npg

ORIGINAL ARTICLE

A nonsynonymous SNP in *BANK1* is associated with serum LDL cholesterol levels in three Korean populations

Kyung-Won Hong[1], Jieun Lyu[1], So Hyun Lee[1], Bo Youl Choi[2], Sung Soo Kim[1] and Yeonjung Kim[1]

Serum levels of lipids, such as cholesterol and triglycerides, are heritable risk factors for cardiovascular disease and targets for therapeutic intervention. Because previous genome-wide association studies (GWASs) did not target functional genetic variants, we employed an alternate approach using nonsynonymous single-nucleotide polymorphisms (SNPs) to identify functional genetic variants associated with the regulation of serum lipid levels. We selected 3667 healthy individuals from a rural community-based cohort (CAVAS; Cardio Vascular disease Association Study) of the Korean Genome and Epidemiology Study project. We analyzed demographic and lifestyle information, lipid measurements and genotypes using the Illumina-1M SNP chip. For genotyping, we isolated 11 558 nonsynonymous SNPs and conducted a linear regression analysis with four lipid traits (total, high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterols and triglycerides). Significantly associated SNPs were validated in two independent Korean populations, Korean Association Resource (KARE) ($n = 4116$) and Health Examinee (HEXA) ($n = 2178$). Of the 11 558 SNPs, one SNP (rs3733197) from the CAVAS was significantly associated with serum LDL cholesterols (beta ± s.e. = 4.67 ± 0.94, P-value = 1.0×10^{-6} and Bonferroni corrected P-value = 0.012). The replication results of HEXA and KARE were beta ± s.e. = 2.88 ± 1.12, P-value = 0.016 and beta ± s.e. = 1.26 ± 0.97, P-value = 0.196, respectively. An overall meta-analysis of the three data sets revealed beta = 2.98 ± 0.57, P-value = 6.19×10^{-7}. The rs3733197 is located in the coding region of *BANK1* (B-cell scaffold protein with ankyrin repeats 1), and the minor allele (A) resulted in the replacement of the Alanine at position 383 with Threonine.

INTRODUCTION

Serum concentrations of lipids, such as cholesterol and triglycerides, are heritable risk factors for cardiovascular disease and targets for therapeutic intervention.[1,2] To date, 19 genome-wide association studies (GWASs) and over 200 single-nucleotide polymorphisms (SNPs) have been reported for lipid traits and are listed online in the catalog of published GWASs.[3] Recently, Kim *et al.*[4] reported a GWAS of metabolic traits, including lipid levels, in East Asians[4] that implicated 18 lipid level influencing loci, including three novel loci—MYL2, OAS3 and C12orf51.

In conventional approaches, multiple testing of several hundred thousand SNPs makes the detection of an association difficult, and the variants detected by Kim *et al.*[4] cannot be categorized into functional variants such as nonsynonymous SNPs (nsSNPs). [4] By selecting SNPs according to their functional significance, it is possible to overcome the problem of an overwhelming number of SNPs and to identify the candidate SNPs for experimental validation.[5] The nsSNPs are located in coding regions and result in amino-acid changes in the protein products of genes; they are thought to be the class of SNPs that have

the greatest impact on phenotype. Several studies have suggested the impact of amino-acid allelic variants on protein structure and function and on human disease.[6–8]

Here, we used an alternative approach based on nsSNP analysis of GWAS data. A total of 11 558 nsSNPs, distributed genome wide, were analyzed for an association with lipid levels in a new Korean cohort (CAVAS; Cardio Vascular disease Association Study), which is a subset of the Korean Genome and Epidemiology Study (KoGES), and tested for data replication in another two independent populations.

MATERIALS AND METHODS

Study subjects
Study subjects were selected from a cohort of the ongoing KoGES. Participants were recruited among the residents of three cities located in rural areas of Korea: Yang-pyeong, Nam-won and Go-ryeong. In this study, the study population termed CAVAS included at the time a total of 4052 participants (aged 38–89 years) who were recruited from 2004 to 2008. Written informed consent was obtained from all participants, and this research project was approved by the institutional review board of the Korea National Institute of Health.

[1]Division of Epidemiology and Health Index, Center for Genome Science, Korea Centers for Disease Control and Prevention, Cheongju-si, Korea and [2]Department of Preventive Medicine, College of Medicine, Hanyang University, Seoul, Korea
Correspondence: Dr Y Kim, Division of Epidemiology and Health Index, Center for Genome Science, Korea Centers for Disease Control and Prevention, #200 Osongsaengmyeong2-ro, Osong-eup, Cheongju-si, Chungcheongbuk-do 363-951, Korea.
E-mail: kimye@korea.kr

The CAVAS subjects did not have self-reported metabolic diseases (type 2 diabetes, hypertension or hyperlipidemia), cardiovascular diseases (myocardial infarction or stroke) or cancers. In addition, they had a normal range of blood pressure (systolic blood pressure <130 mm Hg and diastolic blood pressure <90 mm Hg) and fasting glucose levels (<126 mg dl^{-1}), as measured during health examinations of the cohort members.

For the replication analyses, we used two independent Korean populations, from the Korean Association Resource (KARE) and Health Examinee (HEXA) study cohorts. The KARE is a consortium for analyzing the Ansung–Ansan population, which is a part of the KoGES project, and the HEXA is a cohort from two metropolitan areas (Seoul and Pusan), which is part of the KoGES project. The initial populations of the KARE and HEXA cohorts included 8842 and 3703 patients, respectively, and we applied the same inclusion criteria that were used for the CAVAS cohort. Ultimately, 3679 and 2123 individuals, respectively, were selected for the replication study. Detailed descriptions of the KARE and HEXA cohorts have been previously reported elsewhere.[4,9]

Parameters were measured for fasting serum lipid levels (HDL cholesterol, triglycerides and total cholesterol). Biochemical measurements were obtained in the morning before the first meal of the day. The concentration of LDL cholesterol was calculated using Friedewald's formula (LDL cholesterol = total cholesterol − HDL cholesterol − (triglycerides/5)). Missing values were assigned for the individuals with TG >400 mg dl^{-1}.

Genotyping and quality control
Genomic DNA samples, isolated from peripheral blood drawn from CAVAS cohort participants, were genotyped using the Illumina Human 1M-duo Beadchip (Illumina Inc., San Diego, CA, USA). A total of 4052 samples were genotyped using 500 ng genomic DNA with an Illumina chip. The Bayesian Robust Linear Modeling using Mahalanobis Distance (BRLMM) genotyping algorithm was used for genotype calling of 1 010 624 SNPs.[10] Subjects with genotype accuracies below 98% and high missing genotype call rates ($\geqslant 4$%),

high heterozygosity (>30%) or inconsistency in sex were excluded from subsequent analyses. Individuals were related individuals whose estimated identity-by-state values were high (>0.80). The methods used to estimate heterozygosity and identity-by-state have been described elsewhere.[10] After these quality control steps, 3667 samples were selected. Markers with a high missing gene call rate (>5%), low minor allele frequency (<0.01) or significant deviation from Hardy–Weinberg equilibrium ($P<1\times10^{-6}$) were excluded, leaving a total of 723 056 markers to be examined in 3667 individuals.

Functional SNP selection
We previously reported nsSNP GWAS results for blood pressure regulation and described the nsSNP selection methods.[11] Briefly, using BIOMART version 0.7 (http://www.ensembl.org/biomart/), nsSNPs were selected from the 723 056 SNPs based on the data in dbSNP v.135. BIOMART is a web-based tool that allowed us to use information from the latest release of a database to select SNPs limited to a specific nonsynonymous category.

Estimation of genetic variance
The genetic variances were computed via GCTA v1.24,[12] which is a tool for estimating the proportion of phenotypic variance that is explained by genome-wide SNPs for complex traits.[13] First, we estimated the pairwise genetic relationship using the make-grm option for the nsSNPs and all SNPs on the array. We then estimated the proportion of phenotypic variance contributed by the nsSNPs and all SNPs, respectively, based on the restricted maximum likelihood.[14]

Replication study
Because the KARE and HEXA cohorts were genotyped using the Affymetrix 5.0 and 6.0 SNP arrays (Affymetrix, Santa Clara, CA, USA), respectively, only one-third of the SNPs on the Illumina Beadchip were available. For this reason, we used imputed SNPs reported in a previous study.[4,9] Briefly, we used the IMPUTE program (http://mathgen.stats.ox.ac.uk/impute/) for SNP imputation based on the International HapMap data (http://hapmap.ncbi.nlm.nih.gov/; phase 2, release 22 k NCBI build 36 and dbSNP build 126), including 2.2 million SNPs from 90 individuals from the JPT and CHB populations for use as a reference panel.[4,14,15] After excluding imputed SNPs with low genotype information content (<0.05), a posterior probability score <0.90, a call rate <0.90, a minor allele frequency <0.01 or Hardy–Weinberg equilibrium $P<1\times10^{-6}$, a total of 1 573 409 SNPs for the KARE cohort and 1 984 393 SNPs for the HEXA cohort were selected.

Statistical analysis
In our quantitative analysis of HDL cholesterol, LDL cholesterol, triglyceride and total cholesterol levels, linear regression was used by controlling for the age, sex and body mass index of the cohort subjects. To understand the environmental effect, we added smoking and alcohol drinking habits as the covariates.

Statistical analyses were performed using PLINK (version 1.07).[16] We conducted GWASs by 723 056 SNPs (all SNPs) and 11 558 nsSNPs. All tests were based on an additive model, and *P*-values were not adjusted for multiple tests. Significantly associated SNPs were examined in the replication study. The inverse-variance meta-analysis method, assuming fixed effects, was used to generate meta-analysis statistics. Cochran's *Q*-test was used to assess between-study heterogeneity.[17] All meta-analysis calculations were performed using PLINK (version 1.07).[16] Possible damaging effects of nonsynonymous SNPs on protein structure were predicted using PolyPhen 2,[18] SIFT,[19] Mutation Taster[20] and LRT predictor.[21]

Population stratification was tested by a principal component analysis using the PLINK.[16] To prevent overrepresentation of regions with more redundant SNPs, we used the indep-pairwise command in PLINK[16] to reduce linkage disequilibrium between the remaining variants by eliminating any SNP that had a pairwise $r^2>0.3$ with any other SNP in a 1500 bp window (step size, 150 bp). This reduced the CAVAS data set to 148 902 SNPs; subsequently, we selected 19 159 SNPs commonly presented in KARE and HEXA, and 545 SNPs commonly presented in HapMap 3 ($n=957$) in which the plink format was obtained from GitHub (https://github.com/gabraham/flashpca). Using the

Table 1 Basic patient characteristics and lipid measurements

	Discovery population	Validation population	
	CAVAS	KARE	HEXA
Categorical variables			
Sex			
Male/female	1350/2243	1861/2255	843/1335
Smoking			
Never (%)	2420 (67.7)	1918 (46.9)	1550 (71.4)
Quit (%)	545 (15.2)	252 (6.1)	335 (15.4)
Current (%)	609 (17.1)	1923 (47.0)	286 (13.2)
Alcohol consumption			
Never (%)	1777 (49.7)	2456 (60.5)	999 (46.0)
Quit (%)	377 (10.5)	547 (13.4)	86 (4.0)
Current (%)	1421 (39.8)	1068 (26.1)	1078 (50.0)
Continuous variables (mean ± s.d.)			
Age (year)	59.8 ± 10.1	49.2 ± 8.0	51.0 ± 7.6
Height (cm)	157.5 ± 8.5	160.3 ± 8.4	161.2 ± 8.0
Body mass index (kg m^{-2})	23.8 ± 3.1	24.0 ± 2.9	23.4 ± 2.7
Total cholesterol (mg dl^{-1})	195.9 ± 35.6	187.1 ± 33.4	196.0 ± 33.5
HDL cholesterol (mg dl^{-1})	45.5 ± 10.9	45.3 ± 9.9	56.2 ± 13.4
Triglyceride (mg dl^{-1})	135.9 ± 84.6	142.4 ± 89.7	107.2 ± 62.3
LDL cholesterol (mg dl^{-1})	124.1 ± 32.0	114.2 ± 30.1	118.5 ± 30.1

Abbreviations: CAVAS, Cardio Vascular disease Association Study; HDL, high-density lipoprotein; HEXA, Heath Examinee; KARE, Korean Association Resource; LDL, low-density lipoprotein.
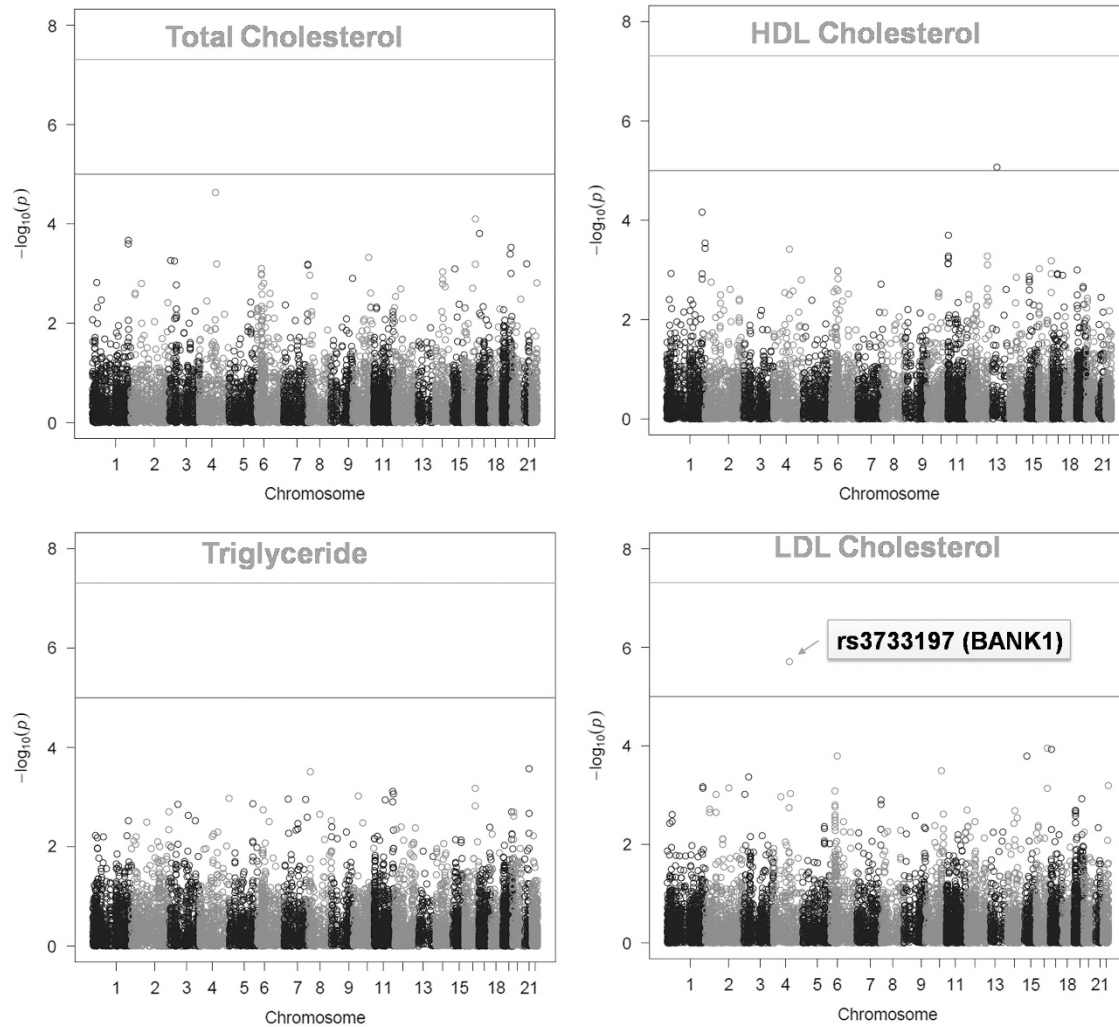
Figure 1 A Manhattan plot of nonsynonymous single-nucleotide polymorphisms (SNPs) from the genome-wide association studies (GWASs) that are associated with lipid traits. SNPs are plotted based on their physical chromosomal positions (horizontal axis) together with their $-\log_{10}(P\text{-values})$ in the GWAS (vertical axis). The blue line indicates the suggestive threshold and the red line indicates the genome-wide significance threshold of $P = 5 \times 10^{-8}$. HDL, high-density lipoprotein; LDL, low-density lipoprotein. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

Table 2 The proportion of phenotypic variance explained by genotypic variance (Vg/Vp)

| Variables | Vp | Nonsynonymous SNPs (n = 11 558) | | All SNPs (n = 723 056) | |
		Vg	Vg/Vp	Vg	Vg/Vp
HDL cholesterols	112.09 (2.63)	0.67 (3.01)	0.006 (0.027)	1.93 (7.94)	0.018 (0.071)
LDL cholesterols	940.42 (22.21)	39.79 (26.97)	0.042 (0.029)	50.84 (72.79)	0.054 (0.077)
Triglycerides	6791.311 (159.125)	0.007 (179.45)	0.000001 (0.0264)	0.007 (551.51)	0.000001 (0.081)
Total cholesterols	1181.39 (27.68)	21.99 (32.47)	0.019 (0.027)	56.87 (86.72)	0.048 (0.073)

Abbreviations: HDL, high-density lipoprotein; LDL, low-density lipoprotein; SNPs, single-nucleotide polymorphisms; Vg, genetic variance; Vp, phenotypic variance.

545 common SNPs, we computed the principal components by Plink –pca command.

RESULTS

Discovery GWAS by all SNPs

The clinical characteristics of the CAVAS, KARE and HEXA cohorts are described in Table 1. In the CAVAS cohort, we conducted the conventional GWASs for total cholesterol, HDL cholesterol, LDL

cholesterol and triglyceride levels by 723 056 SNPs. Manhattan plots and quantile–quantile plot (QQ plot) were illustrated in Supplementary Figures S1–S4. Unfortunately, there was no SNP that passed the genome-wide significant level (P-values $< 5 \times 10^{-8}$). The genomic inflation factors from 1.000 (triglycerides) to 1.045 (LDL cholesterol) were not significantly deviated from the expected line.

We then decided to test the previously reported SNPs to be associated with lipid traits from GWAS catalogs (Supplementary

Table 3 Significant results from the nsSNP GWAS for LDL cholesterol levels and validation in two independent populations (the KARE and HEXA cohorts)

CHR	SNP	BP	Gene symbol	Amino-acid change	A1	Covariates	Cohort	n	MAF	BETA	s.e.	P-value	Cochran's Q	Heterogeneity
4	rs3733197	103058310	BANK1	Ala383Thr	A	Age, sex and BMI	CAVAS	3667	0.17	4.67	0.98	1.0×10^{-6}		
							KARE	4116	0.16	1.26	0.97	0.196		
							HEXA	2178	0.17	2.89	1.20	0.016		
							Meta-analysis	9961		2.98		6.19×10^{-7}	0.037	69.57
						Age, sex, BMI and alcohol drinking	CAVAS			4.79	0.98	9.8×10^{-7}		
							KARE			1.38	0.88	0.118		
							HEXA			2.97	1.20	0.013		
							Meta-analysis			2.93		3.6×10^{-7}	0.035	70.19
						Age, sex, BMI and smoking	CAVAS			4.80	0.98	1.0×10^{-6}		
							KARE			1.23	0.89	0.165		
							HEXA			3.07	1.20	0.011		
							Meta-analysis			2.89		5.2×10^{-7}	0.026	72.6
						Age, sex, BMI, smoking and alcohol drinking	CAVAS			4.81	0.98	8.7×10^{-7}		
							KARE			1.31	0.89	0.139		
							HEXA			2.99	1.20	0.013		
							Meta-analysis			2.92		4.1×10^{-7}	0.030	71.58

Abbreviations: A1, minor allele; BP, base pair; CAVAS, Cardio Vascular disease Association Study; CHR, chromosome; Cochran's Q, population heterogeneity *P*-value of meta-analysis; HEXA, Heath Examinee study; KARE, Korean Association Resource; *n*, individual numbers; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

Table S1).[22] Among 242 SNPs reported for the four lipid traits (total cholesterol, HDL cholesterol, LDL cholesterol and triglyceride), we included 122 SNPs available for our genotype data, of which small fraction (23 SNPs) showed *P*-value < 0.05. (Supplementary Table S1).

Alternative approach GWAS by nsSNPs

In the CAVAS cohort, we conducted the alternative approach GWAS for total cholesterol, HDL cholesterol, LDL cholesterol and triglyceride levels using 11 558 nsSNPs. For multiple comparisons, we detected significant associations based on the Bonferroni correction criteria (P-value < 4.3×10^{-6} (0.05/11 558)). We plotted $-\log_{10}(P)$ values against chromosomal position on Manhattan plots (Figure 1), and all significant associations are described in Table 3.

Through the discovery GWAS by using the nsSNPs, we identified one SNP (rs3733197) for LDL cholesterol levels that passed the Bonferroni correction criteria. Individuals with the minor allele of rs3733191 showed significantly increased LDL cholesterol levels compared with those with the major allele (beta ± s.e. = 4.67 ± 0.94, P-value = 1.0×10^{-6}). The rs3733197 nsSNP is located in the coding region of *BANK1* (B-cell scaffold protein with ankyrin repeats 1) gene-coding region, and the minor allele replaces the alanine at position 383 with threonine. Interestingly, SNP rs3733197 was previously reported to be a candidate variant for systemic lupus erythematosus,[23,24] psoriasis[25] and systemic sclerosis,[26,27] which are all diseases that are affected by serum cholesterol levels. [28]

To validate this association, we genotyped rs3733197 in other independent populations (the KARE and HEXA cohorts). The SNP for LDL cholesterol was significantly replicated in the HEXA cohort (beta ± s.e. = 2.88 ± 1.12, P-value = 0.016), but was not significant in the KARE cohort (beta ± s.e. = 1.26 ± 0.97, P-value = 0.196). Furthermore, the meta-analysis showed a significant association; P-value = 6.19×10^{-7}, but the Cochran's Q (0.013) and Heterogeneity I (69%) were also significant.

Estimation of the phenotypic variance explained by genetic variances (Vg/Vp)

The Vg/Vp were estimated by 723 056 all SNPs and 11 558 nsSNPs by restricted maximum likelihood method[14] and the results were described in Table 2. The all SNPs examined were ~70 times more frequently found than the nsSNPs; however, compared wlith those explained by all SNPs, the Vg/Vp by nsSNPs was similar for LDL cholesterols (4.2% in nsSNPs vs 5.4% in all SNPs), a third proportions for total cholesterols (1.9% in nsSNPs vs 4.8% in all nsSNPs) and for HDL cholesterols (0.6% in nsSNPs vs 1.8% in all SNPs). Therefore, we assumed that the nsSNPs can explain the large proportion of phenotypic variances of lipid traits compared with the all SNPs.

DISCUSSION

The nsSNP rs3733197 discovered in our study has never been reported in previous GWAS for the lipid traits. Our study subjects were selected from the general population cohorts (CAVAS, KARE and HEXA), who showed normal characteristics in any of the intermediate phenotypes for metabolic diseases (that is, diabetes, hypertension and hyperlipidemia) and cancers. When we ran all SNP-GWAS using the subjects without exclusion of diabetes or hypertension, we found no significantly associated SNP (data not shown). We, therefore, suggest that nsSNP association might be attributable to lipid level in the healthy populations, and the effects could be diluted by the other risk genetic variant effects in the mixed populations.

The minor allele frequencies of SNP rs3733197 in three cohorts were similar (see Table 3), but the Cochran's *Q* (0.013) and Heterogeneity I (69%) were also significant. Therefore, we tried to find whether there was any population stratification that could lead us to spurious association. First of all, we conducted the Principal Component analysis by using the genome-wide SNPs of three cohorts (CAVAS, KARE and HEXA). As shown in Supplementary Figure S5, the genotypes of those cohorts were similarly aggregated, indicating no differences of the genetic constructions. Next, we looked into the environmental factors such as the smoking and alcohol drinking habits, which might differentially effect on the association results. The KARE cohort had more proportion of current smokers. We, therefore, tested the association between rs3733197 and LDL cholesterols using the smoking and alcohol drinking as the covariates. The results showed slightly increased significance for all three cohorts, but the *P*-value of KARE was still non-significant (see Table 3).

BANK1 has four functional domains (DBB, ANK1, ANK2 and interaction with ITPR2); rs3733197 is located in the ANK2 domain, which is known to function as protein–protein interaction motif. On the basis of the Ala to Thr substitution caused by rs3733197, this SNP was predicted to be 'probably damaging' to the BANK1 protein structure as the Poly-Phen risk score was 0.983. The Poly-Phen risk score indicates the Bayes posterior probability that a given mutation is damaging. Although PolyPhen risk score suggested a high probability to reduce protein stability or function,[16] other prediction softwares showed the mild effect of the variants for which the score of SIFT was 0.27 (tolerate),[19] and the score of Mutation Taster predicted the harmless and polymorphism.[20] The score of LRT predictor showed neutral.[21]

BANK1 is predominantly expressed in B cells, but low expression can be detected in other cell populations.[29] BANK1 functions in B-cell receptor-induced calcium mobilization from intracellular stores.[30] In addition, the protein can promote tyrosine phosphorylation of inositol 1,4,5-triphosphate receptors.[31] Mice homozygous for a *Bank1* knock-out allele exhibit enhanced B-cell responses.[32]

Increased cholesterol levels are associated with a greater abundance of B cells.[32] In our GWAS, individuals with the minor allele of rs3733197 showed significantly increased cholesterol levels. Therefore, it is possible that the genotype effect of rs3733197 may contribute to B-cell activation by interactions with cholesterol levels.

Interestingly, a recent clinical study indicated that statins (a class of cholesterol-lowering drugs), which were originally developed to treat lipid disorders,had immune-modulating effects in systemic lupus erythematosus patients.[33] Systemic lupus erythematosus [MIM:152700] is a chronic autoimmune disorder of the connective tissues, such as the skin, joint, and kidney. Therefore, our results suggest that the effects of rs3733197 and statins on systemic lupus erythematosus may be related, and an improved understanding of the role of LDL cholesterol in these contexts should be a goal of future studies.

In conclusion, we identified a novel functional variant of *BANK1* that affects the serum lipid levels in three Korean cohorts. Our findings suggest several hypotheses for the effects of BANK1 on several immune disorders and the interactions of *BANK1* alleles with cholesterol levels. Furthermore, we expect our strategy that focuses on nsSNPs to be an efficient way to discover functional associations from whole-genome data sets.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, N. & Stokes, J. 3rd Factors of risk in the development of coronary heart disease-six year follow-up experience. The Framingham Study. *Ann. Intern. Med.* **55**, 33–50 (1961).
2 Beekman, M., Heijmans, B. T., Martin, N. G., Pedersen, N. L., Whitfield, J. B. & DeFaire, U. *et al.* Heritabilities of apolipoprotein and lipid levels in three countries. *Twin Res.* **5**, 87–97 (2002).
3 Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M. & Koseki, M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
4 Kim, Y. J., Go, M. J., Hu, C., Hong, C. B., Kim, Y. K. & Lee, J. Y. *et al.* Large-scale genomewide association studies in east Asians identify new genetic loci influencing metabolic traits. *Nat. Genet.* **43**, 990–995 (2011).
5 Schork, N. J., Fallin, D. & Lanchbury, J. S. Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin. Genet.* **58**, 250–264 (2000).
6 Sunyaev, S., Ramensky, V. & Bork, P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet.* **16**, 591–597 (2000).
7 Wang, Z. & Moult, J. SNPs protein structure and disease. *Hum. Mutat.* **17**, 263–270 (2001).
8 Ferrer-Costa, C., Orozco, M. & de la Cruz, X. Characterization of sequence and structure properties. *J. Mol. Biol.* **315**, 771–786 (2002).
9 Cho, Y. S., Go, M. J., Kim, Y. J., Heo, J. Y., Oh, J. H. & Ban, H. J. *et al.* A large-scale genomewide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41**, 527–534 (2009).
10 Rabbee, N. & Speed, T. P. A genotype calling algorithm for affymetrix SNP arrarys. *Bioinformatics* **22**, 7–12 (2006).
11 Hong, K. W., Jin, H. S., Lim, J. E., Cho, Y. S., Go, M. J. & Jung, J. *et al.* Nonsynonymous single nucleotide polymorphisms associated with blood pressure and hypertension. *J. Hum. Hypertens.* **24**, 763–774 (2010).
12 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
13 Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K. & Nyholt, D. R. *et al.* Common SNPs explain a large proportion of heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
14 Hong, K. W., Lim, J. E., Kim, Y. J., Cho, N. H., Shin, C. & Oh, B. KARE genomewide association study of blood pressure using imputed SNPs. *Genomics Inform.* **8**, 103–107 (2010).
15 Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
16 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L.,L. & Ferreira, M. A. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2009).
17 Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
18 Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A. & Bork, P. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
19 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
20 Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
21 Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
22 Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P. & Junkins, H. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **22**, D1001–D1006 (2014).
23 Castillejo-Lopez, C., Delgado-Vega, A. M., Wojcik, J., Kozyrev, S. V., Thavathiru, E. & Wu, Y. Y. *et al.* Alarcon-Riquelme, genetic and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK. *Ann. Rheum. Dis.* **71**, 136–142 (2012).
24 Fan, T., Tao, J. H., Zhang, L. P., Li, L. H. & Ye, D. Q. The association between BANK1 and TNFAIP3 gene polymorphisms and systemic lupus erythematosus: a meta-analysis. *Int. J. Immunogenet.* **38**, 151–159 (2011).
25 Zhang, X., Fei, Z., Wan, J., Xu, J., Yu, B. & Guan, M. Association analysis of BANK1 gene with psoriasis in Southern Han Chinese. *Int. J. Immunogenet.* **38**, 507–512 (2011).
26 Dawidowicz, K., Dieude, P., Avouac, J., Wipff, J., Hachulla, E. & Diot, E. *et al.* Allanore, Association study of B-cell marker gene polymorphisms in European Caucasian patients with systemic sclerosis. *Clin. Exp. Rheumatol.* **29**, 839–842 (2011).

27 Coustet, B., Dieude, P., Guedj, M., Bouaziz, M., Avouac, J. & Ruiz, B. *et al.* C8orf13-BLK is a genetic risk locus for systemic sclerosis and has additive effects with BANK1: results from a large French cohort and meta-analysis. *Arthritis Rheum.* **63**, 2091–2096 (2011).

28 Olusi, S. O. & George, S. Prevalence of LDL atherogenic phenotype in patients with systemic lupus erythematosus. *Vasc. Health Risk Manag.* **7**, 75–80 (2011).

29 Kozyrev, S. V., Abelson, A. K., Wojcik, J., Zaghlool, A. & Linga Reddy, M. V. *et al.* Alarcon-Riquelme, Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 211–216 (2008).

30 Yokoyama, K., Su, I. H., Tezuka, T., Yasuda, T., Mikoshiba, K. & Tarakhovsky, A. *et al.* BANK regulates BCR-induced calcium mobilization by promoting tyrosine phosphorylation of IP3 receptor. *EMBO J.* **21**, 83–92 (2002).

31 Aiba, Y., Yamazaki, T., Okada, T., Gotoh, K., Sanjo, H. & Ogata, M. *et al.* BANK negatively regulates Akt activation and subsequent B cell responses. *Immunity* **24**, 259–268 (2006).

32 Steinberg, D. Atherogenesis in perspective: hypercholesterolemia and inflammation as partners in crime. *Nat. Med* **8**, 1211–1217 (2002).

33 Martin, S. S., Blumenthal, R. S. & Miller, M. LDL cholesterol: the lower the better. *Med. Clin.* **96**, 13–26 (2012).