinal data available on each subject). Marker informativity was determined by a distance metric, D, within a simplex determined by the probability that a sibling pair shares i alleles IBD f_i, where

$$
D = \begin{cases} \frac{1}{3}\left(1 - 2\hat{f}_0 / (1 - \hat{f}_1)\right)^2 + \frac{1}{4}\left(1 - 2\hat{f}_1\right)^2 & \text{if } \hat{f}_1 < 1 \\ \frac{1}{4} & \text{otherwise.} \end{cases}
$$

Trait-based weights for b_{i1}, b_{i2}, and b_{i3} were calculated based on the inverse sum of the variance of each parameter estimate and the residual variance,

$$w(b_{il}) = \frac{1}{\text{var}(b_{il}) + s_i^2}, \text{ for } l = 1, 2, 3.$$

Trait-based weights for $\log s_i^2$ were calculated based on a first order approximation of the inverse of the variance of $\log s_i^2$,

$$w(\log s_i^2) = \frac{1}{\text{var}(\log s_i^2)} \approx n_i - 3.$$

A combined weight that included information on both trait and marker informativity was obtained by taking the geometric mean of the marker and trait weights. We tested for the presence of significant linkage using usual right-tailed t-test statistics $\hat{\beta}_2 / \hat{\sigma}_2$, where $\hat{\beta}_2$ is an estimate of β_2, and $\hat{\sigma}_2^2$ is the corresponding estimated variance of $\hat{\beta}_2$.

Results and Discussion

All results presented below include the covariates mentioned above in their models. However, the significance of these covariates will not be discussed because it was not the design of the analysis to obtain interpretable parameter estimates, but rather to test for linkage. Criteria for suggestive and significant evidence for linkage were chosen as markers having *p*-values lower than 10^{-4} and 10^{-5}, respectively. Using these criteria, the most significant results for mean SBP were found on chromosome 12 near marker GATA47F05 (*p* = 0.00000029), on chromosome 15 between markers GATA22F01 and GATA27A03 (*p* = 0.00000021), and on chromosome 17 near marker GATA28D11 (*p* = 0.0000048), and on chromosome 20 near marker GATA47F05 for both slope and curvature of SBP (*p* = 0.000042, *p* = 0.0000028), as shown in Table 1

Table 1: Regions with suggestive or significant evidence of linkage to SBP

Trait	Weights[A]	Chromosome: Region	Marker Name(s)	Maximum[B] -log$_{10}$ (*p*-value)
SBP mean (**b**$_1$)	t, **c**	5: 24–36 cM	GATA3E10	5.09 at 32 cM
SBP mean (**b**$_1$)	t, m, **c**	12: 68–82 cM	GGAT2G06	6.54 at 78 cM
SBP mean (**b**$_1$)	t, m, **c**	15: 111–134 cM	GATA22F01, GATA27A03	6.68 at 129 cM
SBP mean (**b**$_1$)	**c**	16: 133–139 cM		4.34 at 137 cM
SBP mean (**b**$_1$)	t, **c**	17: 107–119 cM	GATA28D11	5.32 at 109 cM
SBP slope (**b**$_2$)	**m**	20: 70–78 cM	GATA47F05	4.38 at 72 cM
SBP curvature (**b**$_3$)	u, t, **m**, c	20: 70–82 cM	GATA47F05	5.55 at 74 cM
SBP mean (**b**$_1$)	**m**, c	22: 35–48 cM	GATA6F05, GATA11B12	4.85 at 38 cM

[A]u, unweighted model; t, trait weights; m, marker weights; c, combined weights. [B]Reported maximum -log *p*-values refer to the corresponding weighting scheme indicated in bold.

and Figure 1.

Although it was one of our goals, the analysis performed for the mean SBP is not directly comparable to the linkage scan performed by Levy et al. [2]. A fundamental source of differences is the approach taken to correct for right-censoring of the SBP phenotype for individuals treated for hypertension. Levy et al. [2] used the full Framingham data set on 87,840 examinations on 10,313 individuals to estimate the trait distribution of SBP and applied an empirical adjustment to the SBP of treated individuals conditional on age, sex, and cohort. However, our analysis was limited to the 2413 individuals available as part of the Framingham linkage data, so an equivalent method of adjustment was not deemed feasible. (However, the Genetic Analysis Workshop13 contribution by Diego and Almasy reports that they did duplicate the imputation procedure described by Levy et al., and were successful in reproducing the most significant linkage finding.) Instead, we chose to adjust measurements of SBP of treated individuals by increasing their SBP using an approximate estimator based on their measured SBP plus a constant, their last untreated measurement, and the study-defined clinical threshold for being hypertensive. Adding a constant to a treated SBP measurement reflects a belief that hypertensive treatment has some degree of efficacy. Utilizing a retrospective measure of the last untreated SBP controls for many latent individual-specific factors, and makes the mild assumption of a non-decreasing trend in SBP over time for those who are eventually treated for hypertension. The final criteria enforced the minimum clinical criterion for diagnosis of hypertension, presuming that any subject that was being treated for hypertension had a SBP high enough to warrant
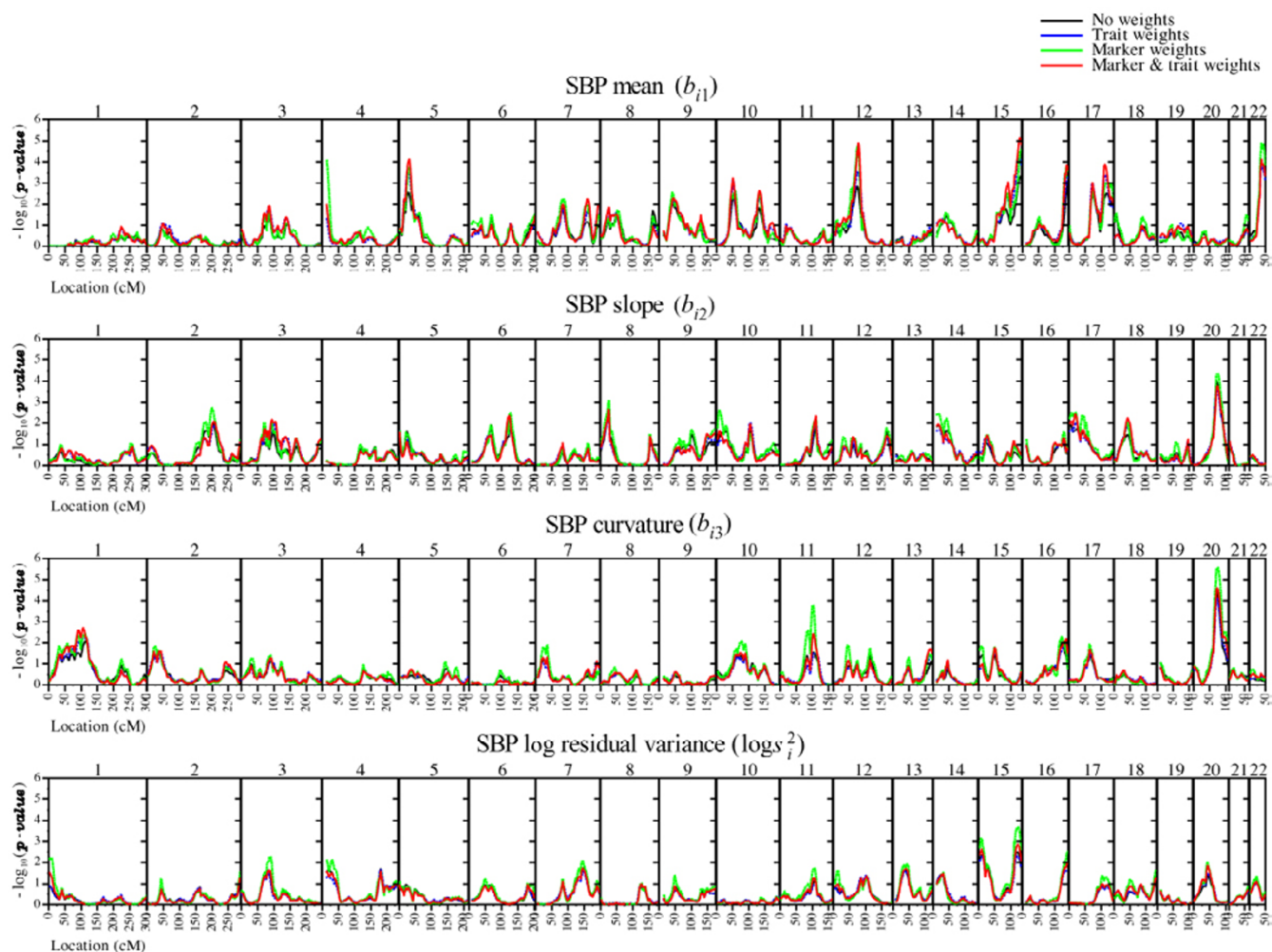
Figure 1
SBP genome scan results

treatment. We believe that these three criteria provide a better estimate of the SBP without treatment for hypertension than methods utilizing any of the criteria individually, and possibly even the imputation method of Levy et al.

Conclusions
Weighting our regression models by marker and trait informativity increased the power to detect linkage (Figure 1). In several of the significant regions, the *p*-value was decreased by at least an order of magnitude when including marker, trait, or both weights. Though we were not able to conduct an analysis directly comparable to that performed by Levy et al. [2], we present significant results that confirm a genetic component to hypertension and suggest the location of those genes to be on chromosomes

5, 12, 15, 17, and 20, and suggestive evidence of genes on chromosomes 16 and 22. Further, we present straightforward yet effective methods to adjust for treatment effects and to increase the power of linkage tests with siblings of varying marker and trait informativity.

Acknowledgments

References
1. Timberlake DS, O'Connor DT, Parmer RJ: **Molecular genetics of essential hypertension: recent results and emerging strategies.** *Curr Opin Nephrol Hypertens* 2001, **10:**71-79.

2. Levy D, DeStefano AL, Larson MG, O'Donnell CJ, Lifton RP, Gavras H, Cupples LA, Myers RH: **Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the Framingham heart study.** *Hypertension* 2000, **36:**477-83.
3. Dawber TR, Meadors GF, Moore FEJ: **Epidemiological approaches to heart disease: the Framingham Study.** *Am J Public Health* 1951, **41:**279-286.
4. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP: **An investigation of coronary heart disease in families: the Framingham Offspring Study.** *Am J Epidemiol* 1979, **110:**281-290.
5. Robson DS: **A simple method for constructing orthogonal polynomials when the independent variable is unequally spaced.** *Biometrics* 1959, **15:**187-191.
6. Statistical Solutions Ltd: **S.A.G.E. Statistical Analysis for Genetic Epidemiology, Release 4.3.** *Cork, Ireland, Statistical Solutions Ltd* 2002.
7. Shete S, Jacobs S, Elston RC: **Adding further power to the Haseman and Elston method for detecting linkage.** *Genet Epidemiol* 1999, **17:**194.
8. Elston RC, Buxbaum S, Jacobs KB, Olson JM: **Haseman and Elston revisited.** *Genet Epidemiol* 2000, **19:**1-17.
9. Xu X, Weiss S, Wei LJ: **A unified Haseman-Elston method for testing linkage with quantitative traits.** *Am J Hum Genet* 2000, **67:**1025-1028.