*Corresponding author:
E-mail: wchung@ssu.ac.kr

Wonil Chung and Hyunji Hwang
contributed equally to this work.

Bayesian analysis of longitudinal traits in the Korea Association Resource (KARE) cohort

Wonil Chung[1,2]*, Hyunji Hwang[1], Taesung Park[3]

[1]Department of Statistics and Actuarial Science, Soongsil University, Seoul 06978, Korea
[2]Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[3]Department of Statistics, Seoul National University, Seoul 08826, Korea

Various methodologies for the genetic analysis of longitudinal data have been proposed and applied to data from large-scale genome-wide association studies (GWAS) to identify single nucleotide polymorphisms (SNPs) associated with traits of interest and to detect SNP-time interactions. We recently proposed a grid-based Bayesian mixed model for longitudinal genetic data and showed that our Bayesian method increased the statistical power compared to the corresponding univariate method and well detected SNP-time interactions. In this paper, we further analyze longitudinal obesity-related traits such as body mass index, hip circumference, waist circumference, and waist-hip ratio from Korea Association Resource data to evaluate the proposed Bayesian method. We first conducted GWAS analyses of cross-sectional traits and combined the results of GWAS analyses through a meta-analysis based on a trajectory model and a random-effects model. We then applied our Bayesian method to a subset of SNPs selected by meta-analysis to further discover SNPs associated with traits of interest and SNP-time interactions. The proposed Bayesian method identified several novel SNPs associated with longitudinal obesity-related traits, and almost 25% of the identified SNPs had significant p-values for SNP-time interactions.

Keywords: Bayesian mixed model, KARE data, longitudinal data, obesity-related traits

Introduction

Genome-wide association studies (GWAS) have been extensively performed to identify single nucleotide polymorphisms (SNPs) associated with cross-sectional traits, but GWAS on longitudinal traits have been underexplored [1–5]. Longitudinal traits dynamically change over time, controlled by both genetic and environmental factors. Multiple measurements at various time points under varying environmental conditions are collected as longitudinal traits; thus, conducting a GWAS for longitudinal traits requires considering the covariance among observations at all-time points. Multivariate linear mixed models have been widely adopted for longitudinal GWAS, and efficient multivariate algorithms for longitudinal GWAS have been developed to reduce the computational complexity for SNP-inferred kinship matrix [3,6]. Furthermore, several Bayesian methods such as time-varying-coefficient regression [7], a simple regression-based method [8], and a Gaussian process-based model [9] have been developed for analyzing such longitudinal traits.

Recently, for the analysis of longitudinal genetic data, we developed a Bayesian mixed model with a built-in variable selection feature based on a grid-based covariance estimation approach [10]. The proposed Bayesian method modeled multiple candidate SNPs simultaneously and allowed SNP-SNP and SNP-time/environment interactions, enabling us to detect SNPs with time-varying effects that were of great scientific and medical interest. The proposed grid-based approach modeled the covariance structure nonparametrically; not only was this parsimonious in estimating the covariance matrix, but it also enabled the flexible approximation of any type of covariance structure by employing a reasonable number of grid points. The number of grid points was set in advance, but the deviance information criterion (DIC) [11] and simplified Bayesian predictive information criterion (BPIC) [12] can be utilized to select the optimal number of grid points. Through simulation studies, we showed that the proposed Bayesian method using all-time points outperformed the ordinary Bayesian method with one or all-time points, and statistical power increased as the data had more samples, smaller number of SNPs, a lower proportion of causal SNPs, and larger trait-heritability.

In this paper, we analyze longitudinal obesity-related traits such as body mass index (BMI), hip circumference (HIP), waist circumference (WST), and waist-hip ratio (WHR) from Korea Association Resource (KARE) data to further evaluate the proposed Bayesian method. These traits are well-known markers of obesity across all ages and continuously change over time. We first conduct cross-sectional GWAS analysis at each time point and combine the results of the GWAS analysis of all-time points using a meta-analysis based on a trajectory model [13] and a random-effects model [14]. We then apply our Bayesian mixed model to further discover SNPs associated with the traits and SNP-time/environment interactions. The paper is organized as follows. In Methods section, we describe the KARE data, GWAS analysis, meta-analysis and summarize our grid-based Bayesian mixed model for longitudinal genetic data. In the Results section, we present the analysis results for longitudinal obesity-related traits such as BMI, HIP, WST, and WHR from KARE data and evaluate the proposed Bayesian method. We conclude the paper with a discussion on Bayesian longitudinal analysis and future work.

Methods

Study participants

The Korean Genome and Epidemiology Study (KoGES) was a large prospective cohort study that was initiated to solve public health issues and prepare for personalized and preventive health care in Korea [15]. The purpose of the KoGES was to build a genome epidemiological study platform for various researchers and examine the genetic and environmental etiology of complex traits and common diseases such as type 2 diabetes, hypertension, obesity, and metabolic syndrome in Korea. Data collection was initiated in 2001, and follow-up examinations for the participants have been conducted every 2 years. In particular, the availability of GWAS data and repeatedly measured traits in the KoGES Ansan-Ansung study facilitated the identification of genetic variants associated with various disease traits. The GWAS for the KoGES was known as the KARE cohort, which was launched to perform GWAS to discover the underlying genetic variants associated with diverse complex traits and diseases. The KARE data included 10,038 unrelated individuals aged 40–69 years, assembled through the KoGES Ansan-Ansung study, representing urban and countryside populations, respectively [16]. Our analyses utilized the KARE dataset from the baseline (2001–2002) to the sixth (2013–2014) follow-up.

Genotype data

The 10,038 participants in the KARE were genotyped at ~500,000 SNPs using the Affymetrix Genomewide Human SNP array 5.0. For SNP quality control (QC), we removed SNPs with minor allele frequency (MAF) < 0.05, genotype calling rates < 95%, and Hardy-Weinberg equilibrium p-values < 10^{-6}, resulting in 4,518,929 SNPs. For sample QC, we only preserved participants with consistent sex and calling rates > 90%, resulting in 9,331 individuals. The SNP QC and sample QC were performed using PLINK (v1.90) [17]. We then imputed the genotyped SNP data using the Beagle 5.0 software [18] after the SNP and sample QCs. We restricted the analysis to Hapmap3 SNPs with MAF > 0.05 and imputation $R^2 > 0.80$, which consisted of 3,848,960 SNPs.

GWAS analysis

In the GWAS for cross-sectional obesity-related traits such as BMI, HIP, WST, and WHR, we performed linear mixed model (LMM) association analysis implemented in the BOLT-LMM (v2.3) software [19] using genotyped and imputed genetic variants from the KARE data [20]. The current analysis adjusted for age, age squared, sex, drinking, smoking status, and 10 genotype principal components (PCs). In order to obtain independent SNPs (i.e., a smaller number of clumps of correlated SNPs), we reduced the genome-wide scan using the PLINK clumping function based on empirical estimates of linkage disequilibrium (LD) between genetic variants loci (with four main flags: --clump-p1 5e-8 --clump-p2 1e-5 --clump-r2 0.1 --clump-kb 1000). The clumping

procedure found index SNPs ($p < 5 \times 10^{-8}$) that were independent from each other and formed clumps of other SNPs that were located within 1,000 kb from each index SNP and in LD with the index SNP based on $r^2 > 0.1$. Clumps annotated to the same gene were further combined to the clump with the most significant index SNP.

Meta-analysis using trajectory model

To discover SNPs associated with the overall trend and trajectory of phenotypes, we prepared two different longitudinal outcomes (the average and trajectory), which were obtained from longitudinal BMI, HIP, WST, and WHR measurements (baseline to sixth follow-up). To assess the average outcome, we calculated the average BMI, HIP, WST, and WHR measurements over the six time points and then used the average values as the phenotype for GWAS analysis. To assess the trajectory outcome, we utilized the methodology in Gouveia et al. [13] designed to study cognitive trajectories. The LMM function implemented in the R package lme4 [21] was used to obtain the per-individual trajectory for all obesity-related traits. Specifically, we fitted the following mixed model for subject i at time j: $y_{ij} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i})x_{1ij} + \sum_{h=2}^{q}\beta_h x_{hij} + e_{ij}$ ($i = 1, \ldots, n, j = 1, \ldots, 6$) where y_{ij} is a phenotypic value; u_{0i} is a random effect allowing variation around the intercept β_0; u_{1i} is a random effect allowing variation around the slope for age covariate, β_1; x_{hij} ($h = 1, \ldots, q$) are q covariates including age ($h = 1$) with fixed effects β_h; and e_{ij} is a random error. We assumed that the random intercept (u_{0i}) and the random slope (u_{1i}) were independent of each other and normally distributed. The models were adjusted for age squared, sex, drinking, and smoking status. We then used the predicted random slope (u_{1i}) from the model as the phenotype for a GWAS analysis of trajectory outcomes. All GWAS results for average and trajectory outcomes were obtained using BOLT-LMM software. In the GWAS for trajectory outcomes, the models were adjusted for only 10 genotype PCs to avoid over-adjustment.

Meta-analysis using a random-effects model

We utilized two different meta-analysis methods based on a fixed-effects model (Meta-Fixed) and a random-effects model (Meta-Random) to combine GWAS results for cross-sectional traits across all-time points (baseline to sixth follow-up). The Meta-Fixed method, as exemplified by the inverse-variance weighted method [22] and weighted sum-of-z-scores method [22], assumed that the magnitude of the true effects was common or fixed in every GWAS study and did not model heterogeneity. p-values for all GWAS analyses were converted to z-scores and then combined with different weights for each study according to their sam-

ple sizes. Meanwhile, the Meta-Random method proposed by DerSimonian and Laird [23] and Lee et al. [14] assumed that the true effect size of each study was sampled from an underlying distribution and modeled heterogeneity, explicitly. The Meta-Random method had two modifications: (1) accounting for correlation among test statistics based on the Lin and Sullivan method [24], and (2) focusing on heterogeneous effects conditioned on the Meta-Fixed method. The Meta-Random method can obtain higher power to detect heterogeneous effects than the Meta-Fixed method and it allows the testing of heterogeneous effect sizes between individual test statistics.

Grid-based Bayesian mixed model

To further discover SNPs associated with traits of interest and SNPs interacting with time/environmental factors, we considered the following Bayesian mixed model:

$$(1)$$

$$y_i = \mu_i + x_i^t\beta^t + x_i^g\beta^g + x_i^{gg}\beta^{gg} + x_i^{gt}\beta^{gt} + p_i v_i + e_i = \mu_i + x_i\beta + p_i v_i + e_i (i = 1, \ldots, n),$$

where $y_i = (y_{i1}, \ldots, y_{in_i})^T$ is an $n_i \times 1$ phenotype vector of individual i; $\mu_i = \mu 1_{n_i}$ is an $n_i \times 1$ overall mean vector; $x_i = (x_i^t, x_i^g, x_i^{gg}, x_i^{gt})$ is the design matrix corresponding to q time/environmental covariates, p SNPs, two-way interactions among p SNPs (resulting in total of $p(p-1)/2$ terms), and pq SNP-time/SNP-environment interactions; $\beta = (\beta^{tT}, \beta^{gT}, \beta^{ggT}, \beta^{gtT})^T$ is a vector of genetic effects, time/environmental effects, epistasis effects, and SNP-time/environment interactions; e_i is an $n_i \times 1$ vector of random errors with $e_i \sim N(0, \sigma^2 I_{n_i})$. For a given trait, we had n individuals, where individual i has phenotypic values measured at n_i time points ($i = 1, \ldots, n$), and the total number of observations was $N = \sum_{i=1}^{n} n_i$. To model the correlation among the repeated measurements of the same individual, we partitioned the observed time interval by k pre-specified grid points, $t = (t_1, \ldots, t_k)^T$, and defined v_i as a $k \times 1$ vector of random effects at the grid time points with $v_i \sim N(0, D)$, where D is a $k \times k$ covariance matrix. We also defined the incidence matrix p_i as $p_i = (p_{i1}^T, \ldots, p_{in_i}^T)^T$. If all subjects had k observations measured exactly on the k grid time points, then p_i became an identity matrix. We applied a linear interpolation procedure to any observation that did not fall on any one of k grid time points. When the jth measurement of individual i fell at time t, which was in between the grid points t_r and t_{r+1} ($t_r \leq t \leq t_{r+1}$),

we set $p_{ij} = (0_{(r-1)}^T, \frac{t_{r+1}-t}{t_{r+1}-t_r}, \frac{t-t_r}{t_{r+1}-t_r}, 0_{(k-r-1)}^T)$. When t = t_r, we get $p_{ij} = (0_{(r-1)}^T, 1, 0_{(k-r)}^T)$.

For Bayesian estimation of the mixed-effects model (1), we ap-

plied the modified Cholesky decomposition of Chen and Dunson [25] to the $k \times k$ covariance matrix D, resulting in the decomposition, $D = \Delta \Psi \Psi^T \Delta$, where Δ is a nonnegative $k \times k$ diagonal matrix and Ψ is a $k \times k$ lower triangular with 1's in the diagonal elements. We re-parameterized model (1) as

$$y_i = \mu_i + x_i\beta + p_i\Delta\Psi b_i + e_i (i = 1, \dots, n). \qquad (2)$$

where $b_i = (b_{i1}, \dots b_{ik})^T$ such that $b_{ij} \sim N(0,1)$ and $b_{ij} \perp b'_{ij}(j \neq j')$, $j = 1, \dots, k$.

As priors for Δ and Ψ, we defined two vectors $\delta = (\delta_l : l = 1, \dots, k)^T$ and $\psi = (\psi_{ml} : m = 2, \dots, k; l = 1, \dots, m-1)^T$. The prior distribution for δ is $P(\delta) = \prod_{l=1}^{k} P(\delta_l) = \prod_{l=1}^{k} N^+(\delta_l | m_{l0}, s_{l0}^2)$ where $N^+(\delta_l | m_{l0}, s_{l0}^2)$ was the density of a half normal distribution that was a $N(\delta_l | m_{l0}, s_{l0}^2)$ density truncated below by zero. The prior distribution for ψ was $P(\psi) \stackrel{d}{=} N(\psi_0, R_0)$ where ψ_0 and R_0 are pre-specified hyperparameters. The joint prior distribution of $b = (b_1^T, \dots, b_n^T)^T$ was $P(b) \stackrel{d}{=} N(0, I_{nk})$. The prior for the ath genetic effect was a normal distribution, $P(\beta_a | \gamma_a, \sigma_\beta^2) \stackrel{d}{=} N(0, \gamma_a \sigma_\beta^2)$ and the prior for the variance σ_β^2 was a scaled inverse χ^2 distribution. The prior for the overall mean μ was given by $P(\psi) \stackrel{d}{=} N(\eta_0, \tau_0^2)$. The prior for the residual variance σ^2 was chosen as a scaled inverse χ^2 distribution, $P(\sigma^2) \stackrel{d}{=} \text{inv} - \chi^2(v_\sigma, s_\sigma^2)$. The joint posterior distribution was proportional to the product of the likelihood and the prior distributions of all unknown parameters. We utilized Metropolis-Hastings and Gibbs sampling algorithms, and alternately updated each unknown parameter or set of unknown parameters conditional on all the other parameters and the observed data.

The posterior samples can be used to approximate the posterior distribution of the parameters. The posterior inclusion probability of each SNP was calculated using its inclusion proportion in the Markov chain Monte Carlo samples. The Bayes factor (BF) can be calculated to quantify the evidence for the inclusion of a specific SNP against the exclusion of that SNP. Jeffreys [26] and Yandell et al. [27] suggested the following criteria for judging the significance of each SNP: weak support if the BF falls between 3 and 10; moderate support if the BF falls between 10 and 30; and strong support if the BF is larger than 30 (i.e., $\log_2(BF) > 6.8$). A critical issue with the proposed Bayesian model was how to choose an optimal number of grid points, k. We achieved the goal by evaluating the goodness of the predictive distributions of our Bayesian models. Spiegelhalter et al. [11] proposed the DIC and Ando [12] developed the BPIC. We select the optimal number of grid points for our model by minimizing the DIC or simplified BPIC scores. The proposed methods were implemented in an R package named Gridbayes [10], which is available for download at https://github.com/wonilchung/GridBayes.

Results

GWAS analysis

We conducted a GWAS analysis of KARE data on longitudinal obesity-related phenotypes such as BMI, HIP, WST, and WHR. Seven covariates (age, age squared, sex, area, drinking, smoking status and 10 genotype PCs) were included in the analysis. Among 9,331 individuals, we selected 4,621 individuals who had no missing values for either phenotypes or covariates. The number of measurements for each individual was six, and thus the total number of observations was 27,726. Hapmap3 SNPs with MAF > 0.05 and imputation $R^2 > 0.8$ were retained for the GWAS analyses, resulting in 3,848,960 SNPs.

GWAS analyses for cross-sectional obesity-related phenotypes were performed using LMM association analysis implemented in the BOLT-LMM. To assess the cross-sectional outcomes, we analyzed all measurements (baseline to sixth follow-up) separately for each phenotype. Fig. 1 displays Manhattan plots for the baseline BMI, HIP, WST, and WHR measurements and Supplementary Figs. 1–5 in the supplementary materials show the Manhattan plots for the second, third, fourth, fifth, and sixth measurements. Supplementary Figs. 6–11 show the corresponding Quantile-Quantile (Q-Q) plots. For BMI, one SNP on chromosome 6 (baseline) reached GWAS significance ($p < 5 \times 10^{-8}$), and for HIP, five SNPs on chromosome 11 (baseline), one SNP on chromosome 12 (third follow-up), two SNPs on chromosome 12 (fourth follow-up), and one SNP on chromosome 6 (sixth follow-up) exceeded the threshold for GWAS significance in Supplementary Table 1. For WST, 11 SNPs on chromosome 11 (baseline) and for WHR, 16 SNPs on chromosome 12 (second follow-up), one SNP on chromosome 12 (fifth follow-up), one SNP on chromosome 18 (second follow-up), and one SNP on chromosome 18 (fifth follow-up) reached GWAS significance (see Supplementary Table 1).

Meta-analysis

In order to identify SNPs associated with the overall trend and trajectory of the phenotype, we utilized two types of longitudinal outcomes (average and trajectory), obtained for BMI, HIP, WST, and WHR measurements across consecutive examinations of each individual. For the average outcomes, we calculated the average of all measurements from baseline to sixth follow-up. For the trajectory outcomes, we estimated the per-individual trajectory for all obesity-related traits by fitting a mixed model in the R package
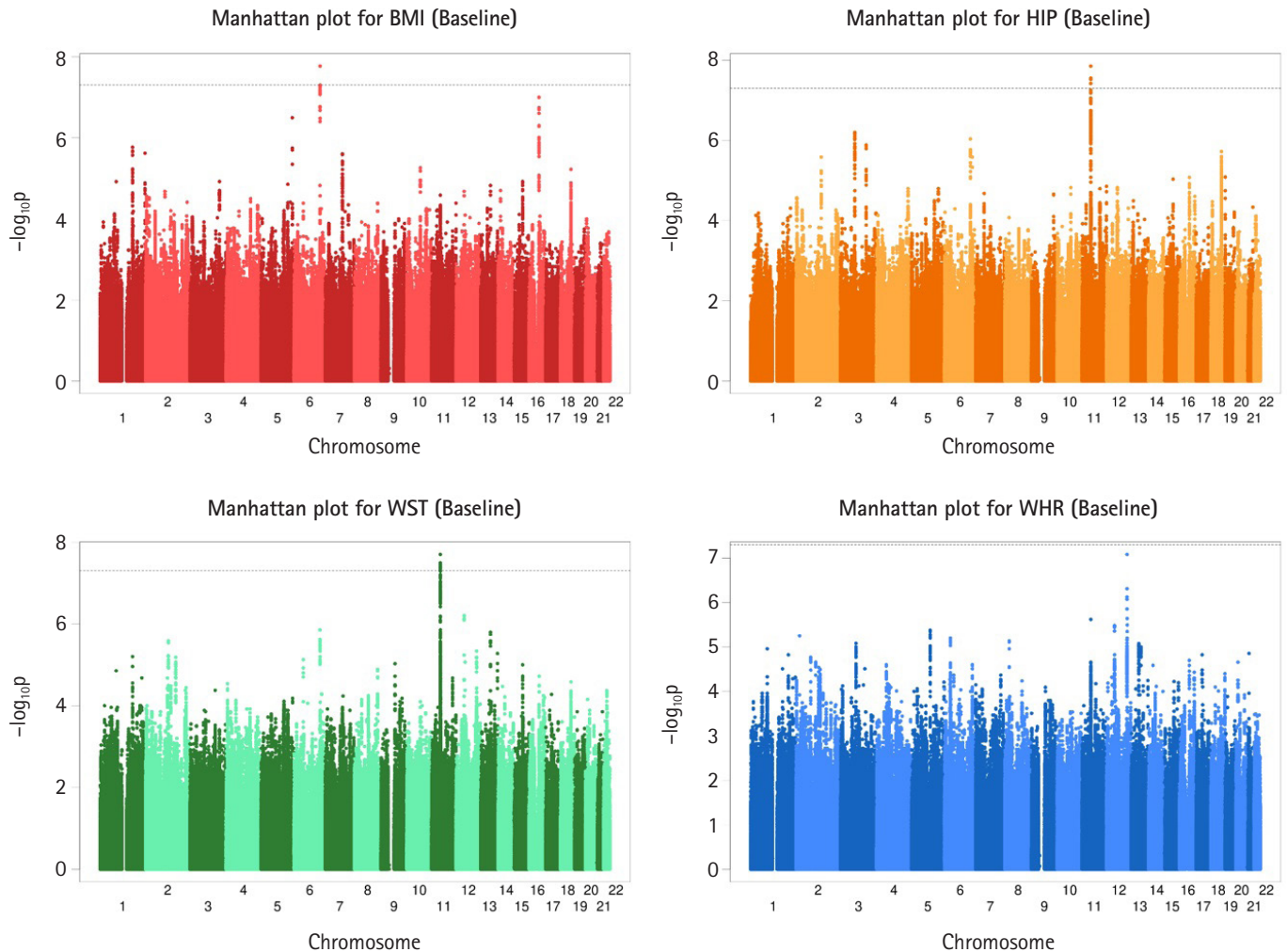
Fig. 1. Genome-wide Manhattan plots of −log₁₀(p-value) for association with baseline body mass index (BMI), hip circumference (HIP), waist circumference (WST), and waist-hip ratio (WHR) measurements using BOLT-LMM. For BMI, one single nucleotide polymorphism (SNP) on chromosome 6 reached genome-wide association studies (GWAS) significance, for HIP, five SNPs on chromosome 11 exceeded the threshold for GWAS significance and for WST, eleven SNPs on chromosome 11 reached GWAS significance.

lme4 [21]. The model contained a random intercept and a slope for age, as well as other covariates such as age squared, sex, area, and drinking and smoking status. The predicted random slopes of all individuals for age were used as the phenotype for GWAS on trajectory outcomes. We performed BOLT-LMM analyses with these average and trajectory outcomes for four obesity-related traits. Supplementary Fig. 12 shows Manhattan plots for the average outcomes, and Supplementary Fig. 13 shows Manhattan plots for the trajectory outcomes. Supplementary Figs. 14 and 15 show the corresponding Q-Q plots. No GWAS significant SNPs were found for either the average or trajectory outcomes, suggesting that a simple average and simple trajectory of longitudinal outcomes cannot effectively combine the cross-sectional traits; there-

fore, more sophisticated approaches are necessary.

To effectively combine all GWAS results for cross-sectional traits, we performed two different meta-analyses (Meta-Fixed and Meta-Random; see the Methods section) using RE2C software [14]. Because the Meta-Random method accounted for correlation among test statistics for all cross-sectional traits, it would be more appropriate than Meta-Fixed method for our analysis. Supplementary Fig. 16 shows Manhattan plots for the Meta-Fixed method, and Fig. 2 shows Manhattan plots for the Meta-Random method. Supplementary Figs. 17 and 18 show the corresponding Q-Q plots. From the results of Meta-Fixed method, for BMI, two SNPs on chromosome 5 and seven SNPs on chromosome 16 reached GWAS significance, and for WHR, five SNPs on chromo-
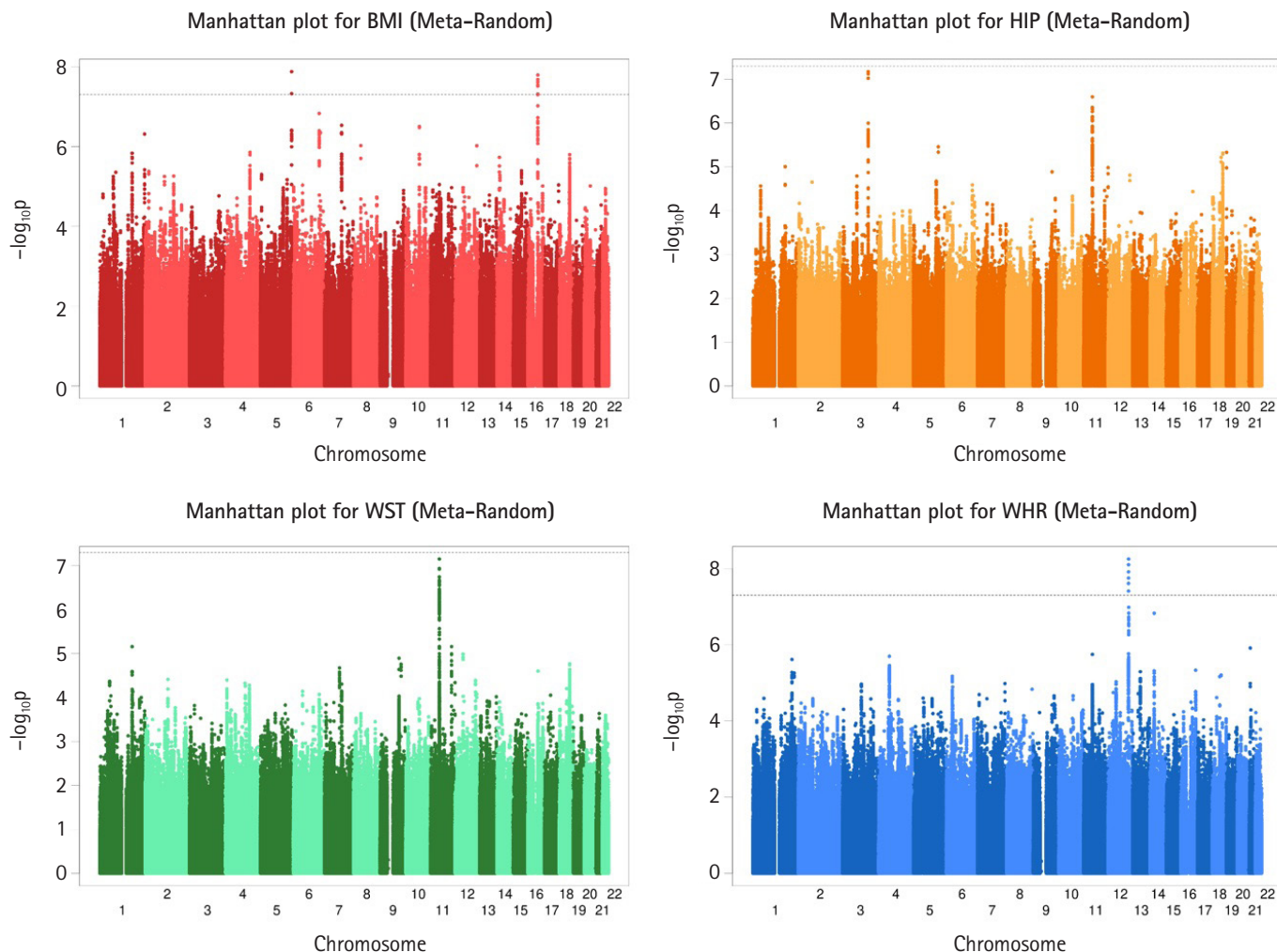
Fig. 2. Genome-wide Manhattan plots of $-\log_{10}$(p-value) for associations with longitudinal body mass index (BMI), hip circumference (HIP), waist circumference (WST), and waist-hip ratio (WHR) measurements using the Meta-Random method. Two single nucleotide polymorphisms (SNPs) on chromosome 5 and seven SNPs on chromosome 16 reached genome-wide association studies (GWAS) significance and for WHR, six SNPs on chromosome 12 reached GWAS significance.

some 12 reached GWAS significance (Supplementary Table 2). From the results of the Meta-Random method, for BMI, two SNPs on chromosome 5 and seven SNPs on chromosome 16 reached GWAS significance, and for WHR, six SNPs on chromosome 12 reached GWAS significance (Supplementary Table 3). Supplementary Fig. 19 shows the scatter plots of $-\log_{10}$(p-value) in GWAS for four longitudinal traits between the Meta-Random method and Meta-Fixed method. As expected, the p-values with the Meta-Random method were smaller than those obtained with the Meta-Fixed method, but there was reasonable concordance between them. In Fig. 3, it can be seen that p-values of the Meta-Random method were more significant than those of the baseline analysis in general.

Bayesian analysis

For our Bayesian analysis, a subset of SNPs was selected using LD clumping analysis based on the results of the Meta-Random method to only include SNPs that are not highly correlated with each other (with correlation < 0.5 to avoid multicollinearity). For our Bayesian analysis, we picked a list of the 4,124 SNPs from the LD clumping analyses (option: --clump -kb 250 −p1 0.001 −p2 0.01 − r2 0.5) using PLINK (v1.90) based on summary statistics from the Meta-Random analysis for the four obesity-related traits. Again, our Bayesian analysis included age, age squared, sex, area, drinking, smoking status, and 10 genotype PCs as covariates. The number of grid time points was set to 3 for the analysis based on the DIC and simplified BPIC scores with a different number of
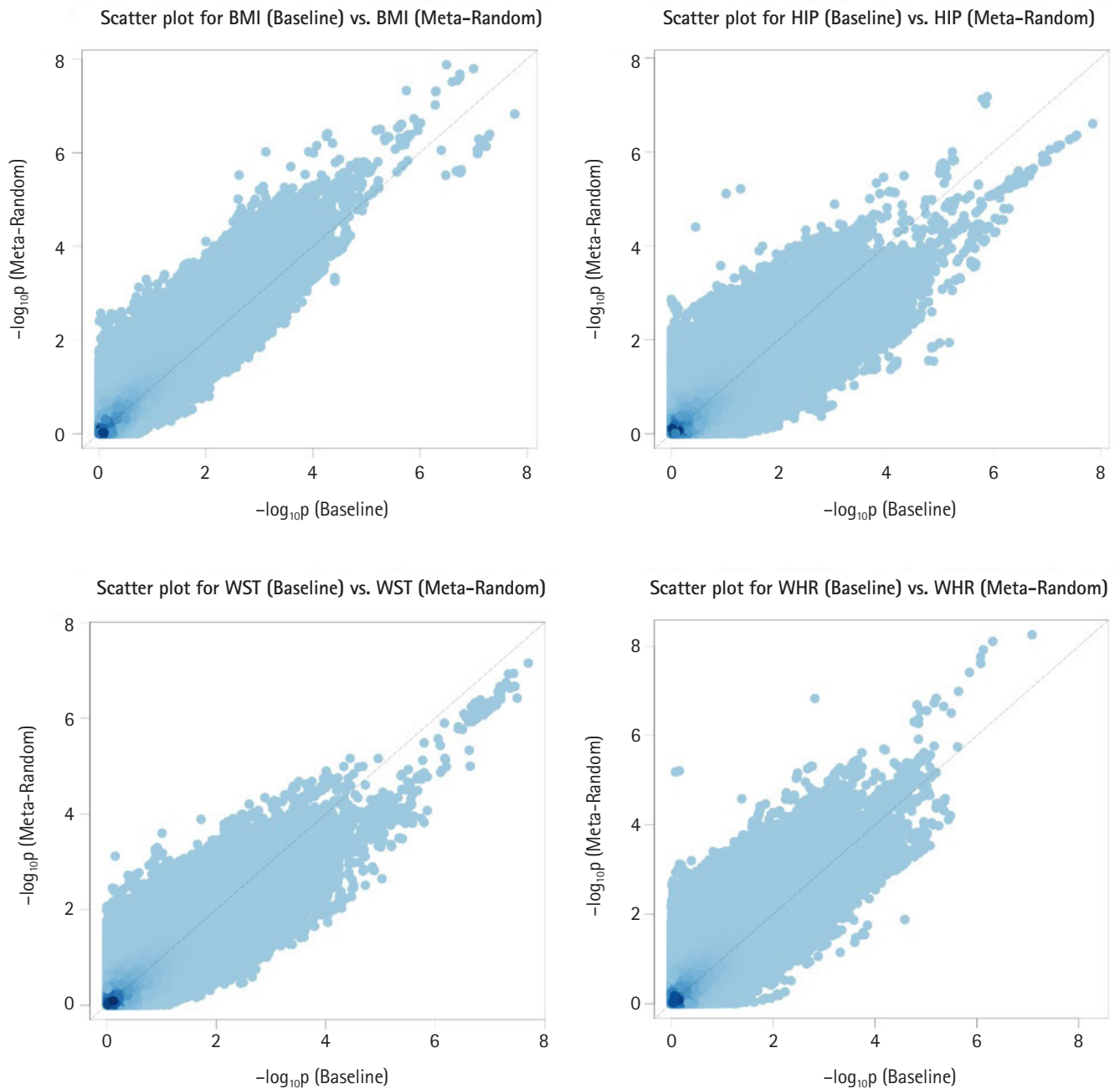
Fig. 3. Scatter plots of −log$_{10}$(p-value) for association for longitudinal body mass index (BMI), hip circumference (HIP), waist circumference (WST), and waist-hip ratio (WHR) measurements between the Meta-Random method and baseline analysis. The p-values of the Meta-Random method were more significant than those of the baseline analysis in general.

grid points (results not shown). Fig. 4 shows the one-dimensional genome-wide profiles of 2log(BF) for the combined effects (main, epistasis, and SNP-age interactions). Based on the criteria suggested in Jeffreys [26] and Yandell et al. [27], we found 112 SNPs with strong signals (i.e., $\log_2$(BF) > 6.8) for BMI, 20 SNPs for HIP, 10 SNPs for WST, and 6 SNPs for WHR in Supplementary Table 4.

We compared the results of our Bayesian analysis with p-values from the baseline, average, trajectory, and Meta-Random methods and discovered various newly detected novel loci and reasonable concordance among different approaches.

Also, our Bayesian method can discover SNPs interacting with environmental covariates including age, we needed to investigate
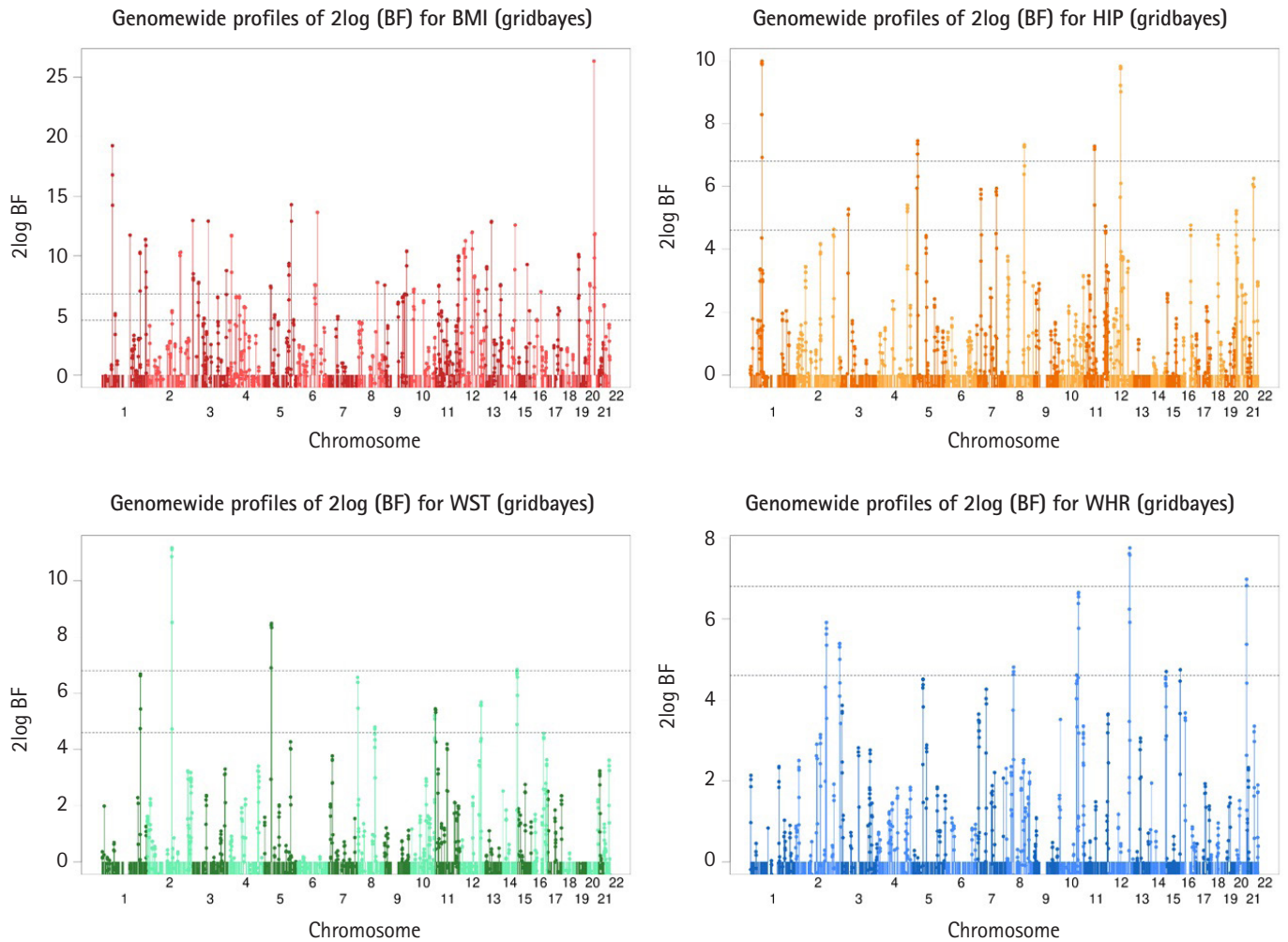
Fig. 4. Genome-wide profiles of 2log(BF) for all combined effects with body mass index (BMI), hip circumference (HIP), waist circumference (WST), and waist–hip ratio (WHR) measurements from Gridbayes. The two dashed horizontal lines represent the genome–wide thresholds for moderate (Bayes factor [BF]=10) strong (BF=30) genome–wide associations. We found 112 single nucleotide polymorphisms (SNPs) with strong signals for BMI, 20 SNPs for HIP, 10 SNPs for WST, and six SNPs for WHR.

whether the identified SNPs interacted with age covariate. For SNP-age interaction, we fitted the LMM with each of the four traits (BMI, HIP, WST, and WHR) as a response variable and eight covariates (SNP, age, age squared, sex, area, drinking and smoking status, SNP-age interaction) as predictors: $y_{ij} = (\beta_0 + u_{0i}) + \sum_{h=1}^{7} \beta_h x_{hij} + e_{ij} (i = 1, \ldots, n, j = 1, \ldots, 6)$ where y_{ij} is a phenotypic value, u_{0i} is a random effect allowing variation around the intercept β_0, β_h denotes genetic and nongenetic effects for eight covariate values x_{hij} (including SNP-age interaction), and e_{ij} is a random error. Based on p-values for SNP-age interaction in the LMM, we found 30 significant SNPs interacting with age for BMI, three SNPs for HIP, three SNPs for WST, and one SNP for WHR in Table 1, meaning that almost 25% of the identified SNPs using Gridbayes had significant p-values for SNP-age interaction. Based on the re-

sults of SNP-age interaction, as shown in Table 1, we displayed age-specific changes in BMI for three different genotypes (0, 1, and 2) at four SNPs (Chr5:149489242, Chr6:96377858, Chr10:18571215, and Chr20:40534573), age-specific changes in HIP at one SNP (Chr8:98187086) and age-specific changes in WST at one SNP (Chr2:133806432) in Fig. 5. These figures clearly showed different slopes in phenotypic values over six time points for three genotypes, meaning that there were SNP-age interactions at the identified genetic loci by Gridbayes.

Discussion

In this paper, we performed a meta-analysis based on GWAS results for cross-sectional obesity-related traits such as BMI, HIP,

Table 1. Genome-wide association results for BMI, HIP, WST, and WHR-associated SNPs with a p-value < 0.05 using the LMM method

Trait	Chr	Position	Minor	Major	MAF	Beta	SE	p(LMM)
BMI	13	55546160	C	T	0.3864	−0.1773	0.0467	3.53×10^{-4}
	13	55326244	A	G	0.3059	−0.1164	0.0503	3.53×10^{-4}
	8	108287752	C	T	0.1474	0.1386	0.0638	5.71×10^{-4}
	9	109243327	C	A	0.2209	0.0088	0.0580	1.24×10^{-3}
	12	97721824	G	A	0.3989	−0.0519	0.0484	1.71×10^{-3}
	4	17340809	A	G	0.1842	0.1939	0.0596	1.72×10^{-3}
	5	136606085	C	T	0.4770	−0.1044	0.0442	2.06×10^{-3}
	5	136608677	G	A	0.4774	−0.1075	0.0442	2.32×10^{-3}
	11	126243505	T	C	0.4071	0.1202	0.0480	2.89×10^{-3}
	3	88585081	A	C	0.0772	−0.1467	0.0860	3.94×10^{-3}
	20	37949055	G	A	0.2599	0.0571	0.0503	5.30×10^{-3}
	3	37832172	A	G	0.4460	−0.1209	0.0475	6.85×10^{-3}
	4	17378993	T	C	0.2553	0.1735	0.0520	8.00×10^{-3}
	9	119401114	T	C	0.4033	0.0356	0.0452	1.02×10^{-2}
	20	15637524	T	C	0.4354	0.1340	0.0519	1.18×10^{-2}
	13	105882996	C	T	0.3231	−0.0168	0.0550	1.45×10^{-2}
	20	43260694	C	T	0.1624	−0.1104	0.0613	1.82×10^{-2}
	6	96377858	T	C	0.2070	−0.2001	0.0601	1.96×10^{-2}
	3	186711098	T	G	0.4910	−0.1182	0.0454	2.13×10^{-2}
	5	149489242	T	C	0.1006	0.1105	0.0785	2.25×10^{-2}
	10	18571215	G	A	0.3777	−0.1505	0.0495	2.37×10^{-2}
	20	40534573	T	C	0.3207	−0.1388	0.0488	2.56×10^{-2}
	3	5500434	T	C	0.2601	0.0894	0.0521	2.70×10^{-2}
	5	136484839	G	T	0.2642	0.1446	0.0519	2.76×10^{-2}
	20	16840158	A	G	0.3718	−0.1600	0.0521	3.02×10^{-2}
	13	104745772	A	G	0.0685	0.2781	0.0892	3.19×10^{-2}
	9	119383320	A	C	0.3169	−0.0684	0.0479	3.42×10^{-2}
	1	237458209	G	A	0.4012	−0.1363	0.0493	3.75×10^{-2}
	12	78510165	G	T	0.4001	−0.1467	0.0455	4.02×10^{-2}
	11	20359771	C	A	0.3368	−0.1069	0.0523	4.65×10^{-2}
HIP	5	28989968	G	T	0.1186	−0.1733	0.0700	6.01×10^{-3}
	8	98187086	A	G	0.1649	0.1739	0.0612	8.60×10^{-3}
	1	64914487	G	A	0.0839	0.2533	0.0816	4.83×10^{-2}
WST	2	133232051	A	G	0.0532	0.1073	0.1009	3.05×10^{-4}
	5	41264019	C	T	0.1530	−0.1240	0.0645	5.06×10^{-3}
	5	42569642	C	A	0.4095	−0.1104	0.0457	4.61×10^{-2}
	2	133806432	G	T	0.3763	0.0108	0.0506	4.96×10^{-2}
WHR	12	114085922	C	T	0.2865	0.1432	0.0508	3.23×10^{-2}

BMI, body mass index; HIP, hip circumference; WST, waist circumference; WHR, waist-hip ratio; SNP, single nucleotide polymorphism; LMM, linear mixed model.

WST, and WHR from KARE data and then applied our Bayesian method to a subset of SNPs selected by meta-analysis to further detect causal SNPs and SNPs with time-varying effects. To obtain the meta-analysis results, we first conducted GWAS analyses on obesity-related traits separately for each time point and then combined all GWAS results based on the average, trajectory, Meta-Fixed, and Meta-Random methods. For our Bayesian analysis, we first selected a subset of SNPs based on Meta-Random results and the LD clumping analyses and then applied our Bayesian method to those selected SNPs. We found our Bayesian method newly detected various novel loci and there was a reasonable concordance between our Bayesian analysis and the average, trajecto-
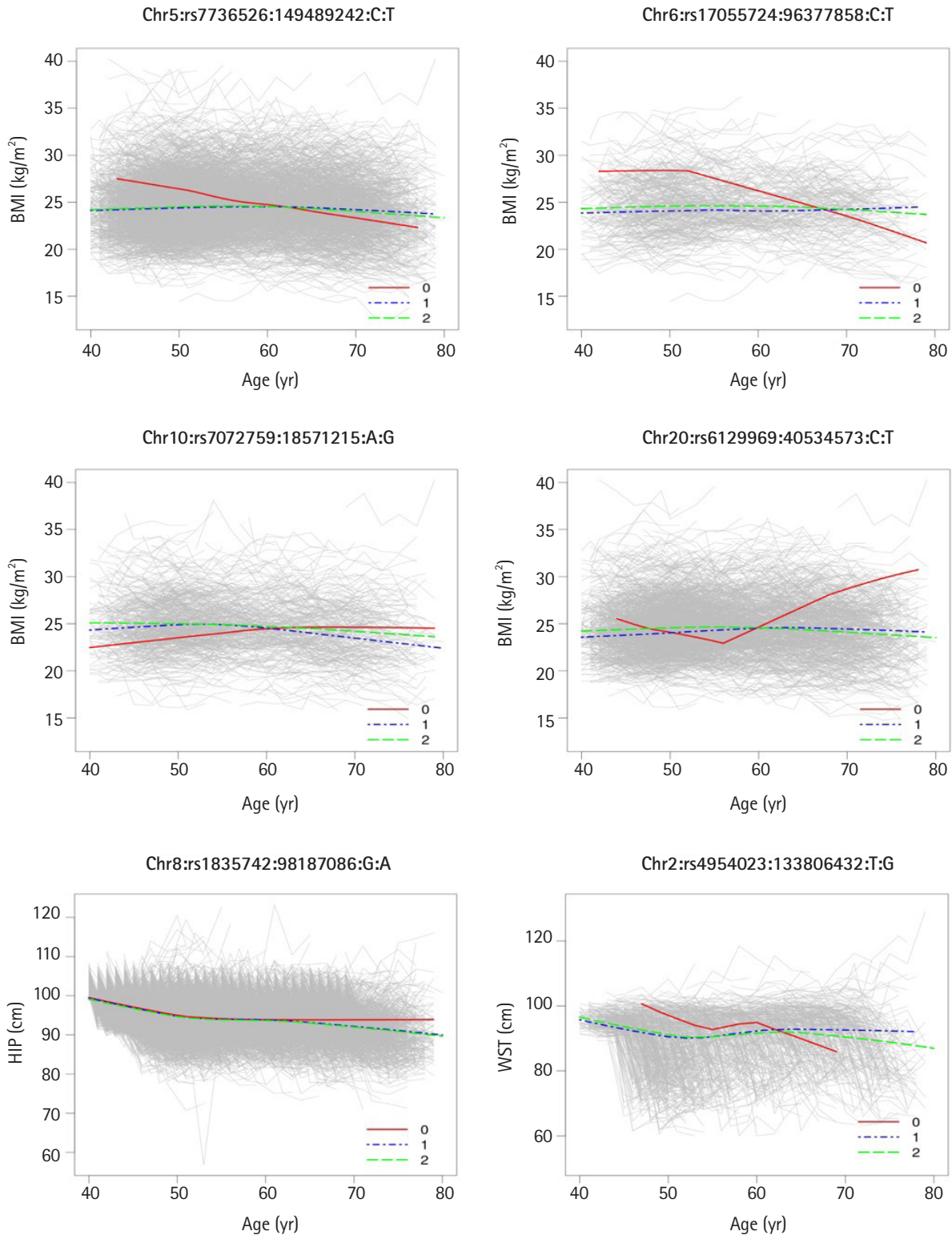
Fig. 5. Time-specific changes in body mass index (BMI) for three different genotypes (0, 1, and 2) at four single nucleotide polymorphisms (SNPs) (Chr5:149489242, Chr6:96377858, Chr10:18571215, and Chr20:40534573), age-specific changes in hip circumference (HIP) at one SNP (Chr8:98187086), and age-specific changes in waist circumference (WST) at one SNP (Chr2:133806432). The figures clearly show different slopes in phenotypic values over six time points for three genotypes, meaning that there are SNP-age interactions at identified the genetic loci.

ry, Meta-Fixed, and Meta-Random methods. We also confirmed that almost 25% of the identified SNPs using our Bayesian method had significant p-values for SNP-age interaction, meaning that our Bayesian method can well detect SNPs interacting with age as well as SNPs associated with traits of interest.

With the limited sample size, we restricted our Bayesian analysis to a subset of SNPs based on the BOLT-LMM analysis and LD clumping analysis. With a sufficient sample size, our method can be applied to all available SNPs. We are currently developing a parallel computing algorithm based on a message passing interface to execute multiple groups of SNPs simultaneously. This will make it feasible to apply our method to large-sample GWAS data.

One limitation of the proposed Bayesian model is its interpretability. It cannot readily conclude whether the identified SNPs had main effects and/or interactions with other covariates including time. Novel findings need to be further evaluated to elucidate their mode of action, as we performed the analysis for SNP-age interaction. If the number of identified SNPs is small, we can estimate the effect of each genotypic combination, which we can use to interpret the genetic mode. Our Bayesian model for GWAS data relied on a set of preselected SNPs. How to select a good set of SNPs, especially those with low marginal effects but high interactions with other SNPs or environmental factors, is challenging and deserves further investigation.

ORCID

Wonil Chung: https://orcid.org/0000-0002-5766-6247
Hyunji Hwang: https://orcid.org/0000-0002-5884-5394
Taesung Park: https://orcid.org/0000-0002-8294-590X

Authors' Contribution

Conceptualization: WC. Data curation: WC, HH, TP. Formal analysis: WC, HH. Funding acquisition: WC, TP. Methodology: WC. Writing - original draft: WC, HH. Writing - review & editing: WC, HH, TP.

Conflicts of Interest

Taesung Park serves as an editor of the Genomics and Informatics, but has no role in the decision to publish this article. All remaining authors have declared no conflicts of interest.

Acknowledgments

Supplementary Materials

Supplementary data can be found with this article online at http://www.genominfo.org.

References

1. Couto Alves A, De Silva NM, Karhunen V, Sovio U, Das S, Taal HR, et al. GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI. Sci Adv 2019;5:eaaw3095.

2. Vicuna L, Barrientos E, Norambuena T, Alvares D, Gana JC, Leiva V, et al. New insights from GWAS on longitudinal and cross-sectional BMI and related phenotypes in admixed children with Native American and European ancestries Preprint at https://doi.org/10.1101/2021.09.24.21263664 (2021).

3. Ning C, Wang D, Zhou L, Wei J, Liu Y, Kang H, et al. Efficient multivariate analysis algorithms for longitudinal genome-wide association studies. Bioinformatics 2019;35:4879-4885.

4. Chung W. Statistical models and computational tools for predicting complex traits and diseases. Genomics Inform 2021;19:e36.

5. Chung W, Chen J, Turman C, Lindstrom S, Zhu Z, Loh PR, et al. Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. Nat Commun 2019;10:569.

6. Ning C, Wang D, Zheng X, Zhang Q, Zhang S, Mrode R, et al. Eigen decomposition expedites longitudinal genome-wide association studies for milk production traits in Chinese Holstein. Genet Sel Evol 2018;50:12.

7. Gong Y, Zou F. Varying coefficient models for mapping quantitative trait loci using recombinant inbred intercrosses. Genetics

2012;190:475-486.

8. Kwak IY, Moore CR, Spalding EP, Broman KW. A simple regression-based method to map quantitative trait loci underlying function-valued phenotypes. Genetics 2014;197:1409-1416.

9. Chung W. Grid-based Gaussian process models for longitudinal genetic data. Commun Stat Appl Methods 2022;29:65-83.

10. Chung W, Cho Y. Bayesian mixed models for longitudinal genetic data: theory, concepts, and simulation studies. Genomics Inform 2022;20:e8.

11. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol 2002;64:583-639.

12. Ando T. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. Biometrika 2007;94:443-458.

13. Gouveia MH, Bentley AR, Leonard H, Meeks KA, Ekoru K, Chen G, et al. Trans-ethnic meta-analysis identifies new loci associated with longitudinal blood pressure traits. Sci Rep 2021;11:4075.

14. Lee CH, Eskin E, Han B. Increasing the power of meta-analysis of genome-wide association studies to detect heterogeneous effects. Bioinformatics 2017;33:i379-i388.

15. Kim Y, Han BG; KoGES Group. Cohort Profile: The Korean Genome and Epidemiology Study (KoGES) Consortium. Int J Epidemiol 2017;46:e20.

16. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ, et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat Genet 2009;41:527-534.

17. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-575.

18. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. Am J Hum Genet 2018;103:338-348.

19. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 2015;47:284-290.

20. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nat Genet 2018;50:906-908.

21. Bates D, Machler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4 Preprint at https://doi.org/10.48550/arXiv.1406.5823 (2014).

22. Cochran WG. The combination of estimates from different experiments. Biometrics 1954;10:101-129.

23. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-188.

24. Lin DY, Sullivan PF. Meta-analysis of genome-wide association studies with overlapping subjects. Am J Hum Genet 2009;85:862-872.

25. Chen Z, Dunson DB. Random effects selection in linear mixed models. Biometrics 2003;59:762-769.

26. Jeffreys H. Theory of Probability. 3rd ed. Oxford: Clarendon Press, 1961.

27. Yandell BS, Mehta T, Banerjee S, Shriner D, Venkataraman R, Moon JY, et al. R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. Bioinformatics 2007;23:641-643.