



NIH PUBLIC ACCESS

## Author Manuscript

*Genet Epidemiol.* Author manuscript; available in PMC 2010 August 2.

Published in final edited form as:

*Genet Epidemiol.* 2009 ; 33(Suppl 1): S93–S98. doi:10.1002/gepi.20479.

## Use of Longitudinal Data in Genetic Studies in the Genome-wide Association Studies Era: Summary of Group 14

Berit Kerner<sup>1</sup>, Kari E North<sup>2</sup>, and M Daniele Fallin<sup>3</sup><sup>1</sup>Department of Psychiatry, University of California, Los Angeles, CA<sup>2</sup>Department of Epidemiology and Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC<sup>3</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

### Abstract

Participants analyzed actual and simulated longitudinal data from the Framingham Heart Study for various metabolic and cardiovascular traits. The genetic information incorporated into these investigations ranged from selected single-nucleotide polymorphisms to genome-wide association arrays. Genotypes were incorporated using a broad range of methodological approaches including conditional logistic regression, linear mixed models, generalized estimating equations, linear growth curve estimation, growth modeling, growth mixture modeling, population attributable risk fraction based on survival functions under the proportional hazards models, and multivariate adaptive splines for the analysis of longitudinal data. The specific scientific questions addressed by these different approaches also varied, ranging from a more precise definition of the phenotype, bias reduction in control selection, estimation of effect sizes and genotype associated risk, to direct incorporation of genetic data into longitudinal modeling approaches and the exploration of population heterogeneity with regard to longitudinal trajectories. The group reached several overall conclusions: 1) The additional information provided by longitudinal data may be useful in genetic analyses. 2) The precision of the phenotype definition as well as control selection in nested designs may be improved, especially if traits demonstrate a trend over time or have strong age-of-onset effects. 3) Analyzing genetic data stratified for high-risk subgroups defined by a unique development over time could be useful for the detection of rare mutations in common multifactorial diseases. 4) Estimation of the population impact of genomic risk variants could be more precise. The challenges and computational complexity demanded by genome-wide single-nucleotide polymorphism data were also discussed.

### Keywords

phenotype definition; trends; risk estimation; growth modeling; sampling of controls

### Introduction

The identification of genetic risk factors is commonly based on case-control or cross-sectional phenotype data in mixed-age population samples. These study designs, however, do not allow for exploration of genetic and environmental risk factors that might influence

---

Corresponding author: M. Daniele Fallin, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Room W6509, Baltimore, MD 21205, [dfallin@jhsph.edu](mailto:dfallin@jhsph.edu).

**Conflict of Interest:** Dr. Kerner is a close collaborator and co-author with Bengt Muthén, who is the creator and distributor of the commercially available computer software program Mplus.

the development of traits over time. If the genetic risk factors are sensitive to time effects, the assignment of case and control status in genetic association studies may be biased.

Longitudinal data analysis in genetic studies is a unique, although rarely implemented, strategy that provides several advantages and challenges. The longitudinal design provides additional information regarding time of onset and therefore may allow for a more precise definition of case and control status in association analysis. The longitudinal information is particularly valuable for traits with variable age of onset and for traits that are heterogeneous with regard to development over time. Longitudinal studies also allow for the prospective measurement of time-varying factors that are not typically included in genetic case control association studies focused solely on genetic effects, even though interactions with time-varying environments may be important. This implies that time-variant covariates can be measured with improved precision in longitudinal data analyses. Finally, the genetic trait of critical interest may not be the occurrence of an event, but rather the trajectory of performance or decline over time. Longitudinal designs are uniquely able to capture this type of genetic trait via trajectory classes or quantitative slopes.

Besides design issues that address phenotype definition and population heterogeneity, longitudinal analysis also provides several advantages for the statistical analysis of genetic data. Instead of binary or cross-sectional quantitative traits, intercepts or slopes of repeated measures can be used as quantitative phenotypes, which may capture certain characteristics of population subgroups with unique genetic or environmental risk factors. The analysis could capitalize on non-linear shapes of the growth curves, knot-points, or other features of trajectories that may capture variability or stability over time.

In the Genetic Analysis Workshop 16 (GAW16), participants in Group 14 explored the advantages and challenges of a design that integrated either actual or simulated longitudinal phenotype data of the Framingham Heart Study (FHS) with genome-wide single-nucleotide polymorphism (SNP) data. The participating groups took a variety of approaches capitalizing on the additional information provided by this longitudinal study design. The approaches included use of incidence-density sampling rather than cross-sectional case-control selection, linear mixed modeling [McLean et al., 1991], generalized estimating equations [Zeger and Liang, 1986], multivariate linear growth curve modeling [Duncan et al., 1999], growth mixture modeling (GMM) [Muthén, 2004; Muthén and Asparouhov, 2008], population attributable risk fraction (PAF) based on survival functions under the proportional hazards model, and multivariate adaptive splines for the analysis of longitudinal data (MASAL) [Zhang, 1997; Zhang, 2004]. These contributions will be described in the rest of the paper.

## Methods

### Data

The Framingham Heart Study (FHS) is the oldest prospective longitudinal cohort study of cardiovascular risk factors in the U.S. [Cupples et al., 2007]. The original FHS cohort included presumably unrelated healthy individuals from Framingham, MA. About 5,209 subjects (Original Cohort) were recruited between 1948 and 1953 and were followed by biannual exams for potential cardiovascular risk factors (Cohort 1). The FHS was extended (Offspring Cohort) to ascertain children and spouses of the Original Cohort, enrolling an additional 5,124 individuals between 1971 and 1975 (Cohort 2). Finally, 4,095 children of the Offspring Cohort (the grandchildren of the original participants, Generation 3) extended the study over three generations (Cohort 3). Participants were examined in regular intervals; however, because the study spanned three generations, the age at each exam was not uniform. For GAW16, longitudinal data were available only for Cohort 1 and Cohort 2. In

contrast to other cohort studies, the participants of the FHS in the combined set of cohort collections are related to one another. Indeed, some families even consist of several hundred individuals. Each investigative team that participated in Group 14 selected different subsets of the data to address their specific hypothesis of interest, although all topics were related to longitudinal data analysis. To highlight their contributions, Group 14 participants' names are given in bold throughout.

In this paper, we will review how the Group 14 investigators applied general longitudinal data analyses: in all individuals in Cohorts 1 and 2 [Luan et al., 2009; Zhu et al., 2009], male individuals from Cohort 1 and Cohort 2 [Kerner and Muthén, 2009], incident cases and controls [Fradin and Fallin, 2009], unrelated individuals from Cohort 2 [Roslin et al., 2009; Yan et al., 2009], and select persons from Cohort 2 and Cohort 3 [Park et al., 2009] (Table I). Lastly, one group used simulated data with known results in order to compare the power of three different methodological approaches [Chang et al., 2009].

### Modeling family relationships

A key issue in the analysis of FHS data is the modeling of the non-independence amongst relative pairs. Four of the eight Group 14 contributions directly modeled the family structure of the data [Kerner and Muthén, 2009; Luan et al., 2009; Park et al., 2009; Zhu et al., 2009]. Zhu et al. [2009] split the families into sibships to reduce computational complexity. Three groups selected only unrelated individuals [Fradin and Fallin, 2009; Roslin et al., 2009; Yan et al., 2009].

### Genotypes and phenotypes

Integrating longitudinal data analysis with genome-wide SNP data caused computational complexities and challenges that were addressed by the different groups in unique ways (Table I). Several groups selected only SNPs or chromosomes with prior evidence of association to the trait of interest to demonstrate their approach to longitudinal data [Chang et al., 2009; Fradin and Fallin, 2009; Kerner and Muthén, 2009; Luan et al., 2009; Yan et al., 2009]. Three groups used genome-wide SNP data [Park et al., 2009; Roslin et al., 2009; Zhu et al., 2009]. The traits included in the analysis covered a wide range of metabolic and cardiovascular traits measured on FHS participants including systolic blood pressure (SBP), type 2 diabetes mellitus, coronary artery calcification, body mass index (BMI), weight, metabolic syndrome (MBS), high-density lipoprotein levels (HDL), low-density lipoprotein levels (LDL), fasting triglyceride levels (TG), and coronary artery disease (CHD) (Table I). The data were integrated either in a one-step approach, in which genotype and phenotype data were modeled simultaneously, or in a two-step approach that first modeled the phenotype and then integrated the genotype information in an independent analysis (Table I). All groups incorporated covariates into their analyses.

### Software

Several computer programs were used for the analysis including SAS [Chang et al., 2009; Fradin and Fallin, 2009; Luan et al., 2009; Park et al., 2009], Mplus [Chang et al., 2009; Kerner and Muthén, 2009; Roslin et al., 2009], PLINK 9 [Roslin et al., 2009], GOLDENHELIX [Kerner and Muthén, 2009], Stata [Luan et al., 2009], and MASAL [Zhu et al., 2009] (Table I).

### Overall Analytical Strategies and Results

The specific scientific focus of each group varied, with two overarching themes: more precise definition of the genetic phenotype in a longitudinal context and application of various statistical approaches for assessing the relationship between genes and longitudinal

phenotypes. The particular topics included assessment of changes in cross-sectional genetic effects over time, bias reduction in control selection using incidence density sampling, direct incorporation of genetic data in longitudinal modeling approaches, and the exploration of population heterogeneity with regard to longitudinal trajectories.

Because every group used a different analytical approach and a unique design to answer a very specific scientific question, a comparison of the methods and the results across projects was not feasible. We therefore will describe each study separately highlighting the significant results.

### Comparison of cross-sectional trends over time

Park et al. [2009] focused on the comparison of genetic association results for the synthetic trait MBS at multiple visits over time via cross-sectional association analyses. The investigators found that the significance of the SNP association increased over time. They were able to replicate some of the significant associations previously reported by FHS and identified new loci as well (Table II). They emphasized that attention to timing of case-control analysis and to consistency in trend of effects over time are important advantages of longitudinal data.

### Nested studies from longitudinal data collection

Fradin and Fallin [2009] compared two commonly employed methods of selecting age-matched controls for nested case-control studies, case exclusion sampling (CE), in which both prevalent and incident cases are included but only those free of disease at censoring are used as controls, and incidence density sampling (ID), in which only incident cases are included and controls are selected from all participants remaining at risk for the disease at the time a case occurs. The latter allows for those who develop disease at a later time to act as controls for cases occurring earlier in the study, which has been shown to be less biased than the case-exclusion approach in other areas of epidemiology. They used both the FHS data as well as simulated data and demonstrated that ID sampling was indeed associated with less bias in odds-ratio estimates than CE, although CE appears to be more powerful due to the upward bias of point estimates. They concluded that the ID sampling was an appropriate option for nested case-control genome-wide association studies and that this design could be a very efficient approach to obtain unbiased estimates of relative risk associated with genetic variants, especially when age is a strong risk factor for a disease phenotype. However, the small number of incident cases in the study limited this approach.

### Time-dependent PAF

Yan et al. [2009] focused on the population effect of candidate polymorphisms for CHD in the FHS Cohort 2 and the change of the population effect over time. They estimated the time-dependent PAF for each SNP using the Cox proportional hazards model and then estimated the PAF for all significant SNPs combined. Because the current PAF estimation does not account for age-of-onset data, this group was able to extend the current PAF analysis and to create a more comprehensive estimate of population impact over the life-course of disease by incorporating the age of onset of CHD into the formulation of the hazard function. They also explored the association with the risk score, which was constructed by summing the number of risk alleles across three CHD susceptibility SNPs (rs1333049 close to the *CDKN2A/2B* gene, rs618675 in the *GJA4* gene, and rs1376251 in the *TAS2R50* gene) out of 23 that had been significantly associated with the incident CHD (Table II). The risk score was significantly associated with incident CHD risk ( $p=0.0004$ ). They concluded that this novel tool for population impact may improve the understanding of genetic risk factors at the population level.

## Use of trajectories as genetic phenotype

**Splines**—Zhu et al. [2009] explored trajectories as genetic phenotype in a genome-wide modeling approach in independent individuals using MASAL [Zhang, 1997; Zhang, 2004]. This approach uses a nonparametric regression (forward/backward regression) to estimate regression coefficients that estimate gene, environment, time, and interaction effects. This method can accommodate time-varying covariates, and test for interactions between genes and environmental factors as well as between time and covariates. The group identified 13 significant SNP associations for Cohort 2 and 6 SNP associations for the Cohort 1 (Table II), as well as significant SNP-SNP and SNP-environment interactions. However, the significant SNPs in Cohort 1 and 2 samples did not overlap.

**Quantitative traits (intercepts and slopes)**—Luan et al. [2009] took a candidate gene approach to the longitudinal data, estimating the SNP effects on BMI and weight over time. The group fitted a mixed effects model and compared the intercepts and slopes of the growth curves in groups stratified by the genotypes of known risk alleles. This analysis confirmed the effect of the risk alleles of the SNPs in cross-sectional data (intercept results) and demonstrated a significant effect of the SNPs on the slopes of the developmental trajectories as well.

Roslin et al. [2009] attempted to incorporate the information provided by the longitudinal design into genome-wide SNP association analyses. For this approach they first estimated intercepts and slopes for four different traits in separate multivariate growth models of: 1) SBP, 2) HDL, 3) LDL, and 4) TG. Due to the computational resources required to process data for millions of SNP genotypes, the group followed a two-step design, by first estimating the individual intercepts and slopes of the multivariate growth curves and then regressing these (scalar) features on the SNP genotypes. Slopes and intercepts were regarded as separate traits in eight separate linear regression genetic analyses. The group found a significant association of SNP rs599839 on chromosome 1p13 with the intercept of the trait LDL but not with the slope. Marker rs765547 was associated with both the HDL slope and the TG slope (Table II). This marker is located about 41 kb downstream from the gene lipoprotein lipase (*LPL*) on chromosome 8p21.

**Trajectory classes**—Chang et al. [2009] used the genotype information of previously associated SNPs to predict latent class growth curves and latent class membership (the probability that a genotype affects the probability of the trajectory component membership). In order to accomplish this goal, they compared three different approaches, 1) the likelihood ratio test statistic (LRTS), 2) a direct test of genetic model coefficients, and 3) the chi-square test classifying subjects based on the trajectory model's posterior Bayesian probability. The group found that the LRTS was not usable due to non-normal distribution of the outcome and non-independence of the individuals in the study. The other two tests were satisfactory. Power was still substantial when markers near the gene rather than the gene itself were used. For markers near the actual gene, there was somewhat greater power for the direct test of the coefficients and lesser power for the posterior Bayesian probability chi-square test. Time-varying covariates did not increase overall power due to instability.

## Heterogeneity addressed by growth mixture modeling

Kerner and Muthén [2009] explored heterogeneity in the data with respect to SBP trajectories over time using GMM in Mplus. In order to incorporate genotype information into the model, they took a two-step approach. First, they estimated the class membership probability based on SBP development over time and then they used the latent class membership probability as phenotype in a quantitative trait analysis. This approach allowed for a more homogeneous classification of individuals with regard to the phenotype and it

facilitated the identification of an association with a rare coding variant in a subgroup of individuals. They concluded that GMM could be a useful approach to phenotype heterogeneity in mixed age samples for traits with strong age effects.

## Discussion

GAW16 participants revisited the FHS data for their analysis and participants of Group 14 in particular focused on the longitudinal aspect of the data. In previous workshops, the longitudinal focus had been on the use of intercepts and slopes as phenotypes in genetic analysis [Gauderman et al., 2003]. This year's contributions extended these approaches to include a variety of research questions for which longitudinal data could be useful. This exercise revealed several challenges, often related to handling large amounts of phenotype and genotype data and to limited numbers of events for many phenotypes. As the amount of genetic data in longitudinal studies available to researchers continues to increase, the challenges in attempting to incorporate longitudinal data analysis with genetic analysis will become increasingly critical.

Advantages of longitudinal data analysis include the possibility of finding more homogenous groups of individuals that share a common trajectory. Under this scenario, mean effects appear to be easier to detect than slopes or other trajectory features, and the longitudinal approach, borrowing information on individuals across time, may improve precision over analyses of a single visit. Longitudinal data analysis is advantageous for detecting genes that affect trajectories, rather than simple differences in phenotype values because the trajectory information provides a more specific trait for genetic analysis aimed at detecting such genes. Longitudinal data is also useful for identifying genetic causes that have strong age of onset effects because the use of age at disease onset will be a more specific phenotype for genetic analysis in this context than cross-sectional information about disease status. Certain traits have strong heterogeneity with regard to the underlying pathophysiological conditions dependent on the age of onset. SBP is a classic example. High blood pressure early in life often has its cause in renal pathology, whereas elevated blood pressure later in life is more commonly caused by cardiovascular changes. Attention to timing and trajectory of phenotype can help to clarify these insights.

Even though the longitudinal cohort study design has clear advantages for some questions, there are limitations. For example, analyses that focused on incident cases in cohort studies were underpowered due to a small number of events. Further, the results presented within this analysis group show the current lack of clear analytic strategies to deal with the complex longitudinal data structures. Not a single group was able to capitalize on all the data available: especially the integration of the large number of genotypes in the context of both repeated measures and family relationship. Future work should focus on the development of analytical methods and computer software that can handle these longitudinal data in the context of other complexities that are often found in cohort studies. The lack of software solutions that can handle millions of data points in a practical amount of time led many groups to a two-step design in which the phenotype data were analyzed separately from the genotype data. For example, the Mplus analyses performed by Kerner and Muthen [2009], which allow inclusion of up to 50 auxiliary variables, would have taken approximately 20,000 days (~ 55 years) for one-step analysis of 500,000 markers on a 64-bit 8-core machine. The two-step solution is far from ideal, however, and potentially biased. Other groups limited the analysis to only a small number of selected SNPs, assumed to be genetic risk factors for a given trait or phenotype. This design is also not ideal because it excludes the majority of SNPs that could potentially be important as risk factors in a particular study sample.



The GAW 16 data sets offered a unique opportunity to explore many approaches to longitudinal data analysis. From the diversity of strategies applied and evaluated in GAW16, especially those applied and evaluated as part of Group 14, it became apparent that there are many important methodological approaches available for implementing longitudinal data analysis. Unfortunately, the wide array of analyses performed in Group 14 did not allow for direct comparison of the different approaches. Future studies could suggest a more targeted effort in order to further evaluate the approaches and their usefulness under certain design conditions. Taken together, Group 14 contributions demonstrate the opportunities provided by longitudinal data and highlight the need for a combination of strategies to implement longitudinal data analysis.

## Acknowledgments

We thank all Group 14 participants for discussion during the GAW16 meeting and the GAW16 organizers for providing this opportunity. We also thank NIH for funding of this important exercise and meeting (R01 GM031575). Dr. Kerner's work on this manuscript was partially funded by NIMH K08 MH074057.

## References

- Chang SW, Choi SH, Li K, Saint Fleur R, Huang C, Shen T, Ahn K, Gordon D, Kim W, Wu R, Mendell NR, Finch SJ. Growth mixture modeling as an exploratory analysis tool in longitudinal quantitative trait loci analysis. *BMC Proc.* 2009; 3(Suppl 7):S112. [PubMed: 20017977]
- Cupples LA, Arruda HT, Benjamin EJ, D'Agostino RB Sr, Demissie S, DeStefano AL, Dupuis J, Falls KM, Fox CS, Gottlieb DJ, Govindaraju DR, Guo CY, Heard-Costa NL, Hwang SJ, Kathiresan S, Kiel DP, Laramie JM, Larson MG, Levy D, Liu CY, Lunetta KL, Mailman MD, Manning AK, Meigs JB, Murabito JM, Newton-Cheh C, O'Connor GT, O'Donnell CJ, Pandey M, Seshadri S, Vasan RS, Wang ZY, Wilk JB, Wolf PA, Yang Q, Atwood LD. The Framingham Heart Study 100K SNP genome-wide association study resource: Overview of 17 phenotype working group reports. *BMC Med Genet.* 2007; 8(Suppl 1):S1. [PubMed: 17903291]
- Duncan, TE.; Duncan, SC.; Strycker, LA.; Li, F.; Alpert, A. An introduction to latent variable growth curve modeling: Concepts, issues, and applications. Mahwah: Lawrence Erlbaum; 1999.
- Fradin DD, Fallin MD. Influence of control selection in genome-wide association studies: The example of diabetes in the Framingham Heart Study. *BMC Proc.* 2009; 3(Suppl 7):S113. [PubMed: 20017978]
- Gauderman WJ, Macgregor S, Briollais L, Scurrah K, Tobin M, Park T, Wang D, Rao S, John S, Bull S. Longitudinal data analysis in pedigree studies. *Genet Epidemiol.* 2003; 25(Suppl 1):S18–28. [PubMed: 14635165]
- Kerner B, Muthén BO. Growth mixture modeling in families of the Framingham Heart Study. *BMC Proc.* 2009; 3(Suppl 7):S114. [PubMed: 20017979]
- Luan J, Kerner B, Zhao JH, Loos RJJ, Sharp SJ, Muthén BO, Wareham NJ. A multilevel linear mixed model of the association between candidate genes and weight and body mass index using the Framingham longitudinal family data. *BMC Proc.* 2009; 3(Suppl 7):S115. [PubMed: 20017980]
- McLean RA, Sanders WL, Stroup WW. A unified approach to mixed linear models. *Am Stat.* 1991; 45:54–64.
- Muthén, B. Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan, EL., editor. *Handbook of quantitative methodology for the social sciences.* Newbury Park: Sage Publications; 2004. p. 345-68.
- Muthén, B.; Asparouhov, T. Growth mixture modeling: Analysis with non-Gaussian random effects. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G., editors. *Longitudinal data analysis.* Boca Raton: Chapman & Hall/CRC Press; 2008. p. 143-65.
- Park YM, Province MA, Gao X, Feitosa M, Wu J, Ma D, Rao DC, Kraja AT. Longitudinal trends in the association of metabolic syndrome with 550k single-nucleotide polymorphisms in the Framingham Heart Study. *BMC Proc.* 2009; 3(Suppl 7):S116. [PubMed: 20017981]

- Roslin NM, Hamid JS, Paterson AD, Beyene J. Genome-wide association analysis of cardiovascular-related quantitative traits in the Framingham Heart Study. *BMC Proc.* 2009; 3(Suppl 7):S117. [PubMed: 20017982]
- Yan Y, Hu Y, North KE, Franceschini N, Lin D. Evaluation of population impact of candidate polymorphisms for coronary heart disease in the Framingham Heart Study Offspring Cohort. *BMC Proc.* 2009; 3(Suppl 7):S118. [PubMed: 20017983]
- Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986; 42:121–30. [PubMed: 3719049]
- Zhang HP. Multivariate adaptive splines for analysis of longitudinal data. *J Comput Graph Stat.* 1997; 6:74–91.
- Zhang HP. Mixed effects multivariate adaptive splines model for the analysis of longitudinal and growth curve data. *Stat Methods Med Res.* 2004; 13:63–82. [PubMed: 14746441]
- Zhu W, Cho K, Chen X, Zhang M, Wang M, Zhang H. A genome-wide association analysis of Framingham Heart Study longitudinal data using multivariate adaptive splines. *BMC Proc.* 2009; 3(Suppl 7):S119. [PubMed: 20017984]



Table 1

Summary of data and methods used by Group 14 participants

Group	Cohort	Trait <sup>1</sup>	Genotypes	Analysis approach <sup>2</sup>	Software
Chang et al.	N/A	CAC	Simulated data	L1: GMM	SAS
Fradin et al.	1 and 2	T2D	Selected	C2: log regression, conditional on risk set	SAS
Kerner et al.	1 and 2 (males)	SBP	Chr. 8 (50k)	L2: GMM, QTL analysis	Mplus, Goldenhelix
Luan et al.	1 and 2	BMI, weight	Selected	L1: LMM	Stata, SAS
Park et al.	2 and 3	MBS	All	C1: GEE	SAS
Roslin et al.	2 (unrelated)	HDL, LDL, TG, SBP	500k	L2: MLGM, linear regression	Mplus, PLINK
Yan et al.	2 (unrelated)	CHD	Selected	L1: PAF	SAS
Zhu et al.	1 and 2	TG	Cell	L1: MASAL	MASAL

<sup>1</sup> SNP, single-nucleotide polymorphism; CAC, coronary artery calcification; T2D, type 2 diabetes; SBP, systolic blood pressure; BMI, body mass index; MBS, metabolic syndrome; HDL, high density lipoprotein; LDL, low density lipoprotein; TG, triglycerides; CHD, coronary heart disease

<sup>2</sup> L1, longitudinal analysis, one-step approach; L2, longitudinal analysis, two-step approach; C1, cross-sectional analysis, one-step approach; C2, case-control analysis by time-defined risk sets; GMM, growth mixture modeling; QTL, quantitative trait locus; LMM, linear mixed modeling; GEE, generalized estimating equations; MLGM, multivariate linear growth modeling; PAF, population attributable risk fraction; MASAL, multivariate adaptive splines for the analysis of longitudinal data

Table II

Summary of significant association results by trait, SNP, and chromosomal location

Group	SNP	Chromosome	Location	Gene <sup>1</sup>	Trait <sup>2</sup>	p-Value
<b>Kerner et al.</b>	rs1445404	8q13.3	72429137	<i>EYA1</i>	SBP	1.39×10 <sup>-13</sup>
<b>Luan et al.</b>	rs9939609	16q12.2	52377528	<i>FTO</i>	BMI	8.90×10 <sup>-8</sup>
<b>Park et al.</b>	rs17390167	1p36.32	3084691	<i>PRDM16</i>	MBS	3.80×10 <sup>-6</sup>
	rs13241465	7p14.3	33473074	<i>PTHBI</i>	MBS	1.40×10 <sup>-8</sup>
	rs4236337	7p14.3	33473431	<i>PTHBI</i>	MBS	2.80×10 <sup>-8</sup>
	rs4509212	7p14.3	33473170	<i>PTHBI</i>	MBS	6.20×10 <sup>-8</sup>
	rs318256	8q21.3	90283800		MBS	1.50×10 <sup>-7</sup>
	rs2418441	9q33.1	118114658	<i>PAPPA</i>	MBS	1.20×10 <sup>-8</sup>
	rs1431573	11q24.2	126981969		MBS	1.40×10 <sup>-7</sup>
	rs17037068	12q23.3	104333075		MBS	5.30×10 <sup>-10</sup>
	rs12437159	14q21.3	45882211		MBS	1.00×10 <sup>-9</sup>
	rs12147964	14q21.3	45888772		MBS	7.70×10 <sup>-11</sup>
	rs2899976	14q21.3	45893237		MBS	2.30×10 <sup>-10</sup>
	rs11625735	14q21.3	45895075		MBS	1.80×10 <sup>-10</sup>
	rs2010338	14q21.3	45895631		MBS	1.50×10 <sup>-10</sup>
	rs2415974	14q21.3	45906862		MBS	6.80×10 <sup>-11</sup>
	rs8009432	14q21.3	45917176		MBS	3.10×10 <sup>-10</sup>
	rs10498398	14q21.3	45937239		MBS	1.70×10 <sup>-9</sup>
	rs8019899	14q21.3	45952022		MBS	3.00×10 <sup>-9</sup>
	rs11508026	16q13	55556829	<i>CETP</i>	MBS	7.20×10 <sup>-7</sup>
	rs10408896	19p13.2	8124201	<i>FBN3</i>	MBS	8.80×10 <sup>-8</sup>
<b>Roslin et al.</b>	rs765547	8p21.3	19910054	near <i>LPL</i>	HDL slope	2.67×10 <sup>-6</sup>
	rs599839	1p13.3	109623189	<i>PSRC1</i>	LDL intercept	2.62×10 <sup>-9</sup>
	rs765547	8p21.3	19910054	near <i>LPL</i>	TG slope	5.45×10 <sup>-7</sup>
	rs6501683	17q25.1	69551967		TG slope	3.89×10 <sup>-8</sup>
	rs11976165	7p11.2	56667548		SBP intercept	6.88×10 <sup>-8</sup>
<b>Yan et al.</b>	rs618675	1p34.3	34922261	near <i>GJA4</i>	CHD	0.0451

Group	SNP	Chromosome	Location	Gene <sup>1</sup>	Trait <sup>2</sup>	p-Value
	rs1333049	9p21.3	22115003	near <i>CDKN2A2B</i>	CHD	4.33×10 <sup>-2</sup>
	rs1376251	12p13.2	11029619	<i>TAS2RS50</i>	CHD	0.0476
<b>Zhu et al.</b>	rs1321130	1q42.2	229310524		TG	<0.05
	rs11688196	2p23.1	30032850	near <i>ALK</i>	TG	<0.05
	rs9828013	3p25.1	13879501	<i>WNT7A</i>	TG	<0.05
	rs16860145	3q13.2	114038583	<i>CD200R2</i>	TG	<0.05
	rs5015152	3q26.31	175530123	near <i>NLGN1</i>	TG	<0.05
	rs6835031	4q22.1	90213948	near <i>TIGD2</i>	TG	<0.05
	rs41442345	4q24	102899436	near <i>BANK1</i>	TG	<0.05
	rs17117113	5q33.2	154575179		TG	<0.05
	rs4367528	8q12.3	62539995	<i>RLBP1L1</i>	TG	<0.05
	rs9643584	8q13.2	68702746	<i>CPA6</i>	TG	<0.05
	rs17630545	8q23.3	113334545	<i>CSMD3</i>	TG	<0.05
	rs1361536	9q21.31	80971505		TG	<0.05
	rs4074863	10q26.3	133187679		TG	<0.05
	rs2514930	11q14.3	90572945		TG	<0.05
	rs17783132	14q24.3	75059376	<i>BATF</i>	TG	<0.05
	rs11150610	16p11.2	31241237	<i>ITGAM</i>	TG	<0.05
	rs4984982	16p13.3	910099	<i>LMFI</i>	TG	<0.05
	rs7204454	16q23.3	81216195		TG	<0.05
	rs16995794	20q13.2	49217529	<i>RPSAPI</i>	TG	<0.05

<sup>1</sup> *EYAL1*, eyes absent homolog 1 (brachio-oto-renal syndrome); *FTO*, protein fto (fat mass and obesity-associated protein); *PRDM16*, PR domain zinc finger protein 16; *CETP*, cholesteryl ester transfer protein precursor (lipid transfer protein D); *PTH1L*, protein PTH1L (parathyroid hormone-responsive B1 gene protein); *PAPA*, pappalysin-1 precursor (insulin-like growth factor-dependent IGF-binding protein 4 protease); *FBN3*, fibrillin-3 precursor; *LPL*, lipoprotein lipase precursor; *PSRC1*, proline/serine-rich coiled-coil protein 1; *CDKN2A2B*, cyclin-dependent kinase inhibitor 2A/2B; *GJA4*, gap junction beta-4 protein; *TAS2RS50*, taste receptor type 2 member 50; *RLBP1L1*, retinaldehyde-binding protein 1-like protein 1; *CD200R2*, cell surface glycoprotein CD200 receptor 2 precursor; *WNT7A*, protein Wnt-7a precursor; *BANK1*, B-cell scaffold protein with ankyrin repeats; *ITGAM*, integrin alpha-M precursor; *NLGN1*, neuroligin-1 precursor; *CSMD3*, CUB and sushi domain-containing protein 3 precursor; *TIGD2*, trigger transposable element-derived protein 2; *LMFI*, lipase maturation factor 1; *RPSAPI*, ribosomal protein SA pseudogene 1; *BATF*, basic leucine zipper transcriptional factor ATF-like; *ALK*, ALK tyrosine kinase receptor precursor; *CPA6*, carboxypeptidase A6 precursor

<sup>2</sup> SBP, systolic blood pressure; BMI, body mass index; MBS, metabolic syndrome; HDL, high density lipoprotein; LDL, low density lipoprotein; TG, triglycerides; CHD, coronary heart disease