

ORIGINAL ARTICLE

Genome-wide association study of serum lipids confirms previously reported associations as well as new associations of common SNPs within *PCSK7* gene with triglyceride

Makoto Kurano¹, Kazuhisa Tsukamoto², Shigeo Kamitsuji³, Naoyuki Kamatani³, Masumi Hara⁴, Toshio Ishikawa⁵, Bong-Jo Kim⁶, Sanghoon Moon⁶, Young Jin Kim⁶ and Tamio Teramoto⁷

Previous reports including genome-wide association studies (GWASs) have described associations of serum lipids with genomic variations. In the present study, we examined the association of ~2.5 million single-nucleotide polymorphisms (SNPs) from 3041 Japanese healthy volunteers obtained from the Japan Pharmacogenomics Data Science Consortium (JPDS) database with serum lipids. We confirmed the previously reported associations of 14 SNPs in 5 regions for low-density lipoprotein (LDL) cholesterol, 23 SNPs in 12 regions for high-density lipoprotein (HDL) cholesterol, 16 SNPs in 6 regions for triglyceride and 5 SNPs in 1 region for phospholipid. Furthermore, we identified 16 possible novel candidate genes associated with LDL cholesterol, HDL cholesterol or triglycerides, where SNPs had P -values of $< 1 \times 10^{-5}$. Further replication analyses of these genes with Korean data revealed significant associations of SNPs located within the *PCSK7* gene and triglyceride ($P_{\text{meta}} = 7.98 \times 10^{-9}$ and 1.91×10^{-8} for rs508487 and rs236911, respectively). These associations remained significant even by the conditional analysis adjusting for three neighboring variations associated with triglyceride. Our present data suggest that *PCSK7* as well as *PCSK9* may be associated with lipids, especially triglyceride, and may serve as a candidate for a new drug target to treat lipid abnormality syndromes.

Journal of Human Genetics (2016) 61, 427–433; doi:10.1038/jhg.2015.170; published online 14 January 2016

INTRODUCTION

Cardiovascular atherosclerotic diseases are the leading cause of death not only in developed countries but also in developing countries. Among the several risk factors for atherosclerotic diseases, dyslipidemia has been regarded as the most crucial one, and genetic factors as well as the environmental factors contribute to the variances of lipid traits.¹ For each component of dyslipidemia, monogenic mutations associated with Mendelian traits have been extensively pursued, and many mutations in genes that cause overt dyslipidemia have been found; for example, the deficiency of low-density lipoprotein (LDL) receptor in familial hypercholesterolemia² and the deficiency of cholesteryl ester transfer protein (CETP) in familial hyperalphalipoproteinemia.³

However, causes of most dyslipidemic cases are polygenic rather than monogenic. Therefore, association studies including genome-wide association studies (GWASs) have been performed to identify many loci associated with lipids traits.^{4–15} Subsequent experimental

approaches demonstrated the functional roles of some candidate genes, such as *GalNac transferase (GALNT2)*, *Sort1*,¹² and *Trib1*.¹⁶

Although many candidate genes for dyslipidemia have been proposed as a result of the above-mentioned studies, most of the GWASs were performed in western populations, and only a few studies^{14,17,18} have comprehensively examined genetic factors associated with lipid traits in East Asian populations. It is important to perform comprehensive studies to investigate the genetic factors for dyslipidemia in different populations because genetic backgrounds, lifestyle and lipid profiles are likely to be different between Western and East Asian populations. Actually, Kim *et al.*¹⁴ identified novel loci that had not been found in GWASs performed in western populations. Moreover, the Japanese population has been genetically isolated and the lipid traits of Japanese are special in that they are characterized by high concentrations of high-density lipoprotein cholesterol (HDL-C). Therefore, in order to enrich our knowledge about lipid-associated genes and to identify new drug targets to treat dyslipidemia, it is

¹Department of Clinical Laboratory Medicine, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; ²Department of Metabolism, Diabetes and Nephrology, Aizu Medical Center, Fukushima Medical University, Fukushima, Japan; ³StaGen, Statistical Genetics Analysis Division, Tokyo, Japan; ⁴Department of Medicine IV, Mizonokuchi Hospital, Teikyo University School of Medicine, Kawasaki, Japan; ⁵Department of Internal Medicine, Teikyo University School of Medicine, Tokyo, Japan; ⁶Division of Structural and Functional Genomics, Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do, Korea and ⁷Department of Internal Medicine, Teikyo University School of Medicine, Tokyo, Japan

Correspondence: Professor K Tsukamoto, Department of Metabolism, Diabetes and Nephrology, Aizu Medical Center, Fukushima Medical University, 21-2 Maeda, Tanisawa, Kawahigashi, Aizu-Wakamatsu City, Fukushima 969-3492, Japan.

E-mail: kazut@fmu.ac.jp

Received 28 August 2015; revised 23 November 2015; accepted 13 December 2015; published online 14 January 2016

important to replicate the previously reported genetic associations and to identify novel genetic factors utilizing GWAS in the Japanese population.

In the present study, we analyzed genome-wide data obtained from 3041 healthy Japanese volunteers in the Japan Pharmacogenomics Data Science Consortium (JPDSC) project¹⁹ to examine whether we could replicate previously described genetic associations with the four lipid traits, that is, LDL cholesterol (LDL-C), HDL-C, triglyceride and phospholipid. Afterwards, any associated single-nucleotide polymorphisms (SNPs) in previously unrecognized loci were also tested for associations in a Korean population.

MATERIALS AND METHODS

Database

All the data used in the present study were obtained from the JPDSC data set composed of 3041 healthy Japanese subjects. Detailed data about the data set are described elsewhere.¹⁹ The same database has been used for pharmacogenomic studies,^{20–22} genomic study for human leukocyte antigen²³ and a study for genetic factors affecting electrocardiographic parameters.²⁴

The database contains the genotypes of ~2.5 million SNPs from 2994 Japanese healthy volunteers. DNA and clinical information were obtained in two phases. The first set of samples was collected from 2000 to 2003 and the second set of samples was collected from 10 geographic regions in Japan. All the samples were collected from the Japanese population as validated by principal component analysis. Written informed consent was obtained from all the subjects, and this study was approved by the ethical committee of JPDSC.

Genome-wide association study

The final number of subjects included in the GWAS was 2994. Genomic DNAs were genotyped on an Illumina Human Omni 2.5.8 BeadChip (San Diego, CA, USA), and SNPs used in the GWAS were selected according to the following criteria: SNP call rates ($\geq 95\%$), the Hardy–Weinberg equilibrium test ($P \geq 0.01$) and minor allele frequency ($\geq 1\%$). Total number of the SNPs on the autosomes after quality control was 1 294 241. Lipid measures were adjusted for independent variables within each phase by the linear regression model. The clinical test data were log-transformed and standardized. The standardized residuals in each phase were combined and used as quantitative phenotypes for the association analysis, performed with PLINK (version 1.1.3).²⁵ The significance of the association between a trait and genotype was evaluated using the Wald test. Genome-wide significance was inferred at $P < 5 \times 10^{-8}$.

Replications of previously reported SNPs

Among the SNPs reported to be associated with LDL-C, HDL-C, triglyceride and phospholipid, the associations with P -values $< 5 \times 10^{-8}$ in previous reports were examined. As GWAS platforms were different between ours and those used in previous reports, some SNPs previously reported to be associated with lipid traits were not present in our data. In those cases, we searched for proxy SNPs with high r^2 values (not lower than 0.8) with the reported SNPs. When we attempted to replicate n different associations for a trait, the P -value threshold of $0.05/n$ was used according to the Bonferroni correction. The direction of each association was carefully examined to see whether it was the same in both the previous report and the present study. When a proxy SNP rather than the same SNP was used in the replication, it was carefully examined to see whether the directions of the effects are concordant between the previous study and our study.

Replications of novel SNPs

SNPs with P -values under the prespecified threshold of 1×10^{-5} were tested for replication using data sets obtained from the Korea Association Resource ($n=6805$) data set that was a part of the Korean Genome and Epidemiology Study cohort.¹⁴ Genomic DNAs were genotyped on an Affymetrix 5.0 SNP

array (Santa Clara, CA, USA). The covariates used for the correction in the replication study were the same between our data and the Korean data.

As we attempted to replicate 15 different SNPs except for those within *PCSK7*, the P -value threshold of $0.05/15 = 0.0033$ was used according to the Bonferroni correction. For 6 SNPs within *PCSK7*, $0.05/6 = 0.0083$ was used. The direction of each association was carefully examined to see whether the direction of the association in the previous report was the same as that in the present study. Meta-analysis was performed using the inverse variance method.

Statistical analysis

Association test and quality control were performed with PLINK (version 1.1.3), and all the other statistical analyses were performed using R environment (version 2.15.0; Vienna, Austria).

RESULTS

Lipid phenotypes in JPDSC data

Metabolic trait data and other data were obtained from 3041 healthy subjects.¹⁹ The data were obtained in two different phases. Phase 1 data were obtained from 1032 subjects and phase 2 data were obtained from 2009 subjects. Supplementary Tables 1 and 2 show the summarized data separately for males and females. As shown in these tables, LDL-C and HDL-C data were obtained only from phase 2 (1001 males and 1006 females), whereas the other data including total cholesterol, triglyceride and phospholipid levels were obtained from both phase 1 and phase 2. Examination of the patterns in histograms for age, body mass index (BMI), LDL-C, HDL-C, total cholesterol, triglyceride, phospholipid, systolic blood pressure and diastolic blood pressure indicated that log-transformed rather than raw values generally fit closer to normal distributions (Supplementary Figures 1A–K). In addition, two-dimensional plots for each of the traits against BMI indicated that many of the pairs of variables are significantly correlated with each other, and log-transformed rather than raw values generally showed linear relationships (Supplementary Figures 2A–H and Supplementary Table 3). We analyzed extensively the relationships between two different variables and noticed that, in general, each pair of variables correlated with each other and log-transformed rather than raw values showed linear relationships (Supplementary Figures 3A–D indicate representative results). Therefore, log-transformed values were used for the analyses in most of the traits in this study.

Characterization of the phenotype data

We extensively analyzed the data to determine appropriate conditions for the association study. Multivariate regression analysis using age and BMI as independent variables and one of the clinical laboratory data, or systolic or diastolic blood pressure as a dependent variable, showed that BMI had positive effects on LDL-C, total cholesterol, triglyceride, phospholipid, urate, HbA1c, systolic pressure and diastolic pressure in both males and females (Supplementary Table 4). Note that all the variables and values were log-transformed before use. For HDL-C, BMI had significant negative effects in both males and females, whereas it had a positive effect on urate only in males (Supplementary Table 4). Age had positive effects on all the traits described in Supplementary Table 4 except for HDL-C and urate in males, and phospholipid in females.

Replication of previous associations

Clinical test data were log-transformed and standardized as described in the Materials and Methods. The covariates used for the adjustment included alcohol ingestion, smoking, log-transformed BMI, gender, log-transformed albumin, log-transformed HbA1c,

log-transformed alanine transaminase, log-transformed aspartate transaminase, log-transformed γ -glutamyl transpeptidase, log-transformed creatinine and eigenvalues of the top three components from the principle component analysis using Eigenstrat.²⁶ Using the results of the analysis, we examined whether previously described SNPs associated with LDL-C, HDL-C, triglyceride or phospholipid are also associated in individuals of Japanese ancestry (Table 1).

Low-density lipoprotein cholesterol. Manhattan plots and QQ plots for LDL-C are shown in Supplementary Figures 4A and B. We attempted to examine 95 SNPs reported to be associated with LDL-C selected from the GWAS catalog. When the SNPs reported in the previous studies were included in our platform, we used those SNPs; however, SNPs that were in tight linkage disequilibrium with the reported SNPs were used otherwise as proxy SNPs. Among the 95 SNPs examined, 57 SNPs in our platform were found to be either the same SNPs or the proxy SNPs. Applying the Bonferroni correction method for multiple comparisons, the corrected threshold of $0.05/57 = 8.7 \times 10^{-4}$ was applied. Among the 57 SNPs, 14 showed nominal *P*-values under 0.05 (Table 1) that were mapped to 6 regions (*CELSR2-PSRC1-SORT1*, *APOB*, *HMGCR*, *TIMD4-HAVCR1*, *HP-HPR-DHX38* and *APOE-APOC1-APOC2*). In all the 14 SNPs, the directions of the associations were concordant between the previous reports and the present report. Of the 14 SNPs, 12 SNPs were significant even after applying Bonferroni correction (Table 1). Although previously reported to be associated with LDL-C located in the 3' direction of *PCSK9*,²⁷ rs72669744 was not significantly associated with LDL-C in the present study, and rs17111474 located in the 5' direction of the same gene showed a rather low *P*-value (4.24×10^{-4}).

High-density lipoprotein cholesterol. Manhattan plots and QQ plots for HDL-C are shown in Supplementary Figures 4C and D. We attempted to examine 119 SNPs selected from the GWAS catalog reported to be associated with HDL-C. Among the 119 SNPs examined, 71 SNPs in our platform were found to be either the same SNPs or the proxy SNPs. Applying the Bonferroni correction method for multiple comparisons, the corrected threshold of $0.05/71 = 7.0 \times 10^{-4}$ was applied. Among the 71 SNPs, 24 showed nominal *P*-values under 0.05 (Table 1) that were mapped to 13 regions (*MACF1-PABPC4*, *ZNF648*, *GALNT2*, *LPL*, *ABCA1*, *FADS1-FADS2-FADS3*, *MMAB-MVK*, *SCARB1*, *LIPC*, *CETP*, *CT-CHOLF-PRMT8*, *LCAT* and *HNF4A*). In 23 SNPs out of 24, the directions of the associations were concordant between the previous reports and the present report. Of the 23 SNPs, 8 SNPs were significant even after applying the Bonferroni correction (Table 1).

Triglyceride. Manhattan plots and QQ plots for triglyceride are shown in Supplementary Figures 5A and B. We attempted to examine 99 SNPs from the GWAS catalog reported to be associated with triglyceride. Among the 99 SNPs examined, 60 SNPs in our platform were found to be the same SNPs or the proxy SNPs. Applying the Bonferroni correction method for multiple comparisons, the corrected threshold of $0.05/60 = 8.3 \times 10^{-4}$ was applied. Among the 60 SNPs, 19 showed nominal *P*-values under 0.05 (Table 1) that were mapped to 8 regions (*ANGPTL3*, *GCKR*, *RPL26P19-HAVCR1*, *TBL2-MLXIPL*, *LPL*, *XKR6-AMAC1L2*, *TRIB1* and *APOA5-APOA4-APOC3-APOA1*). In 16 SNPs out of 19, the directions of the associations were concordant between the previous reports and the present report. Of the 16 SNPs, 12 SNPs were significant even after applying the Bonferroni correction (Table 1).

Phospholipid. Manhattan plots and QQ plots for phospholipid are shown in Supplementary Figures 5C and D. We attempted to examine 63 SNPs from the GWAS catalog reported to be associated with phospholipid. Among the 63 SNPs examined, 36 SNPs in our platform were found to be either the same SNPs or the proxy SNPs. Applying the Bonferroni correction method for multiple comparisons, the corrected threshold of $0.05/36 = 1.38 \times 10^{-3}$ was applied. Among the 36 SNPs, 5 showed nominal *P*-values under 0.05 (Table 1) that were mapped to the same 11q12.2 region (*MYRF-TIME258-FEN1-FADS1*). In all the five SNPs, the directions of the associations were concordant between the previous reports and the present report. Of the five SNPs, none remained significant after applying the Bonferroni correction (Table 1).

Possible novel loci associated with the lipid traits

Using a prespecified *P*-value threshold of 1×10^{-5} , we searched for novel genetic loci harboring significant SNPs associated with LDL-C, HDL-C and triglyceride using Korean data. Except for the association between *PCSK7* locus and triglyceride, we selected each one SNP for a locus; that is, four SNPs for LDL-C, seven SNPs for HDL-C and four SNPs for triglyceride. For the replication of the associations between *PCSK7* and triglyceride, six SNPs were submitted as this gene was of interest. Although a rare variant within the *PCSK7* gene (rs142953140) has been reported to be associated with HDL-C, LDL-C and triglyceride by the data from exome sequencing,²⁸ the frequency of that variation was only 0.2%, and no common SNP has been reported to be associated with SNPs at this locus. All the SNPs except for *PCSK7* failed to be replicated (Supplementary Table 5). However, the associations between *PCSK7* and triglyceride were successfully replicated by the Korean data (Table 2); three out of the six SNPs showed significant *P*-values (< 0.0083) against the threshold considering the multiple comparisons as described in the Materials and Methods. Furthermore, in two of the SNPs (rs508487 and rs236911), the *P*-values of the meta-analysis were lower than the threshold of 5×10^{-8} used as the criteria for genome-wide significance (Table 2).

DISCUSSION

A genome-wide association approach is an established methodology that enables the identification of genetic loci associated with complex traits and diseases unconstrained by prior knowledge. In the fields of lipidology, series of elegant studies utilizing GWAS have successfully identified many lipid-associated genetic variants,^{4–15} although these genetic variants account for approximately only one-third of the genetic variance of lipid traits.²⁹ Furthermore, most of the previous GWASs were performed in western populations, and it is important to conduct GWAS in non-western populations to examine whether the identified associations are shared by different ethnic groups and to find novel loci associated with the lipid traits.

In the present study, we examined the associations of ~ 2.5 million SNPs with serum lipid concentrations using the data from 2994 Japanese healthy volunteers obtained from the JPDSC database. We first examined whether the associations of SNPs with LDL-C, HDL-C, triglyceride and phospholipid observed among individuals of European, African and Asian ancestries are replicated in our data (Table 1). The results have shown that some of the associations were replicated in our data. Furthermore, for all the replicated associations that were significant even after Bonferroni correction, the directions were concordant between GWAS catalog and our database. These results indicate that the associations in GWAS catalog are reliable, and that each of these SNPs that showed concordance after Bonferroni correction possesses universal significance on the lipids. Many of such

Table 1 Replication analyses of previously reported genes associated with lipid traits in GWAS catalog utilizing JPDS data

Trait	Region	Reported gene(s)	Mapped gene	SNPs	Risk allele	β	P-value	Proxy ^a	Minor	Major	Coefficient	s.e.	P-value	Bonferroni ^b	Sig ^c
LDL	1p13.3	CELSR2	CELSR2	rs660240	A	-0.04	1.00E-26		A	G	-0.25	0.059	4.56E-05	T	-/-
LDL	1p13.3	CELSR2	CELSR2	rs646776	G	-0.16	2.00E-12		G	A	-0.24	0.059	4.93E-05	T	-/-
LDL	1p13.3	CELSR2, PSRC1, SORT1	CELSR2	rs12740374	T	-0.23	2.00E-42	rs629301	C	A	-0.25	0.061	3.22E-05	T	-/-
LDL	1p13.3	CELSR2, PSRC1	PSRC1	rs599839	G	-0.16	1.00E-33		G	A	-0.23	0.059	7.86E-05	T	-/-
LDL	1p13.3	CELSR2, PSRC1, SORT1	CELSR2	rs629301	G	-5.65	1.00E-170		C	A	-0.25	0.061	3.22E-05	T	-/-
LDL	2p24.1	APOB	APOB	rs1367117	A	4.05	4.00E-114		A	G	0.13	0.061	3.19E-02	T	+/+
LDL	5q13.3	HMGCR	HMGCR	rs12916	C	2.45	5.00E-45		A	G	-0.12	0.031	1.19E-04	T	+/+
LDL	5q13.3	HMGCR	HMGCR	rs3846662	G	0.08	2.00E-11		A	G	-0.12	0.031	1.10E-04	T	+/+
LDL	5q13.3	HMGCR	HMGCR	rs3846663	T	0.07	8.00E-12	rs12916	A	G	-0.12	0.031	1.19E-04	T	-/-
LDL	5q33.3	TIMD4, HAVCR1	TIMD4, HAVCR1	rs1501908	G	-0.07	1.00E-11	rs58198139	G	A	-0.15	0.043	2.25E-04	T	-/-
LDL	5q33.3	TIMD4, HAVCR1	TIMD4	rs6882076	T	-1.67	2.00E-22		A	G	-0.15	0.041	2.86E-04	T	-/-
LDL	16q22.2	HP, HPR, DHX38	HPR	rs2000999	A	2	2.00E-22		A	G	0.08	0.031	9.65E-03	T	+/+
LDL	19q13.32	APOE	TOMM40	rs1160985	C	6.77	2.00E-21		A	G	-0.16	0.033	2.69E-06	T	+/+
LDL	19q13.32	APOE, APOC1, APOC2	APOC1	rs4420638	G	7.14	9.00E-147		G	A	0.23	0.054	1.39E-05	T	+/+
HDL	1p34.3	MACF1, PABPC4	PABPC4	rs4660293	G	-0.48	4.00E-10		G	A	-0.08	0.043	4.24E-02	T	-/-
HDL	1q25.3	ZNF648	LOC100130996	rs1689800	G	-0.47	3.00E-10	rs1779824	A	G	-0.08	0.036	2.45E-02	T	-/-
HDL	1q42.13	GALNT2	GALNT2	rs4846914	G	-0.07	2.00E-13	rs2144300	A	G	0.1	0.041	1.69E-02	T	-/-
HDL	8p21.3	LPL	LPL	rs326	T	0.02	1.00E-08	rs327	C	A	0.15	0.041	2.58E-04	T	+/+
HDL	8p21.3	LPL	LPL	rs325	T	-0.05	8.00E-26	rs328	G	C	0.24	0.048	6.47E-07	T	-/-
HDL	8p21.3	LPL	LPL-RPL30P9	rs17482753	T	2.02	3.00E-11		A	C	0.23	0.048	1.33E-06	T	+/+
HDL	8p21.3	LPL	LPL	rs328	G	0.17	9.00E-23		G	C	0.24	0.048	6.47E-07	T	+/+
HDL	9q31.1	ABCA1	ABCA1	rs12686004	T	-0.02	2.00E-18		A	G	-0.08	0.038	3.05E-02	T	-/-
HDL	9q31.1	ABCA1	ABCA1	rs1883025	T	-0.94	2.00E-33		A	G	-0.1	0.036	5.33E-03	T	-/-
HDL	11q12.2	FADS1	FADS1	rs174548	G	-0.01	1.00E-12		C	G	-0.09	0.033	6.81E-03	T	-/-
HDL	11q12.2	FADS1, FADS2, FADS3	FADS1	rs174547	C	-0.09	2.00E-12		G	A	-0.09	0.033	6.86E-03	T	-/-
HDL	11q12.2	FADS1, FADS2, FADS3	FADS1	rs174546	T	-0.73	2.00E-22		G	A	-0.08	0.033	9.05E-03	T	-/-
HDL	12q24.11	MMAB, MVK	MMAB	rs7134594	C	-0.44	7.00E-15		A	G	0.08	0.033	1.38E-02	T	-/-
HDL	12q24.31	SCARB1	NCOR2-SCARB1	rs838880	C	0.61	3.00E-14		A	G	-0.1	0.031	1.10E-03	T	+/+
HDL	15q21.3	LIPC	LIPC	rs1077835	A	-0.17	2.00E-14	rs1800588	A	G	0.12	0.033	2.14E-04	T	-/-
HDL	15q21.3	LIPC	RPL28P4-LIPC	rs1532085	A	1.45	3.00E-96		G	A	-0.11	0.031	7.30E-04	T	+/+
HDL	15q21.3	LIPC	LIPC	rs1800588	T	0.14	2.00E-32		A	G	0.12	0.033	2.14E-04	T	+/+
HDL	16q13	CETP	CETP	rs12708980	C	-0.04	2.00E-28		C	A	-0.17	0.061	4.79E-03	T	-/-
HDL	16q13	CETP	HERPUD1-CETP	rs9989419	G	0.04	1.00E-32		A	G	-0.18	0.036	1.12E-06	T	+/+
HDL	16q13	CETP	HERPUD1-CETP	rs3764261	A	3.39	7.00E-380		A	C	0.32	0.038	1.18E-16	T	+/+
HDL	16q13	CETP, NUP93, SLIC12A3, HERPUD1	CETP	rs1800775	A	3.09	4.00E-93		C	A	-0.08	0.031	1.01E-02	T	+/+
HDL	16q22.1	CTCF, PRMT8	NUTF2	rs2271293	G	-0.13	8.00E-16		A	G	0.16	0.082	4.76E-02	T	-/-
HDL	16q22.1	LOAT	PSKH1	rs16942887	A	1.27	8.00E-33		A	G	0.16	0.082	4.98E-02	T	+/+
HDL	20q13.12	HNF4A	HNF4A	rs1800961	T	-0.19	8.00E-10		A	G	-0.42	0.156	6.83E-03	T	-/-
TG	1p31.3	ANGPTL3	DOCK7	rs1748195	C	7.12	2.00E-10		G	C	-0.1	0.032	0.001273	T	+/+
TG	2p23.3	GCKR	GCKR	rs780092	G	-0.05	5.00E-27		G	A	-0.07	0.029	0.01145	T	-/-
TG	2p23.3	GCKR	GCKR	rs1260326	T	8.76	6.00E-133		G	A	-0.09	0.026	0.00072	T	+/+
TG	2p23.3	GCKR	GCKR	rs780094	T	8.59	6.00E-32		G	A	-0.08	0.026	0.003526	T	+/+

Table 1 (Continued)

Trait	Region	Reported gene(s)	GWAS catalog										Replication			
			Mapped gene	SNPs	Risk allele	β	P-value	Proxy ^a	Minor	Major	Coefficient	s.e.	P-value	Bonferroni ^b	Signif ^c	
TG	5q33.3	RPL26P19, HAVCR1	TIMD4	rs6882076	G	-2.63	4.00E-12		A	G	-0.1	0.034	0.002649		-/-	
TG	7q11.23	MLXIPL	TBL2	rs17145738	G	-7.91	9.00E-59		A	G	-0.14	0.041	0.000647	T	-/-	
TG	7q11.23	TBL2, MLXIPL	TBL2	rs2286276	T	-0.07	1.00E-15		A	G	-0.15	0.042	0.000277	T	-/-	
TG	8p21.3	LPL	LPL	rs326	T	0.04	1.00E-08	rs327	C	A	-0.19	0.034	2.81E-08	T	+/+	
TG	8p21.3	LPL	LPL-RPL30P9	rs17410962	G	-0.11	7.00E-09	rs17410996	C	A	-0.23	0.04	1.75E-08	T	-/-	
TG	8p21.3	LPL	LPL-RPL30P9	rs10503669	A	-0.09	7.00E-39	rs17410996	C	A	-0.23	0.04	1.75E-08	T	-/-	
TG	8p21.3	LPL	LPL-RPL30P9	rs10105606	C	0.07	4.00E-26		A	C	-0.17	0.034	2.24E-07	T	+/+	
TG	8p21.3	LPL	LPL-RPL30P9	rs12678919	G	-13.64	2.00E-115	rs7841189	A	G	-0.23	0.04	1.75E-08	T	-/-	
TG	8p21.3	LPL	LPL-RPL30P9	rs10096633	G	0.17	2.00E-18	rs17482753	A	C	-0.22	0.04	2.26E-08	T	+/+	
TG	8p21.3	LPL	LPL	rs328	G	-0.19	2.00E-28		G	C	-0.22	0.04	4.47E-08	T	-/-	
TG	8p23.1	XKR6, AMACIL2	XKR6	rs7819412	G	-0.04	3.00E-08		A	G	-0.12	0.049	0.01766		-/+	
TG	8q24.13	TRIB1	TRIB1-LINC00861	rs17321515	G	-0.08	4.00E-17		G	A	-0.06	0.026	0.0137		-/-	
TG	11q23.3	APOA1, KIAA0999, LOC645044	SIK3	rs2075292	G	8.7	5.00E-08		A	C	-0.13	0.026	8.80E-07	T	+/+	
TG	11q23.3	APOA5	APOA5	rs651821	T		9.00E-26		G	A	0.32	0.027	1.01E-30	T	+/+	
TG	11q23.3	BUD13, ZNF259, APOA5, APOA4, APOC3, APOA1	RPL15P15, BUD13	rs4938303	T	-0.07	4.00E-21		G	A	0.07	0.026	0.01347		-/-	
PL	11q12.2	FADS1	FADS1	rs174550	T	0.08	1.00E-57		G	A	-0.06	0.026	0.02514		+/+	
PL	11q12.2	FADS1	FADS1	rs174547	T	0.08	4.00E-154		G	A	-0.06	0.026	0.02514		+/+	
PL	11q12.2	FEN1	FEN1	rs4246215	T	-0.07	1.00E-139	rs174546	A	G	-0.06	0.027	0.02521		-/-	
PL	11q12.2	C11orf9	MYRF	rs174535	T	0.07	1.00E-151		G	A	-0.06	0.026	0.02464		+/+	
PL	11q12.2	C11orf10	TMEM258;FEN1; MIR611	rs174538	A	0.08	5.00E-58		A	G	-0.06	0.027	0.03331		+/+	

Abbreviations: GWAS, genome-wide association study; HDL, high-density lipoprotein; JPDSC, Japan Pharmacogenomics Data Science Consortium; LDL, low-density lipoprotein; PL, phospholipid; SNP, single-nucleotide polymorphism; TG, triglyceride.
^aWhen the original SNP was not included in our platform, a proxy SNP with a high-grade linkage disequilibrium was used.
^bThe association judged as significant after Bonferroni correction is marked as 'T'.
^cThe signs of the coefficient that represents the tendency of association between GWAS catalog and our study.

Table 2 Associations of SNPs within *PCSK7* gene in Japanese and Korean samples

SNPs	CHR	Gene	Japanese					Korean				P (meta-analysis) ^a	
			MAF	Minor	β^b	s.e.	P	MAF	β^b	s.e.	P	Association	Heterogeneity
rs508487	11	<i>PCSK7</i>	0.102	T	0.070	0.022	1.45E-03	0.116	0.055	0.011	1.28E-06	7.98E-09	5.31E-01
rs236911	11	<i>PCSK7</i>	0.47	C	0.051	0.013	1.12E-04	0.458	0.030	0.007	1.55E-05	1.91E-08	1.53E-01
rs2239011	11	<i>PCSK7</i>	0.209	G	-0.055	0.016	5.52E-04	0.232	-0.021	0.008	1.05E-02	1.15E-04	5.53E-02
rs3830008	11	<i>PCSK7</i>	0.209	C	-0.054	0.016	6.22E-04	0.233	-0.019	0.008	1.80E-02	2.43E-04	4.71E-02
rs2277287	11	<i>PCSK7</i>	0.205	A	-0.057	0.016	4.21E-04	0.235	-0.019	0.008	1.58E-02	1.79E-04	3.85E-02
rs522645	11	<i>PCSK7</i>	0.349	A	0.068	0.014	9.26E-07	0.297	0.021	0.008	7.37E-03	1.73E-06	3.50E-03

Abbreviations: CHR, chromosome; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

^aP-value obtained by meta-analysis using inverse variance method.

^bRegression coefficient.

genes including *APOB*, *CETP*, and *LPL* have attracted the attention of lipid researchers.

Among possible novel genetic loci associated with LDL-C, HDL-C or triglyceride, association of the SNPs within *PCSK7* and triglyceride were replicated in the Korean data. The present study provided the first evidence that common SNPs within *PCSK7* are associated with a lipid phenotype. In a previous study in which exome sequencing rather than SNP arrays was used for association studies, a rare allele within *PCSK7* with a minor allele frequency of 0.2% had a large effect on serum HDL-C.²⁸ Although several genes including *APOA1/C3/A4/A5* that are known to be involved in the metabolism of HDL-C and triglyceride are located near the *PCSK7* gene, rs142953140 in *PCSK7* was shown to be associated with HDL-C by conditional analysis adjusting for six known variants in the *APOA1/C3/A4/A5* region associated with the lipid.²⁸ In our data, rs522645 remained to be associated with triglyceride ($P=0.0438$) even after adjustment for rs1942478 (within *ZPR1*), rs651821 (within *APOA5*) and rs734104 (within *APOC3*) that are located near *PCSK7* and showed low *P*-values in our data.

PCSK7 (PC7) is a member of the subtilisin-like proprotein convertase family that processes multiple protein precursors. *PCSK9*, a member of the same family, is known to cause autosomal-dominant hypercholesterolemia^{30,31} and hypocholesterolemia.³² Although *PCSK7* is expressed in the colon and lymphoid-associated tissues, it is also expressed to some degree in the liver and the intestine, organs that are largely involved in lipid metabolism.³³ Scamuffa *et al.*³⁴ reported that *PCSK7* knockout mice had no apparent abnormal phenotype, whereas Wetsel *et al.*³⁵ reported that they had impaired cognitive performance. Adeno-associated virus-mediated gene expression of *PCSK7* in mice liver was reported to have affected plasma lipid levels.²⁸ In spite of these elegant studies, the lipid metabolism in mice is quite different from that in humans. For example, unlike humans, mice have *APOBEC* that edits apoB100 mRNA and makes apoB48 mRNA in both liver and intestine. They also lack *CETP* that greatly affects HDL-C levels. Further basic and experimental approaches are necessary to evaluate the importance of *PCSK7* for the modulation of lipid metabolism in humans. In humans, *PCSK7* has been reported to be associated with soluble transferrin receptor level.³⁶ They analyzed whether soluble transferrin receptor level was associated with lipid parameters but found no associations. The mechanism of the association between *PCSK7* and triglyceride in humans remains to be clarified.

In summary, utilizing the JPDS data set, we replicated the previously reported associations of 14 SNPs in 5 regions for LDL-C, 23 SNPs in 12 regions for HDL-C, 16 SNPs in 6 regions for

triglyceride and 5 SNPs in 1 region for phospholipid, and identified 16 possible novel loci associated with the lipid traits. We further replicated the association between *PCSK7* and triglyceride with the Korean data. *PCSK7* may be associated with lipids, especially triglyceride, and may serve as a candidate for a new drug target to treat dyslipidemia.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by a Health and Labor Sciences Research Grant, Japan (Comprehensive Research on Lifestyle-Related Diseases Including Cardiovascular Diseases and Diabetes Mellitus) from the Ministry of Health, Labor, and Welfare of Japan (to TT) and by research grants from Taisho Toyama Pharmaceutical and DAIICHI SANKYO (to KT). This research was also supported by an intramural grant from the Korea National Institute of Health (2014-NI73001-00), and the Korea Genome Analysis Project (4845-301) and the Korean Genome and Epidemiology Study (4845-302) provided data.

- 1 Pilia, G., Chen, W. M., Scuteri, A., Orru, M., Albai, G., Dei, M. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* **2**, e132 (2006).
- 2 Brown, M. S. & Goldstein, J. L. Expression of the familial hypercholesterolemia gene in heterozygotes: mechanism for a dominant disorder in man. *Science* **185**, 61–63 (1974).
- 3 Inazu, A., Jiang, X. C., Haraki, T., Yagi, K., Kamon, N., Koizumi, J. *et al.* Genetic cholesteryl ester transfer protein deficiency caused by two prevalent mutations as a major determinant of increased levels of high density lipoprotein cholesterol. *J. Clin. Invest.* **94**, 1872–1882 (1994).
- 4 Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**, 161–169 (2008).
- 5 Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burt, N. P., Rieder, M. J. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40**, 189–197 (2008).
- 6 Sandhu, M. S., Waterworth, D. M., Debenham, S. L., Wheeler, E., Papadakis, K., Zhao, J. H. *et al.* LDL-cholesterol concentrations: a genome-wide association study. *Lancet* **371**, 483–491 (2008).
- 7 Chasman, D. I., Pare, G., Zee, R. Y., Parker, A. N., Cook, N. R., Buring, J. E. *et al.* Genetic loci associated with plasma concentration of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and Apolipoprotein B among 6382 white women in genome-wide analysis with replication. *Circ. Cardiovasc. Genet.* **1**, 21–30 (2008).
- 8 Aulchenko, Y. S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I. M., Pramstaller, P. P. *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat. Genet.* **41**, 47–55 (2009).
- 9 Hicks, A. A., Pramstaller, P. P., Johansson, A., Vitart, V., Rudan, I., Ugocsai, P. *et al.* Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet.* **5**, e1000672 (2009).
- 10 Chasman, D. I., Pare, G., Mora, S., Hopewell, J. C., Peloso, G., Clarke, R. *et al.* Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet.* **5**, e1000730 (2009).

- 11 Kathiresan, S., Willer, C. J., Peloso, G. M., Demissie, S., Musunuru, K., Schadt, E. E. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**, 56–65 (2009).
- 12 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).
- 13 Lemaitre, R. N., Tanaka, T., Tang, W., Manichaikul, A., Foy, M., Kabagambe, E. K. *et al.* Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet.* **7**, e1002193 (2011).
- 14 Kim, Y. J., Go, M. J., Hu, C., Hong, C. B., Kim, Y. K., Lee, J. Y. *et al.* Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat. Genet.* **43**, 990–995 (2011).
- 15 Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
- 16 Burkhardt, R., Toh, S. A., Lagor, W. R., Birkeland, A., Levin, M., Li, X. *et al.* Trib1 is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. *J. Clin. Invest.* **120**, 4410–4414 (2010).
- 17 Braun, T. R., Been, L. F., Singhal, A., Worsham, J., Ralhan, S., Wander, G. S. *et al.* A replication study of GWAS-derived lipid genes in Asian Indians: the chromosomal region 11q23.3 harbors loci contributing to triglycerides. *PLoS ONE*, **7**, e37056 (2012).
- 18 Chung, S. K., Yu, H., Park, A. Y., Kim, J. Y. & Cha, S. Genetic loci associated with changes in lipid levels leading to constitution-based discrepancy in Koreans. *BMC Complement. Altern. Med.* **14**, 230 (2014).
- 19 Kamitsuji, S., Matsuda, T., Nishimura, K., Endo, S., Wada, C., Watanabe, K. *et al.* Japan PGx Data Science Consortium Database: SNPs and HLA genotype data from 2994 Japanese healthy individuals for pharmacogenomics studies. *J. Hum. Genet.* **60**, 319–326 (2015).
- 20 Tohkin, M., Kaniwa, N., Saito, Y., Sugiyama, E., Kurose, K., Nishikawa, J. *et al.* A whole-genome association study of major determinants for allopurinol-related Stevens-Johnson syndrome and toxic epidermal necrolysis in Japanese patients. *Pharmacogenomics J.* **13**, 60–69 (2013).
- 21 Kaniwa, N., Sugiyama, E., Saito, Y., Kurose, K., Maekawa, K., Hasegawa, R. *et al.* Specific HLA types are associated with antiepileptic drug-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in Japanese subjects. *Pharmacogenomics* **14**, 1821–1831 (2013).
- 22 Chung, W. H., Chang, W. C., Lee, Y. S., Wu, Y. Y., Yang, C. H., Ho, H. C. *et al.* Genetic variants associated with phenytoin-related severe cutaneous adverse reactions. *JAMA* **312**, 525–534 (2014).
- 23 Khor, S. S., Yang, W., Kawashima, M., Kamitsuji, S., Zheng, X., Nishida, N. *et al.* High-accuracy imputation for HLA class I and II genes based on high-resolution SNP data of population-specific references. *Pharmacogenomics J.* **15**, 530–537 (2015).
- 24 Sano, M., Kamitsuji, S., Kamatani, N., Hong, K. W., Han, B. G., Kim, Y. *et al.* Genome-wide association study of electrocardiographic parameters identifies a new association for PR interval and confirms previously reported associations. *Hum. Mol. Genet.* **23**, 6668–6676 (2014).
- 25 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 26 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
- 27 Kettunen, J., Tukiainen, T., Sarin, A. P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L. P. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat. Genet.* **44**, 269–276 (2012).
- 28 Peloso, G. M., Auer, P. L., Bis, J. C., Voorman, A., Morrison, A. C., Stitzel, N. O. *et al.* Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am. J. Hum. Genet.* **94**, 223–232 (2014).
- 29 Willer, C. J. & Mohlke, K. L. Finding genes and variants for lipid levels after genome-wide association analysis. *Curr. Opin. Lipidol.* **23**, 98–103 (2012).
- 30 Abifadel, M., Varret, M., Rabes, J. P., Allard, D., Ouguerram, K., Devillers, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).
- 31 Timms, K. M., Wagner, S., Samuels, M. E., Forbey, K., Goldfine, H., Jammulapati, S. *et al.* A mutation in PCSK9 causing autosomal-dominant hypercholesterolemia in a Utah pedigree. *Hum. Genet.* **114**, 349–353 (2004).
- 32 Shioji, K., Mannami, T., Kokubo, Y., Inamoto, N., Takagi, S., Goto, Y. *et al.* Genetic variants in PCSK9 affect the cholesterol level in Japanese. *J. Hum. Genet.* **49**, 109–114 (2004).
- 33 Seidah, N. G., Hamelin, J., Mamarbachi, M., Dong, W., Tardos, H., Mbikay, M. *et al.* cDNA structure, tissue distribution, and chromosomal localization of rat PC7, a novel mammalian proprotein convertase closest to yeast kexin-like proteinases. *Proc. Natl Acad. Sci. USA* **93**, 3388–3393 (1996).
- 34 Scamuffa, N., Calvo, F., Chretien, M., Seidah, N. G. & Khatib, A. M. Proprotein convertases: lessons from knockouts. *FASEB J.* **20**, 1954–1963 (2006).
- 35 Wetsel, W. C., Rodriguez, R. M., Guillemot, J., Rousselet, E., Essalmani, R., Kim, I. H. *et al.* Disruption of the expression of the proprotein convertase PC7 reduces BDNF production and affects learning and memory in mice. *Proc. Natl Acad. Sci. USA* **110**, 17362–17367 (2013).
- 36 Oexle, K., Ried, J. S., Hicks, A. A., Tanaka, T., Hayward, C., Bruegel, M. *et al.* Novel association to the proprotein convertase PCSK7 gene locus revealed by analysing soluble transferrin receptor (sTfR) levels. *Hum. Mol. Genet.* **20**, 1042–1047 (2011).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)