Original Article

The Impact of Lipid-metabolizing Genetic Polymorphisms on Body Mass Index and Their Interactions with Soybean Food Intake: A Study in a Chinese Population*

WANG Jin Wei[1], TANG Xun[1], LI Na[2], WU Yi Qun[1], LI Shuai[1], LI Jin[1], QIN Xue Ying[1], ZHANG Zong Xin[3], HU Yong Hua[1,#], and CHEN Da Fang[1,#]

1. Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing 100191, China; 2. Fangshan District Bureau of Health, Beijing 102401, China; 3. The First Hospital of Fangshan District, Beijing 102400, China

Abstract

Objective To evaluate the association of known polymorphisms in the lipid metabolic pathway with body mass index (BMI), and estimate their interactions with soybean food intake.

Methods A community-based cross-sectional survey was conducted in a Chinese Han population. BMI, soybean food intake, and single nucleotide polymorphisms of rs599839, rs3846662, rs3846663, rs12916, rs174547, rs174570, rs4938303, and rs1558861 were measured in 944 subjects. A multivariate logistic regression was used to analyze the association of the studied polymorphisms with BMIs. The expectation-maximization algorithm was employed to evaluate the extent of linkage disequilibrium between pairwise polymorphisms. The gene-environment interaction was assessed in the general multifactor dimensionality reduction model.

Results The polymorphisms of rs3846662 and rs3846663 were associated with 10% highest BMIs when comparing to the 10% lowest values both in individuals and haplotype-based association tests. Although no statistically significant gene-environment interactions were found, people with the haplotype composed of C allele in rs3846662 and T allele in rs3846663 and low frequency of soybean intake had significantly higher risk to overweight and obesity as compared with those with the haplotype consisting of T allele in rs3846662 and C allele in rs3846663 and highly frequent soybean food intake, with an odds ratio of 1.64 (95% confidence interval: 1.15-2.34, $P<0.01$) after adjusting for the common confounders.

Conclusion Our study has suggested that rs3846662 and rs3846663 may be the potential candidate polymorphisms for obesity, and their effect on the pathogenesis could be mediated by the frequency of soybean food intake.

Key words: Body mass index; Lipid metabolism; Genetic epidemiology; Haplotypes; Gene-environment interaction

Biomed Environ Sci, 2014; 27(3): 176-185 *doi: 10.3967/bes2014.039* *ISSN: 0895-3988*

www.besjournal.com (full text) *CN: 11-2816/Q*

*This study was supported by the National Basic Research Program of China (973 Program) (2006CB503903) and the National Natural Science Foundation of China (81172744, 81230066).

#Correspondence should be addressed to: HU Yong Hua, Tel: 86-10-82801528, E-mail: yhhu@bjmu.edu.cn; CHEN Da Fang, Tel: 86-10-82801191, E-mail: dafangchen@bjmu.edu.cn

Biographical note of the first author: WANG Jin Wei, male, born in 1985, PHD candidate student at the School of Public Health, Peking University, majoring in genetic epidemiology of chronic diseases.

Received: January 4, 2013; Accepted: July 18, 2013

INTRODUCTION

The obesity epidemic is a major public health challenge faced by both developed and developing countries. The World Health Organization estimated that in 2008 there were 1.6 billion overweight [body mass index (BMI) ≥25] and obese (BMI≥30) adults worldwide[1]. In China, between 1991 and 2006, the prevalence of overweight adults doubled (from 13.5% to 26.7%) and that of obese adults tripled (from 1.1% to 3.2%)[2]. Obesity increases the risk to many chronic diseases, such as type 2 diabetes mellitus, dyslipidemia, hypertension, cardiovascular diseases and several cancers[3-4], thereby lessening life expectancy and causing an increasing number of deaths[5]. Given the huge disease burden resulting from obesity, effective public health interventions are urgently needed.

An imbalance between energy intake and expenditure is the fundamental cause of obesity. Rapid changing lifestyle is believed to be the main reason for the obesity epidemic. Dietary pattern is an important aspect of daily life and is usually distinctive among different regions and ethnic groups. In eastern Asia, especially in China, Japan and Korea, foods made from soybeans are widely consumed. Soy protein contains all essential amino acids and is rich in several nutrients, which can improve the lipid metabolism, like linoleic acid, linolenic acid, lecithin and isoflavone[6]. Furthermore, studies in humans and animals have consistently found that the intake of soy protein provides less energy than that of animal protein, thus potentially contributing to the body weight control[7].

Twin studies have shown that BMI heritability was 65%-80% in a Caucasian population[8], and 61% in a Chinese Han population[9]. Linkage and association studies have located many genes that influence BMI[10]. The largest genome-wide association (GWA) study for BMI to date recruited nearly 250 thousand individuals and covered 2.8 million common variants in the human genome. Thirty-two single nucleotide polymorphisms (SNPs) that reached the genome-wide significance level ($P<5\times10^{-8}$) were identified. Some of the variants were located in or near the genes related to plasma lipid metabolism, e.g. *APOB48R*, *HMGCR,* and *ZNF608*, thus highlighting the role of dyslipidemia in the pathogenesis of obesity[11]. Dyslipidemia and obesity are both essential disorders that define the metabolic syndrome. It is likely that these two disorders have a common basis and it is highly suspected that the basis is insulin resistance[12]. Therefore, genetic variants of the genes involved in lipid metabolism deserve an evaluation of their association with other metabolic syndrome phenotypes, like obesity.

In this study, eight candidate SNPs with the susceptible function to influence the metabolism of plasma cholesterol and triglyceride, were selected from recent GWA studies[13-18]. The purpose of this study was to explore the relationship between these lipid-metabolizing genes and BMI related phenotypes, and also to evaluate the gene-environment interactions with the habit of dietary soy foods intake.

MATERIALS AND METHODS

Study Population

The participants were enrolled from the urban Fangshan district in southwest Beijing. The subjects were all local adult residents with ancestry from the same area, so that their genetic construction was stable and representative of the Han Chinese in northern China. The subjects were selected based on a two-stage sampling strategy. First, three subdistricts that had at least 1 000 households in the residence registration system were selected. Then, participants were recruited in the primary health care centers of the subdistricts. The study was conducted between 2008 and 2010 with the primary purpose of obtaining data on the prevalence rate of overweight and obesity. Finally, 1 233 individuals completed the in-person interview and physical examinations; 978 of them donated a 4 mL blood specimen. The Ethics Committee of the Peking University Health Science Center approved the study and all the subjects provided their written informed consents.

Information Collection

The eligible participants completed a structured questionnaire. Because the exact amount of soy food intake was difficult to estimate, the frequency of intake was measured instead. Other demographic information, sedentary lifestyle information and status of smoking and drinking were collected. Body weight and height were measured in light clothing and without shoes. BMI was calculated as weight divided by the square of height (kg/m^2). The BMI cutoffs to classify overweight and obesity were 25

and 30 kg/m^2 respectively. The plasma concentrations of total cholesterol (TC), low density lipoprotein cholesterol (LDL-C), high density lipoprotein cholesterol (HDL-C) and triglyceride (TG) were assayed in the laboratory of Fangshan First Hospital within 24 h after blood collection. The measurement was performed according to the standard instructions using a Rongsheng measurement kit (Rongsheng Biomedicine Inc., Shanghai, China) in the Hitachi 7180 auto-biochemical analyzer (Hitachi High-Technologies Corp., Tokyo, Japan). The laboratory of Fangshan First Hospital is the member of the Beijing Centre for Clinical Laboratory and the quality of the measurement satisfies the national standard requirements. Dyslipidemia was defined if one of the following disorders was identified according to the Chinese guidelines on prevention and treatment of dyslipidemia in adults: TC≥6.22 mmol/L, LDL-C≥4.14 mmol/L, HDL-C<1.04 mmol/L, and TG≥2.26 mmol/L.

Genotype Determination

Genomic DNA was extracted from peripheral blood leukocytes according to the standard protocol by using a Tiangen DNA extraction kit (Tiangen Biotechnology Inc., Beijing, China). DNA samples were available from 944 subjects. Eight functional SNPs from the recent published GWA studies on lipids profile were evaluated. Genotyping was performed using the Genomelab SNP stream System (Beckman Coulter Inc., Fullerton, CA, USA). First, the genomic sequences encompassing the SNPs of interest were amplified in an 8-plex polymerase chain reaction (PCR) , using 6 ng of DNA. Then, the products were cleaned using shrimp alkaline phosphatase before the single base extension process. The probe sequences for the next extension reaction contained two parts: the 5' end was complementary to one of the 8-tag DNA oligonucleotides previously immobilized in the SNP ware array; and the 3' end was designed to hybridize to the template strand just one base adjacent to the SNP site. In the extension process, one fluorescent dye labeled nucleotide was connected to the probe sequence and this terminator could be detected by a CCD scan after hybridization to the sorted 8-tag DNA oligonucleotides in each of the 384 wells of the array. The fluorescent signals were transferred into the genotype information according to the fluorescent color and intensity. A subsample was genotyped repeatedly and the rate of consistency was above the 99%.

Statistical Analysis

Besides the BMI value itself, we used the other two comparison patterns: BMI ≥25 (overweight/obesity) compared to BMI <25 (normal); and 10% of the highest compared to 10% of the lowest BMI extremes. The comparison of the means of continuous variables was conducted by two-sample *t* test. The characteristics of categorical variables and the Hardy-Weinberg Equilibrium were analyzed using Pearson's χ^2 test. If ordinal variables were met, the Kruskal–Wallis rank sum test was implemented. The association between individual polymorphisms and metabolic disorders was explored by multivariate logistic regression, adjusting for gender, age, sedentary lifestyle, soybean food intake, smoking and drinking status. Odds ratios (ORs) and 95% confidence intervals (CIs) were reported. The descriptive and association analyses were conducted in the SAS 9.1 software (SAS Institute Inc., Cary, NC). The *P* value of bonferroni correction for multiple testing of single marker associations was 0.05/8=0.00625. The estimation of haplotype construction was based on the expectation-maximization (EM) algorithm using the Haploview program, version 4.2 (Broad Institute, Cambridge, MA, USA)[19]. *D*' and r^2 were used to measure the strength of linkage disequilibrium (LD) between pairwise SNPs. The program PHASE, version 2.1 (Department of Statistics, University of Washington, Seattle), was employed to reconstruct haplotypes from the existing genotype data[20]. The major haplotypes with a frequency of more than 5% were listed and used to explore the haplotype-phenotype associations. Gene-environment interactions were evaluated based on the multifactor dimensionality reduction (MDR) principle[21]. Briefly, this nonparametric method reduces the high-dimensional data into one dimension with two groups: high risk and low risk groups. Therefore, the curse of the dimensionality problem in the logistic regression was avoided. In this study, we used generalized multifactor dimensionality reduction (GMDR) as an alternative to MDR to include the important variables as covariates. Two to four factor models were constructed and the 'est model' was defined as the model having the highest testing balanced accuracy. The significant model was judged by the *P* value from sign statistical test. The power for the single marker association and gene-environment

interaction was calculated by the Quanto software, version 1.2.4 (Department of Preventive Medicine, University of Southern California).

RESULTS

Demographic and Behavioral Characteristics and the Lipid Profile of the Study Population

Nine hundred and forty four Chinese Han subjects were included in the final analysis. They were aged from 18 to 86, with a mean of 40.23, and 74.58% of the subjects were male. The demographic characteristics, distributions of behavioral risk factors and lipids profile of the subjects are listed in Table 1. Compared with those with normal BMI, the subjects with BMI≥25 kg/m^2 had a higher mean age and a higher percentage of male. Also, they reported more cigarette smoking and alcohol drinking, longer time for sitting per day, and less intake of soy foods (P<0.01 for all the comparisons). Additionally, the subjects with BMI≥25 kg/m^2 had higher levels of plasma TG, TC, LDL-C, and HDL-C concentrations than those with BMI<25 kg/m^2 (P<0.01 for all the comparisons).

Table 1. Demographic, Behavioral Characteristics and Lipid Profile of the Study Population, by BMI Status

Variable	BMI<25 kg/m^2	BMI≥25 kg/m^2	*P* Value
	Mean±SD/No. of Persons (%)	Mean±SD/No. of Persons (%)	
n	636	308	
Age (years)	39.15±11.45	42.46±10.59	<0.01
Gender (male)	440 (69.18)	264 (85.71)	<0.01
Cigarette smoking (yes)	249 (39.15)	180 (58.44)	<0.01
Alcohol drinking (yes)	284 (44.65)	174 (56.49)	<0.01
Time spent sitting*			<0.01
<6 h/day	196 (30.82)	50 (16.23)	
6-9 h/day	343 (53.93)	100 (32.47)	
>9 h/day	97 (15.25)	158 (51.30)	
Soy consumption frequency*			<0.01
<1 time/week	410 (64.47)	228 (74.03)	
1-6 times/week	161 (25.31)	68 (22.08)	
≥1 time/day	65 (10.22)	12 (3.90)	
TG (mmol/L)	1.29±0.91	2.18±1.77	<0.01
TC (mmol/L)	4.84±0.99	5.35±1.06	<0.01
LDL-C (mmol/L)	2.84±0.79	3.11±0.75	<0.01
HDL-C (mmol/L)	1.41±0.34	1.26±0.30	<0.01
Dyslipidemia (yes)	260 (40.88)	227 (73.70)	<0.01

Note. BMI: body mass index; SD: standard deviation. *Comparison between BMI groups was conducted by the Kruskal-Wallis rank sum test.

The Associations between Genetic Variations and Dyslipidemia and BMI Categories

Eight candidate SNPs entered our analyses. All of them, except the rs12916, conformed to the Hardy-Weinberg equilibrium (P>0.001). Although the deviation from the Hardy-Weinberg Equilibrium can be explained by the random fluctuations because of the limited sample size, the rs12916 was not included in the following haplotype construction and interaction analysis. The relevant information on the polymorphisms is available in Table 2. We evaluated the effect of T allele for the risk of dyslipidemia and BMI categories for each SNP using the multivariate logistic regression in the additive genetic model. The increasing dosage of T allele of rs599839 increased the risk of dyslipidemia with an OR of 1.48 (95% CI: 1.05-2.07, P=0.03), while that of rs12916 and rs1558861 showed the effect of protection against dyslipidemia with ORs of 0.79 (95% CI: 0.66-0.96, P=0.02) and 0.77 (95% CI: 0.61-0.99, P=0.04), respectively. None of the studied SNPs was associated with overweight/obesity. Given the sample size of our study and the minor allele frequency of 0.45 (as that for the T allele of rs3846662 and C allele of rs3846663), the power to detect the genetic association under the expected relative risks of 1.1, 1.2, and 1.3, were 16.07%, 45.16%, and 75.23%, respectively. In order to improve the ability to detect low genetic effect, we conducted the association using the 10% highest BMIs as the outcome relative to the 10% lowest BMIs. As a result, the copy number of T allele of rs384662 and rs12916 in the *HMGCR* gene was associated with the BMI extremes with ORs of 0.48 (95% CI: 0.27-0.85, P=0.01) and 0.47 (95% CI: 0.27-0.83, P=0.01), respectively, while the reverse association was observed for the rs3846663 in the same gene with an OR of 2.15 (95% CI: 1.21-3.79, P=0.01). The above associations were all adjusted for the demographic and behavioral covariates, but none of the P values of the associations was under the bonferroni correcting significance criterion.

Haplotype Associations

Six candidate variants in our study associated with three haplotype blocks: rs3846662 and rs3846663 in the *HMGCR* gene formed the first block with the major haplotype phases of CT and TC (frequency>5%), rs174547 and rs174570 in the *FADS1* and *FADS2* gene formed the second block with the major haplotype phases of CT and TC (frequency>5%), and rs4938303 and rs1558861 that were close to the *BUD13-ZNF259-APOA5-A4-C3-A1* gene region formed the third block with the major haplotype phases of CC, CT, and TT (frequency>5%). The measures of linkage disequilibrium are provided in Figure 1. The block consisting of rs3846662 and rs3846663 was inversely associated with BMIs in the extremes value comparison with an OR of 0.39 (95% CI: 0.21-0.73, P<0.01). Another block containing rs4938303 and rs1558861 was found to be related with a reduced risk of dyslipidemia with an OR of 0.77 (95% CI: 0.59-1.00, P=0.05) (Table 3).

Gene-environment Interactions

The gene-environment interactions were examined with respect to the BMIs in both continuous and categorical variable patterns. Given the null findings in both individual variant and haplotype-based associations, the genetic polymorphisms of rs174547 and rs174570 were not included in the interaction analysis. If the variants were located in the same haplotype block, only one of them was allowed to enter the GMDR model once at a time to avoid the correlations within the model. The soy consumption frequency was fixed in every model as the environmental risk factor. The best models with highest testing balanced accuracy among potential variable combinations are listed in Table 4. When the continuous BMI value was chosen as the outcome, the SNP of rs3846662 or rs3846663 was observed in the best interaction model. When the BMI category divided by 25 kg/m^2 was used as the outcome, genetic polymorphism of rs4938303 or rs1558861 entered the best models as well. Additionally, when the analysis was conducted in the subjects with 10% tails of BMIs, every candidate genetic polymorphism was included in the best model. However, none of the best models survived the sign statistical test (P>0.05 for all comparisons) regardless of using which BMI outcome.

Combined Effect of Gene and Environment on BMI

Because of the existence in all the best models of gene-environment interaction and the previous significant findings in genetic association tests, we evaluated the joint effect of the polymorphism of rs3846662 or rs3846663 and the soy consumption frequency for the risk of overweight/obesity. The CC and CT genotypes of rs3846662 and TT and CT genotypes of rs3846663, respectively, were combined

Table 2. Associations of Genetic Polymorphisms with Dyslipidemia and BMI Categories*

Chromosome Location	dbSNP ID	Closest Gene	Minor Allele (MAF)	Pct_missing	HWE P Value	Dyslipidemia		Overweight/obesity		10% BMI Extremes	
						OR (95% CI)	*P* Value	OR (95% CI)	*P* Value	OR (95% CI)	*P* Value
1q25	rs599839	*PSRC1*	C (0.09)	0.32%	0.496	1.48 (1.05,2.09)	0.03	1.12 (0.77,1.63)	0.55	0.84 (0.28,2.53)	0.75
5q13	rs3846662	*HMGCR*	T (0.45)	0	0.014	0.88 (0.72,1.06)	0.18	0.90 (0.73,1.11)	0.31	0.48 (0.27,0.85)	0.01
5q13	rs3846663	*HMGCR*	C (0.45)	0	0.018	1.20 (0.99,1.45)	0.07	1.15 (0.94,1.42)	0.18	2.15 (1.21,3.79)	0.01
5q13	rs12916	*HMGCR*	T (0.44)	0.11%	<0.001	0.79 (0.66,0.96)	0.02	0.88 (0.72,1.08)	0.22	0.47 (0.27,0.83)	0.01
11q12	rs174547	*FADS1*	T (0.43)	0.21%	0.383	1.07 (0.87,1.31)	0.52	0.99 (0.80,1.23)	0.93	0.72 (0.38,1.39)	0.33
11q12	rs174570	*FADS2*	C (0.43)	0.21%	0.184	0.92 (0.75,1.12)	0.4	1.01 (0.81,1.26)	0.91	1.39 (0.72,2.69)	0.32
11q23	rs4938303	*BUD13*	T (0.48)	0.21%	0.817	0.89 (0.73,1.09)	0.26	0.90 (0.73,1.11)	0.33	0.82 (0.44,1.52)	0.53
11q23	rs1558861	*BUD13*	C (0.20)	0.11%	0.667	0.77 (0.61,0.99)	0.04	0.94 (0.72,1.23)	0.65	0.73 (0.36,1.51)	0.40

Note. SNP: single-nucleotide polymorphisms; BMI: body mass index; MAF: minor allele frequency; Pct_missing: percentage of genotypes missing; HWE: hardy-weinberg equilibrium; OR: odds ratio; CI: confidence interval. *Listed *P* values, odds ratios, and 95% confidence intervals were calculated using the genetic additive model of the T allele by multivariate logistic regression model, adjusted for age, gender, cigarette smoking, alcohol drinking, time spent sitting and soy consumption frequency.

Table 3. Haplotype Association for Dyslipidemia and BMI Categories*

SNPs	Haplotypes (freq>5%)	Phenotypes	Freq cases	Freq controls	OR (95% CI)	*P* Value
Block1						
rs174547, rs174570	CT	Dyslipidemia	0.57	0.57	reference	
	TC		0.43	0.43	1.08 (0.89,1.32)	0.43
rs174547, rs174570	CT	Overweight/obesity	0.58	0.57	reference	
	TC		0.42	0.43	1.00 (0.80,1.24)	0.96
rs174547, rs174570	CT	10% BMI extremes	0.57	0.51	reference	
	TC		0.43	0.49	0.75 (0.41,1.39)	0.36
Block2						
rs3846662, rs3846663	CT	Dyslipidemia	0.57	0.52	reference	
	TC		0.43	0.48	0.85 (0.70,1.04)	0.11
rs3846662, rs3846663	CT	Overweight/obesity	0.57	0.54	reference	
	TC		0.43	0.46	0.88 (0.71,1.09)	0.23
rs3846662, rs3846663	CT	10% BMI extremes	0.60	0.50	reference	
	TC		0.40	0.50	0.39 (0.21,0.73)	0.003
Block3						
rs4938303, rs1558861	CC	Dyslipidemia	0.21	0.18	reference	
	CT		0.32	0.33	0.80 (0.60,1.06)	0.12
	TT		0.47	0.49	0.77 (0.59,1.00)	0.05
rs4938303, rs1558861	CC	Overweight/obesity	0.20	0.19	reference	
	CT		0.33	0.32	1.01 (0.75,1.37)	0.95
	TT		0.46	0.49	0.90 (0.68,1.20)	0.49
rs4938303, rs1558861	CC	10% BMI extremes	0.21	0.20	reference	
	CT		0.29	0.33	0.78 (0.35,1.76)	0.55
	TT		0.49	0.48	0.73 (0.35,1.53)	0.40

Note. BMI: body mass index; SNPs: single-nucleotide polymorphisms; OR: odds ratio; CI: confidence interval. *Listed *P* values, odds ratios, and 95% confidence intervals were calculated using the multivariate logistic regression model, adjusted for age, gender, cigarette smoking, alcohol drinking, time spent sitting and soy consumption frequency.

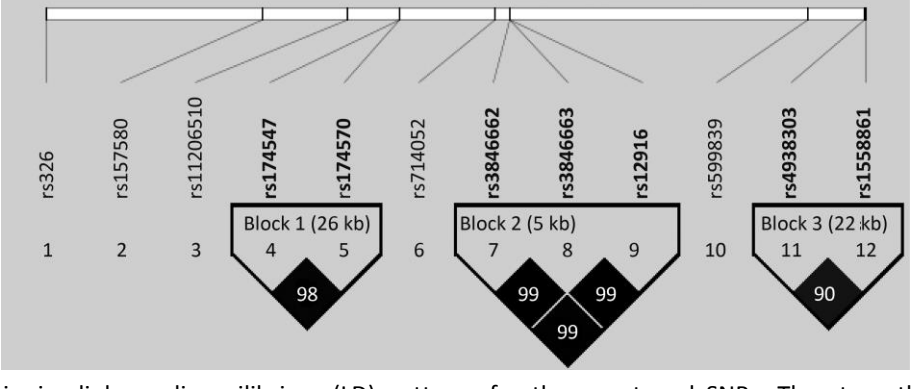

Figure 1. Pairwise linkage disequilibrium (LD) patterns for the genotyped SNPs. The strength of LD measured by D' multiplying 100 was displayed in the black diamonds and seven SNPs in three haplotype blocks were shown.

together to reduce the dimension of the data. In the crude analysis, we observed an increased risk of overweight/obesity for the subjects with homozygous and heterozygous genotypes for the C allele of rs3846662 as well as exposing to <1 time/week of soybean food intake, with an OR of 2.23 (95% CI: 1.22-4.05, *P*<0.01). Regarding the rs3846663, a nearly same size of effect was observed with regard to the combination of carrying homozygous and heterozygous genotypes for the T allele and experiencing the low frequency of soybean food consumption. However, the joint effects for the individual SNPs were no more statistically significant after adjusting for the demographic and behavioral covariates. When the haplotypes consisting of both SNPs were evaluated, exposure to both the haplotype phase of CT and the low frequency of soybean food consumption posed an increased risk for the overweight/obesity with an OR of 1.91 (95% CI: 1.39-2.63, *P*<0.01). The adjustment for the covariates compromised the association, but did not change the statistical significance (*P*<0.01) (Table 5). The power to identify the gene-environment interaction between the haplotype phase of CT and the frequency of soy food intake was estimated to be 63.90% under the parameters observed in our study.

DISCUSSION

The present study did not suggest the direct association between the lipid-metabolizing genetic polymorphisms and the BMIs. When the subjects with extreme BMIs were used, the rs3846662 and rs3846663 polymorphism in the *HMGCR* gene showed the evidence of association with the BMIs. Furthermore, the combined effects were observed between the SNPs and the soybean food intake frequency for the risk of overweight/obesity.

To our knowledge, although many of the lipid-metabolizing genes have been evaluated to see if they are associated with other traits such as coronary artery disease and carotid intima media thickness[16], no study has been conducted to explore their association with obesity. In our study, three susceptible polymorphisms for lipid metabolism were confirmed to be associated with dyslipidemia, however, there is no evidence to support their association with higher BMI. These results may be partly due to the limited statistical power of our study. When we used the 10% BMI extremes only as the outcome variables, in which case the observed association tended to be more significant, the results on all three selected SNPs in the *HMGCR* gene reached the significance level of α=0.05. However, the *P* values did not satisfy the threshold after bonferroni correction for multiple testing.

Table 4. Best Models of Gene-environment Interaction by Generalized Multifactor Dimensionality Reduction for BMI Status*

Included Genetic Factors†	Best Model	Testing bal. accu.	CVC	*P* Value‡
BMI				
rs4938303, rs3846662, rs599839	rs3846662, soy consumption frequency	0.5430	7/10	0.3770
rs4938303, rs3846663, rs599839	rs3846663, soy consumption frequency	0.5516	8/10	0.3770
rs1558861, rs3846662, rs599839	rs3846662, soy consumption frequency	0.5501	7/10	0.3770
rs1558861, rs3846663, rs599839	rs3846663, soy consumption frequency	0.5640	9/10	0.1719
BMI≥25 versus BMI<25				
rs4938303, rs3846662, rs599839	rs4938303, rs3846662, soy consumption frequency	0.4967	10/10	0.6230
rs4938303, rs3846663, rs599839	rs3846663, soy consumption.. frequency..	0.5155	8/10	0.8281
rs1558861, rs3846662, rs599839	rs1558861, rs3846662, soy consumption.. frequency	0.5321	10/10	0.0547
rs1558861, rs3846663, rs599839	rs1558861, rs3846663, soy consumption frequency	0.5307	10/10	0.1719
The highest 10% of BMI versus the lowest 10%				
rs4938303, rs3846662, rs599839	rs4938303, rs3846662, rs599839, soy consumption frequency	0.5605	10/10	0.3770
rs4938303, rs3846663, rs599839	rs4938303, rs3846663, rs599839, soy consumption frequency	0.5457	10/10	0.3770
rs1558861, rs3846662, rs599839	rs1558861, rs3846662, rs599839, soy consumption frequency	0.5996	10/10	0.0547
rs1558861, rs3846663, rs599839	rs1558861, rs3846663, rs599839, soy consumption frequency	0.5785	10/10	0.0547

Note. BMI: body mass index; CVC: cross validation consistency. *Adjusted by age, gender, cigarette smoking, alcohol drinking, and time spent sitting. †Soy consumption frequency as the environmental factor was fixed in the model. ‡Based on sign statistical test.

Table 5. Combined Association of Genotypes and Haplotypes of rs3846662 and rs3846663 and Soy Consumption with Overweight/obesity (BMI ≥ 25 kg/m^2)*

Variable	Soybean Consumption Frequency	Crude		Adjusted†	
		OR (95%CI)	*P* Value	OR (95%CI)	*P* Value
rs3846662					
TT	≥1 times/week	reference		reference	
CC&CT	≥1 times/week	1.43 (0.75,2.70)	0.27	1.15 (0.58,2.31)	0.69
TT	<1 time/week	1.60 (0.81,3.14)	0.17	1.19 (0.57,2.47)	0.65
CC&CT	<1 time/week	2.23 (1.22,4.05)	<0.01	1.69 (0.88,3.25)	0.12
rs3846663					
CC	≥1 times/week	reference		reference	
CT&TT	≥1 times/week	1.43 (0.75,2.70)	0.27	1.16 (0.58,2.32)	0.68
CC	<1 time/week	1.57 (0.80,3.08)	0.19	1.14 (0.55,2.39)	0.72
CT&TT	<1 time/week	2.24 (1.23,4.07)	<0.01	1.72 (0.89,3.30)	0.11
Haplotype phases(rs3846662, rs3846663)					
TC	≥1 times/week	reference		reference	
CT	≥1 times/week	1.28 (0.89,1.85)	0.18	1.20 (0.80,1.80)	0.37
TC	<1 time/week	1.69 (1.22,2.35)	<0.01	1.45 (1.01,2.08)	0.05
CT	<1 time/week	1.91 (1.39,2.63)	<0.01	1.64 (1.15,2.34)	<0.01

Note. OR: odds ratio; CI: confidence interval. * Listed *P* values, odds ratios, and 95% confidence intervals were calculated using the logistic regression model. †Adjusted for age, gender, cigarette smoking, alcohol drinking and time spent sitting.

Haplotypes based on LD are widely spread in the human genome and a few common haplotypes (2-4 per block) can represent a large proportion of the variation of a block[22]. Therefore, the haplotype association would seem to be more powerful in detecting susceptible loci[23]. We constructed the haplotype blocks for the candidate SNPs located nearby and the haplotype associations confirmed the association of rs1558861 with dyslipidemia and rs3846662 and rs3846663 with BMI extremes in the associations for individual genetic markers.

3-hydroxy-3-methylglutaryl (HMG)-coenzyme (CoA) reductase (HMGCR) is a rate-limiting enzyme that regulates cholesterol synthesis and is the target of statin therapy[24]. The *HMGCR* gene is located in the chromosome 5q13-14 region and several common polymorphisms of the gene have been reported to be able to influence the baseline LDL-C level[25]. The SNPs of rs3846662, rs3846663, and rs12916, which are located in the intronic area and 3'UTR of the *HMGCR* gene, are in tight linkage disequilibrium between each other. The variant of rs3846663 was first identified in a GWA study with the effect to lower LDL-C level[26]. Then the SNP of rs3846662 in the vicinity of 3846663 was found to be able to modulate the alternative splicing of exon 13[13]. The deletion of exon13 can influence the stability of *HMGCR*[27], thus leading to more rapid degradation of the protein. Then the increased uptake of LDL-C from plasma will complement the decreased synthesis and maintain the intracellular cholesterol homeostasis[13]. Our study has suggested that the common SNPs of *HMGCR* may relate to the BMIs, which is supported by some other studies. For example, in a GWA study on BMI, a significant polymorphism of rs211234 was identified neighboring the *HMGCR* gene and in tight linkage disequilibrium with the variants located within the *HMGCR* gene[11]. However, the underlying mechanism of the finding was still largely unknown.

Gene-environment interactions are very common in the pathogenesis of common complex disorders like dyslipidemia and obesity. In our study, the dietary intake of soybeans as well as the SNPs associating with dyslipidemia or BMIs, were taken together in the GMDR model. The best models highlighted the existence of potential interactions between soy food consumption and the function of

genetic loci in *HMGCR*. Combined association analyses showed that people who seldom ate soy foods and owned the effect genotypes/haplotypes of *HMGCR* had the largest risk to suffer from overweight/ obesity, further supporting the results from the best models. The proteins contained in soybeans have the complete profile of essential amino acid, while the calories are lower than in the animal proteins. In addition, plenty of the n-6 poly-unsaturated fatty acids (PUFAs), especially the linoleic acid, are contained in the soybeans, which may contribute to the improvement of lipids metabolism and the body weight control[28]. Observational studies have shown that the inclusion of soybeans in the daily diet can improve some components of metabolic syndrome[29] and lower the levels of many inflammation cytokines[30]. Clinical trials have also suggested the positive role of soy meal in reducing the body weight[7]. Animal studies have found that feeding rats with soy protein can change the expression of more than 90 genes involved in the metabolic functions of adipocytes and may reduce the excess accumulation of adipose tissue[31]. The results of our study have suggested that the genetic determinants of lipid metabolism may possibly modify the effect of dietary soy intake, although the exact mechanisms underlying this interaction remain to be elucidated.

Our study has several limitations. First, as the study is based on a cross-sectional survey, and the detected associations might be subject to the measurement errors arising from changes in the environmental factors after the obesity occurrence. Therefore, some of the associations may have been obscured. However, this bias may have had only a minor influence on the overall results, because the observed distributions of susceptible behavioral factors between BMI categories were all statistically significant. Second, we investigated only a limited number of candidate variants identified from GWA studies and only 2-3 polymorphisms per gene, but the complex disorders could be attributable to many genetic polymorphisms with low to modest effect. Thus, the amount of genetic variation of the gene captured by the genotyped variants may be limited. At the same time, the evaluation of the genetic susceptibility to BMIs in the present study may not be comprehensive enough. Third, BMI is a general measurement of body fat mass and some studies have found that adipose tissues in different parts of the body are associated with varying risks for cardiovascular diseases[32]. The anthropometry of abdominal obesity, as well as some intermediate traits representing the visceral fat, may serve as more specific phenotypes to more accurately estimate the effects of genetic factors. Fourth, asking frequency of foods intake was the most popular method used to measure the amount of intakes, but it was just a crude estimation. Further replications of our study by using the more accurate methods, such as 24 h dietary recalls survey and weighing the exact amount of foods, would be desirable. Fifth, the relatively small sample size in our study has led to the limited statistical power to detect the single variant association and gene-environment interaction. Finally, we cannot provide an independent cohort to confirm the findings from the present study so that more robust conclusions can be drawn.

In conclusion, no individual SNP from the studied lipids-metabolizing genes or haplotypes constructed by these SNPs showed the statistically significant association with BMIs. However, two susceptible variants from the *HMGCR* gene were identified, and the combined effect for overweight/obesity between them and soy foods intake frequency was suggested. These findings warranted further examination in the cohorts with larger sample size and more susceptible loci genotyped.

ACKNOWLEDGEMENTS

We thank the staff of the primary health care centers in Fangshan district of Beijing for their work in recruiting the study subjects.

REFERENCES

1. World Health Organization: Global Database on Body Mass Index (BMI). [http://www.who.int/mediacentre/factsheets/fs311/en/index.htm]
2. Cecchini M, Sassi F, Lauer JA, et al. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. Lancet, 2010; 376, 1775-84.
3. Field AE, Coakley EH, Must A, et al. Impact of overweight on the risk of developing common chronic diseases during a 10-year period. Arch Intern Med, 2001; 161, 1581-6.
4. Must A, Spadano J, Coakley EH, et al. The disease burden associated with overweight and obesity. JAMA, 1999; 282, 1523-9.
5. Fontaine KR, Redden DT, Wang C, et al. Years of Life Lost Due to Obesity. JAMA, 2003; 289, 187-93.
6. Lee M and Kim I. Soy protein and obesity. Nutrition, 2003; 16, 459-60.

7. Allison DB, Gadbury G, Schwartz LG, et al. A novel soy-based meal replacement formula for weight loss among obese individuals: a randomized controlled clinical trial. Eur J Clin Nutr, 2003; 57, 514-22.
8. Malis C, Rasmussen EL, Poulsen P, et al. Total and regional fat distribution is strongly influenced by genetic factors in young and elderly twins. Obes Res, 2005; 13, 2139-45.
9. Lee J, Chen L, Snieder H, et al. Heritability of obesity-related phenotypes and association with adiponectin gene polymorphisms in the Chinese national twin registry. Ann Hum Genet, 2010; 74, 146-54.
10. Rankinen T, Zuberi A, Chagnon YC, et al. The human obesity gene map: the 2005 update. Obesity (Silver Spring), 2006; 14, 529-644.
11. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet, 2010; 42, 937-48.
12. Eckel RH, Grundy SM, and Zimmet PZ. The metabolic syndrome. Lancet, 2005; 365, 1415-28.
13. Burkhardt R, Kenny EE, Lowe JK, et al. Common SNPs in HMGCR in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon13. Arterioscler Thromb Vasc Biol, 2008; 28, 2078-84.
14. Waterworth DM, Ricketts SL, Song K, et al. Genetic variants influencing circulating lipid levels and risk of coronary artery disease. Arterioscler Thromb Vasc Biol, 2010; 30, 2264-76.
15. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet, 2009; 41, 56-65.
16. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet, 2009; 41, 47-55.
17. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet, 2008; 40, 189-97.
18. Wallace C, Newhouse SJ, Braund P, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. Am J Hum Genet, 2008; 82, 139-49.
19. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, 2005; 21, 263-5.
20. Stephens M, Smith NJ, and abd Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet, 2001; 68, 978-89.
21. Moore JH, Gilbert JC, Tsai CT, et al. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol, 2006; 241, 252-61.
22. Wall JD and Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet, 2003; 4, 587-97.
23. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science, 2002; 296, 2225-9.
24. Medina MW and Krauss RM. The role of HMGCR alternative splicing in statin efficacy. Trends Cardiovasc Med, 2009; 19, 173-7.
25. Chen YC, Chen YD, Li X, et al. The HMG-CoA reductase gene and lipid and lipoprotein levels: the multi-ethnic study of atherosclerosis. Lipids, 2009; 44, 733-43.
26. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet, 2008; 40, 161-9.
27. Cheng HH, Xu L, Kumagai H, et al. Oligomerization state influences the degradation rate of 3-hydroxy-3-methylglutaryl-CoA reductase. J Biol Chem, 1999; 274, 17171-8.
28. Bhattacharya A, Rahman MM, Sun D, et al. The Combination of Dietary Conjugated Linoleic Acid and Treadmill Exercise Lowers Gain in Body Fat Mass and Enhances Lean Body Mass in High Fat–Fed Male Balb/C Mice. J Nutr, 2005; 135, 1124-30.
29. Mattei J, Hu FB, and Campos H. A higher ratio of beans to white rice is associated with lower cardiometabolic risk factors in Costa Rican adults. Am J Clin Nutr, 2011; 94, 869-76.
30. Azadbakht L, Kimiagar M, Mehrabi Y, et al. Soy consumption, markers of inflammation, and endothelial function: a cross-over study in postmenopausal women with the metabolic syndrome. Diabetes Care, 2007; 30, 967-73.
31. Frigolet ME, Torres N, Uribe-Figueroa L, et al. White adipose tissue genome wide-expression profiling and adipocyte metabolic functions after soy protein consumption in rats. J Nutr Biochem, 2011; 22, 118-29.
32. Matsuzawa Y, Nakamura T, Shimomura I, et al. Visceral fat accumulation and cardiovascular disease. Obes Res, 1995; 3, 645S-7S.