Genomics & Informatics Vol. 8(3) 150-158, September 2010

Application of Structural Equation Models to Genome-wide Association Analysis

Jiyoung Kim[1], Junghyun Namkung[2], Seungmook Lee[1] and Taesung Park[1]*

[1]Department of Statistics, Seoul National University, Seoul 151-747, Korea, [2]Department of Epidemiology and Biostatistics, Case Western Reserve University, 2103 Cornell Road, Cleveland, OH 44106-7281, USA

Abstract

Genome-wise association studies (GWASs) have become popular approaches to identify genetic variants associated with human biological traits. In this study, we applied Structural Equation Models (SEMs) in order to model complex relationships between genetic networks and traits as risk factors. SEMs allow us to achieve a better understanding of biological mechanisms through identifying greater numbers of genes and pathways that are associated with a set of traits and the relationship among them. For efficient SEM analysis for GWASs, we developed a procedure, comprised of four stages. In the first stage, we conducted single-SNP analysis using regression models, where age, sex, and recruited area were included as adjusting covariates. In the second stage, Fisher's combination test was conducted for each gene to detect significant genes using p-values obtained from the single-SNP analysis. In the third stage, Fisher' s exact test was adopted to determine which biological pathways were enriched with significant SNPs. Finally, based on a pathway that was associated with the four traits in common, a SEM was fit to model a causal relationship among the genetic factors and traits. We applied our SEM model to GWAS data with four central obesity related traits: suprailiac and subscapular measures for upper body fat, BMI, and hypertension. Study subjects were collected from two Korean cohort regions. After quality control, 327,872 SNPs for 8842 individuals were included in the analysis. After comparing two SEMs, we concluded that suprailiac and subscapular measures may indirectly affect hypertension susceptibility by influencing BMI. In conclusion, our analysis demonstrates that SEMs provide a better understanding of biological mechanisms by identifying greater numbers of genes and pathways.

*Corresponding author: E-mail tspark@stats.snu.ac.kr
Tel +82-2-880-8924, Fax +82-2-883-6114
Accepted 2 September 2010

Keywords: central obesity, suprailiac, subscapular, body mass index (BMI), hypertension, genome-wide association study (GWAS), structural equation model (SEM), gene-based analysis, pathway-based analysis

Introduction

Genome-wide association studies (GWAS) are one of the major tools used to detect disease susceptibility loci. They have been successful in identifying associations of hundreds of single nucleotide polymorphism (SNPs) with complex traits (Rioux *et al*., 2007; Saxena *et al*., 2007; WTCCC, 2007; Zanke *et al*., 2007). However, testing only for the association of individual SNPs has limitations in unveiling the complex mechanism of genetic structures for complex traits (Lesnick *et al*., 2007). Dissecting biological phenomena and understanding the structure of the complex components comprising a biological pathway are challenging tasks. Commonly, GWASs have reported several significant SNPs from individual SNP analyses. However, complex traits are affected by the joint action of various genes. If only the significant SNPs from the individual SNP analysis are considered, the genetic variants that have joint action in determining traits with small individual contributions will be neglected.

Furthermore, the functions of SNPs are not well conceived in many cases, but the functions of genes and pathways have been better explored. Therefore, gene- and pathway-based analysis provides an easier interpretation to unravel the mechanisms of complex traits (Baranzini *et al*., 2009; Kraft and Raychaudhuri, 2009; Rajagopalan and Agarwal, 2005). Most complex traits arise from complex interactions among multiple genetic factors and environmental factors.

In this study, we applied Structural Equation Models (SEMs) in order to model complex relationships between genetic networks and traits as risk factors (Bollen, 1989). The SEM was originally developed in the field of social science to fit a model with unobserved variables. It is well known that the main advantage of the SEM approach is that it allows us to compare several candidate models. Our application of the SEM to a GWAS enables us to investigate how each risk factor affects a targeted trait directly or through other variables, and SEM is used to represent the relationship among multiple phenotypes. In order to choose the components

used to build a model, it is desirable to consider a higher level of genetic components, such as gene and biological pathways.

However, due to the enormous number of SNPs in a GWAS, it is practically difficult to apply SEMs to GWAS data. Thus, some step-wise procedure for filtering out SNPs is required in order to reduce the burden of computation. Our proposed procedure comprises four stages. At the first stage, single SNP association tests are conducted. The next step is to combine P-values for correlated SNPs in order to represent a gene by using Fisher's combination test (Zaykin *et al*., 2007). Next, Fisher's exact test is employed to find the association of a pathway related with the traits. Finally, SEMs are used to model how each risk factor influences the trait of interest based on the pathway chosen from the previous stages.

The proposed SEM approach is applied to a large-scale GWA dataset (i.e., 8842 samples and 327,872 SNPs), obtained from a Korean population. Especially, our analysis focuses on identifying the relationship between hypertension, obesity, and genetic variants. In hypertension, it is known that greater fatness or obesity in the extreme is associated with greater blood pressure or hypertension in the extreme (Dustan, 1991). Understanding the connections between hypertension and relative obesity is an important step in understanding the biological mechanism between them and providing useful information for the potential treatment. Central body fat distribution is especially associated with hypertension and insulin resistance (Licata *et al*., 1994; Scaglione *et al*., 1995). In this study, we investigated the genetic components associated with four central obesity-related traits by conducting the proposed four-stage analysis for large-scale GWAS data. Our GWAS data were collected from the participants of two cohort regions in Korea. The four traits used for analyses were suprailiac, subscapular, BMI, and hypertension. BMI is used for whole-body obesity, and suprailiac and subscapular represent upper central body fatness.

Methods

Study subjects

The data were collected from a Korea Association Resource (KARE) project that was initiated in 2007 to undertake a large-scale GWA analysis. The 10,038 participants were recruited from two community-based cohorts: Ansung, representing mainly a rural community, and Ansan, representing an urban community--5018 from Ansung and 5,020 from Ansan, aged between 40 and 69 years old (Cho, *et al*., 2009) (Table 1). A total of 8800 participants (4162 men and 4638 women) were included for the GWAS analysis, excluding those who were taking medicine or therapy for lipoprotein levels.

Genotyping and quality control

The DNA samples were isolated from the peripheral blood of participants and genotyped using the Affymetrix Genome-Wide Human SNP array 5.0. The Bayesian Robust Linear Modeling using Mahalanobis Distance (BRLMM) was used for genotype calling. Standard data quality control procedures were applied for the subjects and SNPs, as described in Cho *et al*. (2009). After consideration of the sample and SNP quality controls, a total of 8842 participants and 352,228 SNPs remained for the subsequent analysis.

Obesity-related traits

In this study, four traits were considered: subscapular, suprailiac, body mass index (BMI), and hypertension. Subscapular and suprailiac values are skinfold thickness measurements for upper central body fat distribution around the waist. The subscapular measure is a vertical fold taken one inch to the side of the umbilicus from the abdominal. The suprailiac measure is a diagonal fold taken midway between the hip joint and the bottom of the rib cage. BMI is defined as the individual's body weight divided by the square of his or her height. Hypertensive status was defined as a SBP $<$140 mm Hg and/or DBP $<$90 mm Hg, and the blood pressure was measured in the supine position. One of the well-known risk factors of cardiovascular diseases is obesity. Especially, multiple studies have reported that abdominal obesity is a better predictor of hypertension rather than whole body fat mass (Niskanen *et al*., 2004; Selby *et al*., 1989).

Statistical analysis of genetic association

For each SNP, genetic association analyses were con-

Table 1. Demographic information of study subjects

Features	
Total individuals	8842
Gender (male/female)	4183 (47%)/4659 (53%)
Area (ansung/ansan)	4205 (48%)/4637 (52%)
Age	52.22[1] (±8.91)
Body mass index (BMI)	24.59 (±3.12)
Subscapular skinfold (mm)	23.69 (±10.96)
Suprailiac skinfold (mm)	25.70 (±11.72)

[1]Mean (±Standard deviation).

ducted for individual phenotypes independently using three linear regression models for quantitative traits and a logistic regression model for hypertension. Age, sex, and recruited area (Ansan or Ansung) were included as adjusting covariates in all of the regression models. An additive allelic effect was assumed for the mode of genetic inheritance. Association analyses were conducted using PLINK software (Purcell *et al.*, 2007).

Next, gene-based association tests were conducted using Fisher's combination test on the set of p-values obtained from the SNPs within a gene. The statistic (Z_F) for K p-values obtained from K SNPs in a gene is given by:

$$Z_F = -2\sum_{i=1}^{K} \log P_i ,$$

which follows a χ^2_{2K} distribution when the K P-values are independent (Fisher 1925).

Thirdly, Fisher's exact test, based on the hypergenometric distribution, was employed to search for biological pathways that were enriched in the significantly associated genes. Let N be the total number of genes that are of interest; S be the number of genes that are significantly associated with the disease (nominal P-value $\leq$0.05, by Fisher's combination test), and i and m represent the genes in the pathway and significantly associated genes in the pathway, respectively. The significance of the overrepresentation of a biological pathway is calculated by:

$$P = 1 - \sum_{i=0}^{K} \frac{\binom{S}{i}\binom{N-S}{m-i}}{\binom{N}{m}}$$

Combinatorial analysis using SEMs

SEMs are comprehensive statistical models that allow us to test relations among observed and latent (not observed directly) variables. The SEMs for our genetic networks were defined as follows:

$$y = \Lambda_y \eta + \varepsilon ,$$
$$\eta = B\eta + \varsigma ,$$

where y is a vector representing the observed variables; η is a vector of the latent variables (pathway); Λ_y is a matrix representing the true relationships between the SNP and pathway; and B is a matrix representing the true relationships among the latent variables. Random errors in the equations are represented by ε and ς. To fit this model, we used AMOS, a SEM software solution provided by SPSS (http://www.spss.com/amos/).

Results

Single-SNP analysis using regression models

Of 10,038 participants recruited from two Korean cohort areas, 8800 individuals were included to analyze the association between genetic variants and four central obesity-related traits: suprailiac, subscapular, BMI, and hypertension. The association between each of the 327,872 SNPs and each trait was evaluated via a regression model, adjusted for gender, age, and recruitment area. We used a significance level corresponding to p-value < 1.0e-5 to determine which SNPs were associated with a trait. The detailed results are in Table 2.

For suprailiac, 20 SNPs resulted in a significant association. Of these, rs16906215 in TLR4 showed the strongest association. The TLR4 gene encodes a toll-like receptor protein, which is an important member of the innate immune response. This gene has been reported to be associated with type 2 diabetes (Kolz, 2008). However, the effect of a SNP located far downstream of a gene is hard to interpret.

For subscapular, 5 SNPs from 4 genes met the significance criteria. The four genes are GRIN2A, NBPF21P, LOC1001131027, and FTO, and two SNPs from FTO were identified to be significant. FTO is one of the popular genes that are associated with fat mass and obesity. The rs9939609 allele in the FTO gene was previously reported to be positively related with BMI (Frayling *et al.*, 2007; Willer *et al.*, 2009) and type 2 diabetes (WTCCC, 2007). Rs9939609 also showed a marginally significant association with subscapular, producing a p-value of 1.14e-05 and a significant association with BMI in our analysis. The most significant SNP was located upstream of the NMDA receptor (GRIN2A) gene, and GRIN2A was reported to be correlated with hypertension from a previous GWAS data analysis (Torkamani *et al.*, 2008).

In the analysis of hypertension, four significant SNPs were detected from three genes: ATP2B1, CSK, and PTPN11. The results of ATP2B1 and CSK reproduced the significant relationships between variants of the genes and hypertension that had been previously reported by Hong *et al.*, (2010). In their study, they conducted a meta-analysis using two cohort studies in Korea, including KARE (Hong *et al.*, 2010). ATP2B was also reported to be associated with hypertension in another study (Levy *et al.*, 2009). The key function of ATP2B1 is to control homeostasis of cellular calcium ion levels, which are related with vascular smooth muscle contraction and dilation.

Table 2. SNPs significantly associated with central obesity-related traits

RS number	Chr	P-value	MAF	Gene Symbol[1]	Distance	Description
Suprailiac						
rs16906215	9	1.68E-07	0.025	TLR4	66,490	down 70k
rs7681841	4	7.48E-07	0.012	FBXO8	0	intron
rs10090537	8	9.80E-07	0.011	RIMS2	0	intron
rs3856726	3	1.34E-06	0.012	ATG3	97,246	up 100k
rs4472504	8	1.36E-06	0.033	ZMAT4	0	intron
rs17109716	14	1.89E-06	0.032	NRXN3	0	intron
rs1510447	8	2.13E-06	0.232	SGCZ	0	intron
rs601619	18	2.16E-06	0.012	CCDC102B	0	intron
rs4745034	9	4.08E-06	0.029	TRPM3	0	intron
rs17599042	12	4.89E-06	0.034	MUC19	41,221	down 50k
rs1570064	6	5.25E-06	0.015	RHAG	37,578	up 40k
rs11876341	18	6.01E-06	0.098	MEX3C	76,301	up 80k
rs17168600	7	6.05E-06	0.036	LOC100128217	0	intron
rs2210977	1	6.86E-06	0.019	MARK1	0	intron
rs7583940	2	7.03E-06	0.068	LRPPRC	32,425	up 40k
rs6965746	7	7.47E-06	0.309	SLC25A13	77,756	up 80k
rs17226252	5	8.74E-06	0.017		16,777,215	
rs3103261	2	9.25E-06	0.02	DIS3L2	0	intron
rs1849809	4	9.33E-06	0.245	MGC48628	0	intron
rs7010545	8	9.76E-06	0.235	SGCZ	0	intron
Subscapular						
rs16951883	16	1.06E-06	0.018	GRIN2A	42,168	up 50k
rs17248901	3	2.66E-06	0.03	NBPF21P	9,941	down 10k
rs6561930	13	3.36E-06	0.013	LOC100131027	36,359	down 40k
rs7193144	16	8.94E-06	0.126	FTO	0	intron
rs8050136	16	9.01E-06	0.126	FTO	0	intron
BMI						
rs17178527	6	2.24E-08	0.25	LOC729076	97,011	down 100k
rs9939609	16	1.43E-06	0.127	FTO	0	intron
rs11000212	10	1.45E-06	0.206	DDIT4	78,025	up 80k
rs9926289	16	2.45E-06	0.127	FTO	0	intron
rs8050136	16	2.68E-06	0.126	FTO	0	intron
rs527248	1	2.98E-06	0.237	SEC16B	22,728	down 30k
rs7193144	16	3.3E-06	0.126	FTO	0	intron
Hypertension						
rs17249754	12	1.07E-07	0.374	ATP2B1	10,742	up 20k
rs7136259	12	1.7E-07	0.381	ATP2B1	31,344	up 40k
rs1378942	15	2.81E-07	0.172	CSK	0	intron
rs11066280	12	7.15E-06	0.172	PTPN11	38,753	up 40k

[1]Nearby genes are defined as the closest genes to the SNP within signal boundary or the closest genes within a 200-kb window.
MAF, minor allele frequency; BMI, body mass index.

A total of seven SNPs from four genes were identified for BMI. The four genes are LOC729076, FTO, DDIT4, and SEC16B, and four SNPs, including rs9939609, which was shown for the subscapular analysis, were from the FTO gene. SEC16B polymorphisms were previously reported to be associated with obesity and obesity-induced diabetes (Hotta *et al*., 2009).

Gene-based analysis using Fisher's combination test

Next, we tested for the association of genes using Fisher's combination test, where a set of p-values for SNPs in a gene were considered simultaneously. We mapped the identified SNPs to exon/intron or within the 5-kbp upstream/0.5-kbp downstream regions of the known genes. In total, 31,207 genes were annotated among 327,872 SNPs. The P-values of the genes were calculated by Fisher's combination test. The numbers of

genes that were significantly associated with a p-value less than 1.0e-5 were selected. From the analysis, we could find that a large portion of significant genes contained insignificant SNPs. This shows that joint analysis of multiple loci within a gene can have more power, when a single SNP does not have a strong effect on complex diseases.

For subscapular, 33 genes were found to be associated. For subscapular, 54 genes were detected. Sixty-nine genes showed a strong association with BMI. Hypertension had 75 genes showing a strong association. The gene-based approach detected more genes in common within the four phenotypes than the single-marker association analyses. Three genes, SH3RF3, c12ORF51, and ATG10, were detected to be shared with hypertension and BMI. BMI and suprailiac shared 13 genes, BMI and suprailiac had 8 genes in common. Thus, subscapular and BMI had a much greater number of associated genes in common than suprailiac and BMI. Subscapular and suprailiac had 9 genes in common. BMI, subscapular, and suprailiac had PDIA6 and SNX9 in common (Table 3). The PDIA6 gene encodes a protein disulfide isomerase family member protein and localizes in the endoplasmic reticulum (ER). The role of this gene, related to obesity, has been discussed in the aspect of one of the adipocyte extracellular matrix (ECM) processing enzymes (Mariman and Wang, 2010), which partly explains the association between SNPs in PDIA6 and obesity-related traits in common.

Pathway-based analysis using Fisher's exact test

In order to determine the biological pathways that are associated with the four traits, we conducted tests on whether a pathway was enriched with the significant genes identified from the gene-based analysis. Our pathway-based analysis was conducted using 465 pathways retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG; www.genome.jp/kegg/), BioCarta (www.biocarta.com/), and GenMAPP (http://www.genmapp.org/) databases. Table 4 shows the significant pathways for each phenotype, identified using p-values from Fisher's exact test. Eighteen significant pathways were detected for BMI, 23 significant pathways for hypertension, 16 significant pathways for suprailiac, and 17 significant pathways for subscapular. MAPK signaling pathway was shown to be common for all of the four analyzed traits. Hypertension, subscapular, and suprailiac had calcium signaling pathway in common.

MAPK signaling pathway is one of the most ubiquitous signal transduction systems. The role of JNK-MAPK signaling in obesity was reported by multiple studies (Bost *et al.*, 2005; Hirosumi *et al.*, 2002). In addition, the MAPK signaling pathway is claimed to be related to cardiac hypertrophy through the Grb2 adapter protein and cardiac p38 MAPK signaling. Human cardiac hypertrophy is a common condition that often develops as a by-product of hypertension or valvular heart disease (Zhang *et al.*, 2003).

The calcium signaling pathwayis a well-known target for treatment of hypertension (Berridge, 1994). Recently, calcium signaling in obesity was also studied and reported to have a role in determining the cell fate of adipocytes (Sergeev, 2009).

SEM Analysis

In this analysis, SEMs were used to equate causal relationships between genetic networks and correlated phenotypes based on the analysis results in the previous stages. We constructed SEMs, including the MAPK signaling pathway, which appears to be significantly common for every phenotype. SNPs were chosen within genes from the MAPK signaling pathwaywith p-values less than 1.0e-04 for each phenotype. In the models,

Table 3. Genes detected to be significantly associated with more than two traits from a gene-based association test. Genes having P-values $<$ 1.0E$-$5 from Fisher's combination test were considered to be significant. The numbers of significant genes were 55 for subscapular (SUB), 54 for suprailiac (SUP), 69 for BMI, and 75 for hypertension (HTN)

Traits	Significant genes in common
Obesity	
SUB+SUP	**PDIA6**, **SNX9 LOC**, **SNX9**, FBXL20, SUCLG2, hCG_1981, BMPR1B, COMMD10, DOCK10
BMI+SUB	**PDIA6**, LOC34134, JAKMIP2, PADI2, GLI3, **SNX9 LOC**, PARD3B, **SNX9**, NUP205, FTO, CDC123, TULP4, INDOL1
BMI+SUP	**PDIA6**, DMGDH, SSBP3, **SNX9 LOC**, MACF1, **SNX9**, PGCP, AUTS2
Obesity and hypertension	
BMI+HTN	SH3RF3, C12orf51, ATG10
SUB+HTN	**FBXL20**, HSD17B12
SUP+HTN	DCC, **FBXL20**, SRPK1, FBXL17

Genes significant at more than two traits are written in bold.

Table 4. Pathways significantly enriched for genes associated with one of the four traits. p-values obtained from Fisher's exact test based on hypergeometric distribution

Name of Pathway	p-value	Name of Pathway	p-value
Suprailiac		*Subscapular*	
VEGF_SIGNALING_PATHWAY	0.0007	METHANE_METABOLISM	0.00462
CELL_ADHESION_MOLECULES	0.0091	CARBON_FIXATION	0.00515
ADHERENS_JUNCTION	0.0044	REDUCTIVE_CARBOXYLATE_CYCLE	0.00521
PURINE_METABOLISM	0.0079	THIAMINE_METABOLISM	0.00629
GAP_JUNCTION	0.045	RIBOFLAVIN_METABOLISM	0.0089
COMPLEMENT_AND_COAGULATION_CASCADES	0.0231	VITAMIN_B6_METABOLISM	0.01248
ANTIGEN_PROCESSING_AND_PRESENTATION	0.0043	BIOTIN_METABOLISM	0.01257
RENIN_ANGIOTENSIN_SYSTEM	0.0079	LIPOIC_ACID_METABOLISM	0.01409
CALCIUM_SIGNALING_PATHWAY	0.0079	**CALCIUM_SIGNALING_PATHWAY**	0.01456
HEMATOPOIETIC_CELL_LINEAGE	0.0208	RETINOL_METABOLISM	0.01514
MAPK_SIGNALING_PATHWAY	0.0208	PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	0.01664
B_CELL_RECEPTOR_SIGNALING_PATHWAY	0.0079	LIMONENE_AND_PINENE_DEGRADATION	0.01682
FC_EPSILON_RI_SIGNALING_PATHWAY	0.0128	PHENYLPROPANOID_BIOSYNTHESIS	0.0208
CIRCADIAN_RHYTHM	0.0208	ALKALOID_BIOSYNTHESIS_I	0.02288
LONG_TERM_POTENTIATION	0.0445	AMINOACYL_TRNA_BIOSYNTHESIS	0.0308
		MAPK_SIGNALING_PATHWAY	0.03288
		GLYCAN_STRUCTURES_DEGRADATION	0.04424
BMI		*Hypertension*	
GLYCOLYSIS_AND_GLUCONEOGENESIS	0.0005	HISTIDINE_METABOLISM	0.00007
CITRATE_CYCLE	0.00515	GAMMA_HEXACHLOROCYCLOHEXANE_DEGRADATION	0.00044
PENTOSE_PHOSPHATE_PATHWAY	0.00629	BISPHENOL_A_DEGRADATION	0.00079
INOSITOL_METABOLISM	0.00629	TRYPTOPHAN_METABOLISM	0.00079
PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS	0.0089	PHENYLALANINE_TYROSINE_AND_TRYPTOPHAN_BIOSYNTHESIS	0.00091
FRUCTOSE_AND_MANNOSE_METABOLISM	0.01248	NOVOBIOCIN_BIOSYNTHESIS	0.00148
GALACTOSE_METABOLISM	0.01456	SELENOAMINO_ACID_METABOLISM	0.00208
FATTY_ACID_METABOLISM	0.0208	GLUTATHIONE_METABOLISM	0.00231
SYNTHESIS_AND_DEGRADATION_OF_KETONE_BODIES	0.02521	NAPHTHALENE_AND_ANTHRACENE_DEGRADATION	0.00233
BIOSYNTHESIS_OF_STEROIDS	0.02665	1,4_DICHLOROBENZENE_DEGRADATION	0.00246
BILE_ACID_BIOSYNTHESIS	0.03121	ETHYLBENZENE_DEGRADATION	0.00445
MAPK_SIGNALING_PATHWAY	0.03187	BUTANOATE_METABOLISM	0.00446
ANDROGEN_AND_ESTROGEN_METABOLISM	0.03329	**MAPK_SIGNALING_PATHWAY**	0.0045
OXIDATIVE_PHOSPHORYLATION	0.03953	THIAMINE_METABOLISM	0.00549
UREA_CYCLE_AND_METABOLISM_OF_AMINO_GROUPS	0.0402	VITAMIN_B6_METABOLISM	0.00808
CAFFEINE_METABOLISM	0.04429	BIOTIN_METABOLISM	0.00886
GLUTAMATE_METABOLISM	0.04848	LIPOIC_ACID_METABOLISM	0.01455
		RETINOL_METABOLISM	0.01808
		CALCIUM_SIGNALING_PATHWAY	0.0267
		MONOTERPENOID_BIOSYNTHESIS	0.03128
		SULFUR_METABOLISM	0.03222
		CAPROLACTAM_DEGRADATION	0.04888

the pathway was treated as a latent variable, and SNPs and four phenotypes were the observed variables. Two models were considered, as shown in Fig. 1. Model 1 assumes that suprailiac and subscapular influence BMI and that BMI influences hypertension. Model 2 assumes that suprailiac, subscapular, and BMI influence hypertension simultaneously.

In an SEM, the goodness-of-fit index (GFI) measures the relative differences between the data and estimated values obtained from a model, while the adjusted GFI (AGFI) adjusts the GFI according to the degrees of freedom. If these two measures are close to 1, we can conclude that the model fits the data well. The Akaike information criterion (AIC) is a well-known measure that can be used for model comparisons. The smaller the AIC, the better the model is. Three types of goodness-of-fit measures were evaluated to select the best model of the two models: for model 1, GFI=0.791, AGFI=0.793, and AIC=1132.005; and for model 2, GFI=0.774, AGFI=0.776, and AIC=1290.599. In general, models with GFI and AGFI measures closer to 1 and with smaller AICs are considered to fit well. As GFI and AGFI were close to 1 for both models, both models fit the data well. But, model 1 showed a bigger GFI and

(a) Model 1.

(b) Model 2.

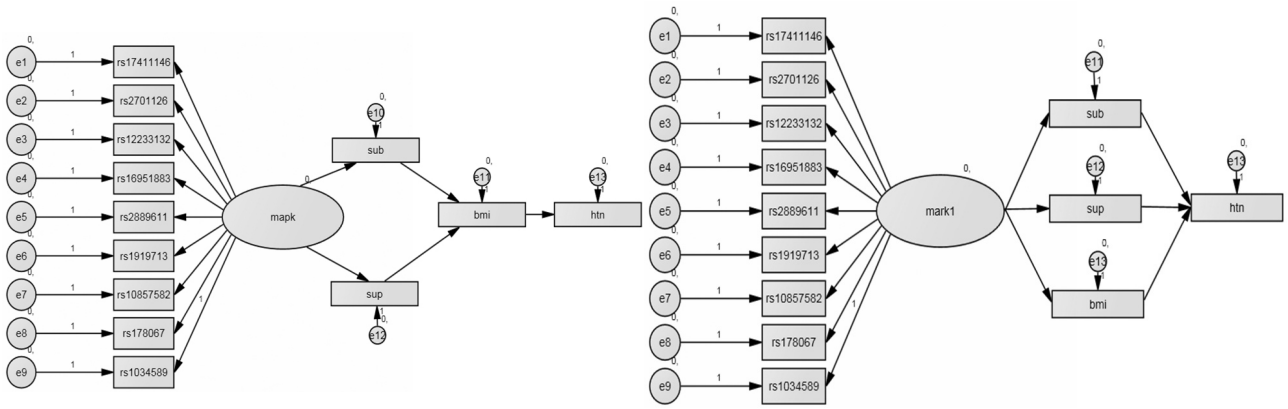

Fig. 1. Structural equation models.

AGFI with a smaller AIC than model 2; model 1 was selected to be the final model. Through a comparison between two models, we concluded that suprailiac and subscapular indirectly affect hypertension susceptibility by influencing BMI.

Discussion

Most GWASs have focused on single-SNP analyses of single traits. The result of a single-SNP analysis is limited to correlations between candidate susceptible loci and single traits. In this study, we conducted a gene- and pathway-based analysis beyond single-SNP analysis. Based on the gene- and pathway-based analysis, we proposed the use of SEMs to construct a model for causal relationships among genetic factors and risk factors in terms of traits through an underlying biological pathway.

The gene-based analysis detected more significant genes than the single-SNP analysis. We demonstrate that gene-based analysis is a powerful method to detect genes that are associated with traits. A gene-based analysis is also easier to interpret, as function and the relevant disease with a gene are better investigated than a SNP.

A pathway-based analysis was conducted to find pathways that were significantly correlated with each phenotype. In our analysis, a pathway that was associated with four traits in common was detected. Pathway-based analysis provides a more comprehensive understanding of the biological process of complex traits than a gene-based or single-SNP analysis.

SEM was used to identify the relationship among risk factors of a complex trait. In this study, SEMs were used to model how upper body fat distribution, represented by suprailiac and subscapular, and average body fat distribution, represented by BMI, were related with hypertension. Although we fit a limited number of SEMs, we demonstrated that SEM analysis is useful in investigating complex biological phenomena, because it allows us to present complex causal relationships in equations and express them in path diagrams and because it deals better with correlated variables that occur frequently in biological data.

However, SEM has some limitations. It requires an initial model to start with. SEM is constructed under strict multivariate normality and independence assumptions among errors. Another difficulty is that run time increases substantially as the number of variables increases.

Acknowledgments

This work was supported by the Consortium for Large-Scale Genome-Wide Association Study, National Research Foundation (KRF-2008-313-C00086), and the Brain Korea 21 Project of the Ministry of Education. The KARE data analyzed in this study were obtained from the Korean Genome Analysis Project (4845-301), which was funded by a grant from the Korea National Institute of Health (Korea Center for Disease Control, Ministry for Health, Welfare and Family Affairs), Republic of Korea. The authors thank Hyung-Lae Kim (Center for Genome Science, National Institute of Health, South Korea), Nam H. Cho (Department of Preventive Medicine, Ajou University, South Korea), and Chol Shin (Department of Internal Medicine, Korea University Ansan Hospital, South Korea) for their great efforts in generating and providing these valuable data.

References

Baranzini, S.E., Galwey, N.W., Wang, J., Khankhanian, P., Lindberg, R., Pelletier, D., Wu, W., Uitdehaag, B.M., Kappos, L., Polman, C.H., Matthews, P.M., Hauser, S.L., Gibson, R.A., Oksenberg, J.R., and Barnes, M.R. (2009). Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum. Mol. Genet.* 18, 2078-2090.

Berridge, M.J. (1994). The biology and medicine of calcium signalling. *Mol. Cell Endocrinol.* 98, 119-124.

Bost, F., Aouadi, M., Caron, L., and Binetruy, B. (2005). The role of MAPKs in adipocyte differentiation and obesity. *Biochimie* 87, 51-56.

Cho, Y.S., Go, M.J., Kim, Y.J., Heo, J.Y., Oh, J.H., Ban, H.J., Yoon, D., Lee, M.H., Kim, D.J., Park, M., Cha, S.H., Kim, J.W., Han, B.G., Min, H., Ahn, Y., Park, M.S., Han, H.R., Jang, H.Y., Cho, E.Y., Lee, J.E., Cho, N.H., Shin, C., Park, T., Park, J.W., Lee, J.K., Cardon, L., Clarke, G., McCarthy, M.I., Lee, J.Y., Oh, B., and Kim, H.L. (2009). A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* 41, 527-534.

Dustan, H. (1991). Obesity and hypertension. *Diabetes Care* 14, 488-504.

Fisher, R.A. (1925). Statistical methods for research workers. Oliver and Lloyd, London.

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W., Shields, B., Harries, L.W., Barrett, J.C., Ellard, S., Groves, C.J., Knight, B., Patch, A.M., Ness, A.R., Ebrahim, S., Lawlor, D.A., Ring, S.M., Ben-Shlomo, Y., Jarvelin, M.R., Sovio, U., Bennett, A.J., Melzer, D., Ferrucci, L., Loos, R.J., Barroso, I., Wareham, N.J., Karpe, F., Owen, K.R., Cardon, L.R., Walker, M., Hitman, G.A., Palmer, C.N., Doney, A.S., Morris, A.D., Smith, G.D., Hattersley, A.T., and McCarthy, M.I. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316, 889-894.

Hirosumi, J., Tuncman, G., Chang, L., Gorgun, C.Z., Uysal, K.T., Maeda, K., Karin, M., and Hotamisligil, G.S. (2002). A central role for JNK in obesity and insulin resistance. *Nature* 420, 333-336.

Hong, K.W., Go, M.J., Jin, H.S., Lim, J.E., Lee, J.Y., Han, B.G., Hwang, S.Y., Lee, S.H., Park, H.K., Cho, Y.S., and Oh, B. (2010). Genetic variations in ATP2B1, CSK, ARSG and CSMD1 loci are related to blood pressure and/or hypertension in two Korean cohorts. *J. Hum. Hypertens.* 24, 367-372.

Hotta, K., Nakamura, M., Nakamura, T., Matsuo, T., Nakata, Y., Kamohara, S., Miyatake, N., Kotani, K., Komatsu, R., Itoh, N., Mineo, I., Wada, J., Masuzaki, H., Yoneda, M., Nakajima, A., Funahashi, T., Miyazaki, S., Tokunaga, K., Kawamoto, M., Ueno, T., Hamaguchi, K., Tanaka, K., Yamada, K., Hanafusa, T., Oikawa, S., Yoshimatsu, H., Nakao, K., Sakata, T., Matsuzawa, Y., Kamatani, N., and, Y.N. (2009). Association between obesity and polymorphisms in SEC16B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population. *J. Hum. Genet.* 54, 727-731.

Kolz, M. (2008). Association between variations in the TLR4 gene and incident type 2 diabetes is modified by the ratio of total cholesterol to HDL-cholesterol. *BMC Med. Genet.* 25, 9.

Kraft, P., and Raychaudhuri, S. (2009). Complex diseases, complex genes: keeping pathways on the right track. *Epidemiology* 20, 508-511.

Lesnick, T.G., Papapetropoulos, S., Mash, D.C., Ffrench-Mullen, J., Shehadeh, L., de Andrade, M., Henley, J.R., Rocca, W.A., Ahlskog, J.E., and Maraganore, D.M. (2007). A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet*, 3, e98.

Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., Aulchenko, Y., Lumley, T., Kottgen, A., Vasan, R.S., Rivadeneira, F., Eiriksdottir, G., Guo, X., Arking, D.E., Mitchell, G.F., Mattace-Raso, F.U., Smith, A.V., Taylor, K., Scharpf, R.B., Hwang, S.J., Sijbrands, E.J., Bis, J., Harris, T.B., Ganesh, S.K., O'Donnell, C.J., Hofman, A., Rotter, J.I., Coresh, J., Benjamin, E.J., Uitterlinden, A.G., Heiss, G., Fox, C.S., Witteman, J.C., Boerwinkle, E., Wang, T.J., Gudnason, V., Larson, M.G., Chakravarti, A., Psaty, B.M., and van Duijn, C.M. (2009). Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* 41, 677-687.

Licata, G., Scaglione, R., Ganguzza, A., Corrao, S., Donatelli, M., Parrinello, G., Dichiara, M.A., Merlino, G., and Cecala, M.G. (1994). Central obesity and hypertension. Relationship between fasting serum insulin, plasma renin activity, and diastolic blood pressure in young obese subjects. *Am. J. Hypertens.* 7, 314-320.

Mariman, E.C., and Wang, P. (2010). Adipocyte extracellular matrix composition, dynamics and role in obesity. *Cell Mol. Life Sci.* 67, 1277-1292.

Niskanen, L., Laaksonen, D.E., Nyyssönen, K., Punnonen, K., Valkonen, V.-P., Fuentes, R., Tuomainen, T.-P., Salonen, R., and Salonen, J.T. (2004). Inflammation, abdominal obesity, and smoking as predictors of hypertension. *Hypertension* 44, 859-865.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559-575.

Rajagopalan, D., and Agarwal, P. (2005). Inferring pathways from gene lists using a literature-derived network of biological relationships. *Bioinformatics* 21, 788-793.

Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W., Shugart, Y.Y., Griffiths, A.M., Targan, S.R., Ippoliti, A.F., Bernard, E.J., Mei, L., Nicolae, D.L., Regueiro, M., Schumm, L.P., Steinhart, A.H., Rotter, J.I., Duerr, R.H., Cho, J.H., Daly, M.J., and Brant, S.R. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates au-

tophagy in disease pathogenesis. *Nat. Genet.* 39, 596-604.

Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J., Hughes, T.E., Groop, L., Altshuler, D., Almgren, P., Florez, J.C., Meyer, J., Ardlie, K., Bengtsson Bostrom, K., Isomaa, B., Lettre, G., Lindblad, U., Lyon, H.N., Melander, O., Newton-Cheh, C., Nilsson, P., Orho-Melander, M., Rastam, L., Speliotes, E.K., Taskinen, M.R., Tuomi, T., Guiducci, C., Berglund, A., Carlson, J., Gianniny, L., Hackett, R., Hall, L., Holmkvist, J., Laurila, E., Sjogren, M., Sterner, M., Surti, A., Svensson, M., Tewhey, R., Blumenstiel, B., Parkin, M., Defelice, M., Barry, R., Brodeur, W., Camarata, J., Chia, N., Fava, M., Gibbons, J., Handsaker, B., Healy, C., Nguyen, K., Gates, C., Sougnez, C., Gage, D., Nizzari, M., Gabriel, S.B., Chirn, G.W., Ma, Q., Parikh, H., Richardson, D., Ricke, D., and Purcell, S. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316, 1331-1336.

Scaglione, R., Ganguzza, A., Corrao, S., Parrinello, G., Merlino, G., Dichiara, M.A., Arnone, S., D'Aubert, M.D., and Licata, G. (1995). Central obesity and hypertension: pathophysiologic role of renal haemodynamics and function. *Int J Obes Relat. Metab. Disord.* 19, 403-409.

Selby, J.V., Friedman, G.D., and Quesenberry, C.P., Jr. (1989). Precursors of essential hypertension. The role of body fat distribution pattern. *Am. J. Epidemiol.* 129, 43-53.

Sergeev, I.N. (2009). Novel Mediators of Vitamin D Signaling in Cancer and Obesity, Immun., Endoc. & Metab. *Agents in Med. Chem.* 9, 153-158.

Torkamani, A., Topol, E.J., and Schork, N.J. (2008). Pathway analysis of seven common diseases assessed by genome-wide association, *Genomics* 92, 265-272.

Willer, C.J., Speliotes, E.K., Loos, R.J., Li, S., Lindgren, C.M., Heid, I.M., Berndt, S.I., Elliott, A.L., Jackson, A.U., Lamina, C., Lettre, G., Lim, N., Lyon, H.N., McCarroll, S.A., Papadakis, K., Qi, L., Randall, J.C., Roccasecca, R.M., Sanna, S., Scheet, P., Weedon, M.N., Wheeler, E., Zhao, J.H., Jacobs, L.C., Prokopenko, I., Soranzo, N., Tanaka, T., Timpson, N.J., Almgren, P., Bennett, A., Bergman, R.N., Bingham, S.A., Bonnycastle, L.L., Brown, M., Burtt, N.P., Chines, P., Coin, L., Collins, F.S., Connell, J.M., Cooper, C., Smith, G.D., Dennison, E.M., Deodhar, P., Elliott, P., Erdos, M.R., Estrada, K., Evans, D.M., Gianniny, L., Gieger, C., Gillson, C.J., Guiducci, C., Hackett, R., Hadley, D., Hall, A.S., Havulinna, A.S., Hebebrand, J., Hofman, A., Isomaa, B., Jacobs, K.B., Johnson, T., Jousilahti, P., Jovanovic, Z., Khaw, K.T., Kraft, P., Kuokkanen, M., Kuusisto, J., Laitinen, J., Lakatta, E.G., Luan, J., Luben, R.N., Mangino, M., McArdle, W.L., Meitinger, T., Mulas, A., Munroe, P.B., Narisu, N., Ness, A.R., Northstone, K., O'Rahilly, S., Purmann, C., Rees, M.G., Ridderstrale, M., Ring, S.M., Rivadeneira, F., Ruokonen, A., Sandhu, M.S., Saramies, J., Scott, L.J., Scuteri, A., Silander, K., Sims, M.A., Song, K., Stephens, J., Stevens, S., Stringham, H.M., Tung, Y.C., Valle, T.T., Van Duijn, C.M., Vimaleswaran, K.S., Vollenweider, P., Waeber, G., Wallace, C., Watanabe, R.M., Waterworth, D.M., Watkins, N., Witteman, J.C., Zeggini, E., Zhai, G., Zillikens, M.C., Altshuler, D., Caulfield, M.J., Chanock, S.J., Farooqi, I.S., Ferrucci, L., Guralnik, J.M., Hattersley, A.T., Hu, F.B., Jarvelin, M.R., Laakso, M., Mooser, V., Ong, K.K., Ouwehand, W.H., Salomaa, V., Samani, N.J., Spector, T.D., Tuomi, T., Tuomilehto, J., Uda, M., Uitterlinden, A.G., Wareham, N.J., Deloukas, P., Frayling, T.M., Groop, L.C., Hayes, R.B., Hunter, D.J., Mohlke, K.L., Peltonen, L., Schlessinger, D., Strachan, D.P., Wichmann, H.E., McCarthy, M.I., Boehnke, M., Barroso, I., Abecasis, G.R., and Hirschhorn, J.N. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* 41, 25-34.

WTCCC (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-678.

Zanke, B.W., Greenwood, C.M., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S.M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E., Ferretti, V., Laflamme, P., Sundararajan, S., Roumy, S., Olivier, J.F., Robidoux, F., Sladek, R., Montpetit, A., Campbell, P., Bezieau, S., O'Shea, A.M., Zogopoulos, G., Cotterchio, M., Newcomb, P., McLaughlin, J., Younghusband, B., Green, R., Green, J., Porteous, M.E., Campbell, H., Blanche, H., Sahbatou, M., Tubacher, E., Bonaiti-Pellie, C., Buecher, B., Riboli, E., Kury, S., Chanock, S.J., Potter, J., Thomas, G., Gallinger, S., Hudson, T.J., and Dunlop, M.G. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* 39, 989-994.

Zaykin, D.V., Zhivotovsky, L.A., Czika, W., Shao, S., and Wolfinger, R.D. (2007). Combining p-values in large-scale genomics experiments. *Pharm. Stat.* 6, 217-226.

Zhang, S., Weinheimer, C., Courtois, M., Kovacs, A., Zhang, C.E., Cheng, A.M., Wang, Y., and Muslin, A.J. (2003). The role of the Grb2-p38 MAPK signaling pathway in cardiac hypertrophy and fibrosis. *J. Clin. Invest.* 111, 833-841.