



Unveiling Genetic Variants Underlying Vitamin D Deficiency in Multiple Korean Cohorts by a Genome-Wide Association Study

Ye An Kim^{1,*}, Ji Won Yoon^{2,*}, Young Lee³, Hyuk Jin Choi², Jae Won Yun³, Eunsin Bae³, Seung-Hyun Kwon³, So Eun Ahn⁴, Ah-Ra Do⁴, Heejin Jin⁵, Sungho Won^{4,5,6}, Do Joon Park⁷, Chan Soo Shin⁷, Je Hyun Seo³

¹Division of Endocrinology, Department of Internal Medicine, Veterans Health Service Medical Center; ²Healthcare System Gangnam Center, Seoul National University Hospital; ³Veterans Medical Research Institute, Veterans Health Service Medical Center; ⁴Department of Public Health Science, ⁵Institute of Health and Environment, Seoul National University; ⁶RexSoft, Inc.; ⁷Department of Internal Medicine, Seoul National University, Seoul, Korea

Background: Epidemiological data have shown that vitamin D deficiency is highly prevalent in Korea. Genetic factors influencing vitamin D deficiency in humans have been studied in Europe but are less known in East Asian countries, including Korea. We aimed to investigate the genetic factors related to vitamin D levels in Korean people using a genome-wide association study (GWAS).

Methods: We included 12,642 subjects from three different genetic cohorts consisting of Korean participants. The GWAS was performed on 7,590 individuals using linear or logistic regression meta- and mega-analyses. After identifying significant single nucleotide polymorphisms (SNPs), we calculated heritability and performed replication and rare variant analyses. In addition, expression quantitative trait locus (eQTL) analysis for significant SNPs was performed.

Results: rs12803256, in the actin epsilon 1, pseudogene (*ACTE1P*) gene, was identified as a novel polymorphism associated with vitamin D deficiency. SNPs, such as rs11723621 and rs7041, in the group-specific component gene (*GC*) and rs11023332 in the phosphodiesterase 3B (*PDE3B*) gene were significantly associated with vitamin D deficiency in both meta- and mega-analyses. The SNP heritability of the vitamin D concentration was estimated to be 7.23%. eQTL analysis for rs12803256 for the genes related to vitamin D metabolism, including glutamine-dependent NAD(+) synthetase (*NADSYN1*) and 7-dehydrocholesterol reductase (*DHCR7*), showed significantly different expression according to alleles.

Conclusion: The genetic factors underlying vitamin D deficiency in Korea included polymorphisms in the *GC*, *PDE3B*, *NADSYN1*, and *ACTE1P* genes. The biological mechanism of a non-coding SNP (rs12803256) for *DHCR7/NADSYN1* on vitamin D concentrations is unclear, warranting further investigations.

Keywords: Vitamin D deficiency; Genome-wide association study; Asians; Genetic predisposition to disease; Polymorphism, single nucleotide

Received: 27 August 2021, **Revised:** 12 October 2021,
Accepted: 19 October 2021

Corresponding author: Je Hyun Seo
 Veterans Medical Research Institute, Veterans Health Service Medical Center,
 53 Jinhwangdo-ro 61-gil, Gangdong-gu, Seoul 05368, Korea
Tel: +82-2-2225-1445, **Fax:** +82-2-2225-3950, **E-mail:** jazmin2@naver.com

*These authors contributed equally to this work.

Copyright © 2021 Korean Endocrine Society

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Vitamin D plays an essential role in bone mineralization, with its deficiency resulting in osteomalacia and rickets [1]. In addition, vitamin D deficiency is also related to various non-skeletal diseases, such as autoimmune disease, infectious disease, cardiovascular disease, and diabetes [2-6]. In recent studies, vitamin D deficiency (25-hydroxy-vitamin D [25(OH)D] ≤ 20 ng/mL) was estimated to be present in 1 billion people worldwide [1,7], constituting a significant public health concern [8]. In the United States, the third National Health and Nutrition Examination Survey showed a prevalence of vitamin D deficiency in approximately 30% of Americans [9]. An even higher prevalence was observed in certain ethnic groups, with $>80\%$ of African American adults and $>60\%$ of Hispanic adults having serum concentrations of 25(OH)D ≤ 20 ng/mL [10]. Among East Asians, 32.1% to 75.2% of Chinese adults and 53.6% of Japanese adults have concentrations of 25(OH)D ≤ 20 ng/mL [11]. The vitamin D deficiency prevalence values are similar among Koreans, being 47.3% in males and 64.5% in females according to the Korea National Health and Nutrition Examination Survey (KNHANES) [12]. Because of its high prevalence and clinical significance for several chronic diseases, research on vitamin D deficiency deserves attention.

Factors that can potentially affect vitamin D levels are mostly related to the environment and nutrition; these include aging, obesity, skin color, dietary intake, exposure to ultraviolet B sunlight, geographical latitude, and dietary supplement intake [8]. Vitamin D levels are, however, also determined by genetic background, as studies on twins have shown 43% to 70% heritability [13,14]. Understanding the genetic variants underlying vitamin D levels could aid vitamin D deficiency screening, which could be applied for providing vitamin D supplementation to high-risk groups. Genome-wide association studies (GWASs) have demonstrated several 25(OH)D-related loci, including those in group-specific component (*GC*), glutamine-dependent NAD(+) synthetase (*NADSYN1*), 7-dehydrocholesterol reductase (*DHCR7*), cytochrome P450 R1 (*CYP2R1*), and cytochrome P450 24A1 (*CYP24A1*) [15-21]. Most of these loci have been identified in subjects of European descent [16], and only the results on polymorphisms near *GC* and *DHCR7* were replicated in populations of African and Hispanic American ancestry [22]. Although *GC* and *DHCR7* variants have been reported in China Kadoorie Biobank [23], little is known about the genetic variants governing vitamin D levels in Koreans. Moreover, considering the high prevalence of vitamin D deficiency in Korean populations, it is

necessary to determine the genetic variants underlying serum 25(OH)D concentration using multi-cohort data. To achieve this goal, we conducted a GWAS in the Korean population using a multi-cohort population. Furthermore, an integrated functional and pathway analysis for vitamin D levels was performed based on single nucleotide polymorphisms (SNPs) combined with GWAS catalog data and the results of this study.

METHODS

Study subjects

Schematic plots of the study design for mega- and meta-analyses are shown in Fig. 1. Data were obtained from three cohorts: the GENIE (Gene-Environmental Interaction and Phenotype; Seoul National University Hospital Healthcare System Gangnam Center, $n=6,579$) cohort [24], the KARE (Korean Association Resource, Ansan/Ansung study; from the Korean Genome and Epidemiology Cohort, $n=5,493$) cohort, and the VHSMC (Veterans Health Service Medical Center; $n=570$) cohort. Each cohort has its own distinct characteristics. The GENIE cohort consists of mostly healthy people who undergo regular check-ups in the healthcare system. The KARE cohort is a representative cohort for genome research in Korea; it is a longitudinal cohort of the Ansan and Ansung communities in Korea. The VHSMC cohort is a hospital-based cross-sectional cohort of elderly men, with disease information. Among 12,642 participants from the three cohorts, 4,856 were excluded, and 7,786 were enrolled. After quality control, 7,590 subjects were finally analyzed (Fig. 1). The study was conducted in compliance with the Helsinki Declaration. The Institutional Review Board of the Seoul National University Hospital approved the storage of blood samples for genetic analysis with informed consent (IRB No. H-1103-127-357), and the Institutional Review Board approved this study protocol (IRB No. 1601-063-734) for the GENIE cohort. The Institutional Review Board of the VHSMC approved the study protocols for the KARE cohort (IRB No. 2019-08-014) and the VHSMC cohort (IRB No. 2020-01-053). The committee of the National Biobank of Korea (KBN-2019-054) and VHS Biobank (VBP-2020-02) approved the use of bioresources for this study.

Biochemical measurements

Serum 25(OH)D levels were measured by radioimmunoassay (DiaSorin Inc., Stillwater, MN, USA) for the GENIE cohort and by chemiluminescent microparticle immunoassay (CMIA) using an Architect i2000SR system (Abbott, Singapore) for the KARE and VHSMC cohorts. Vitamin D deficiency was defined

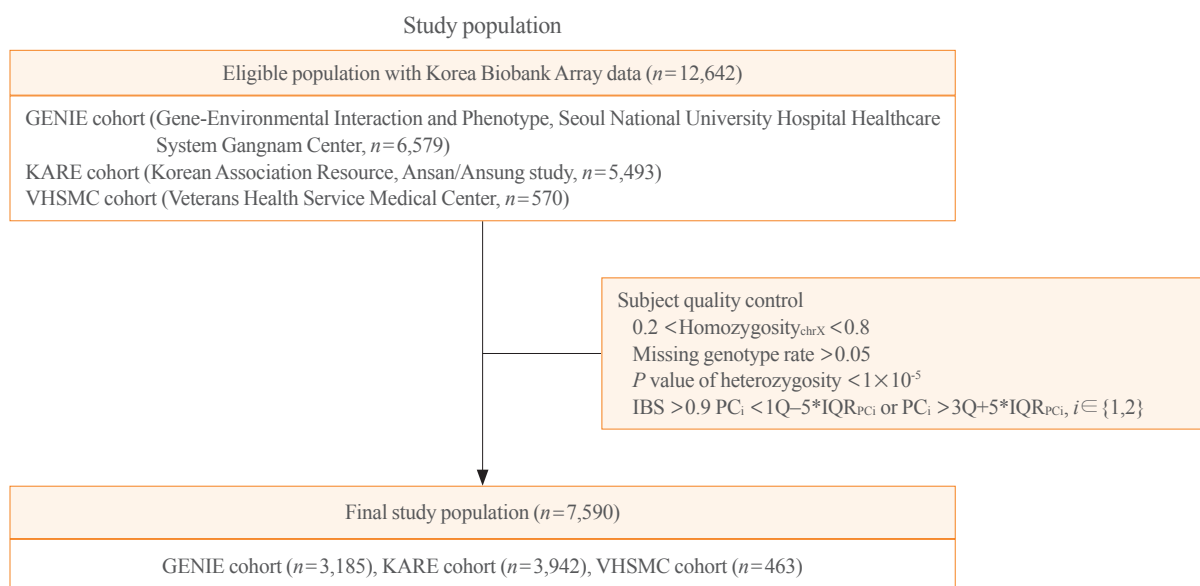


Fig. 1. Schematic plot of the study design: mega- and meta-analysis of genome-wide association study data. IBS, identity-by-state; PC, principal component; IQR, interquartile range.

using two cut-points, ≤ 20 ng/mL for deficiency and ≤ 10 ng/mL for severe deficiency [25,26].

Genotyping

Genomic DNA was separated from venous blood samples, and 100 ng of genomic DNA was genotyped using Korea Biobank Array (KoreanChip) and Affymetrix Axiom version 1.0 or 1.1 (Affymetrix, Santa Clara, CA, USA), which were designed by the Korean National Institute of Health, Korea [27]. Genotypes were called with the K-medoid algorithm to remove the batch effect [28]. The PLINK program version 1.9 (Boston, MA, USA) and ONETOOL [29] were used for data analysis and quality control. Samples meeting any of the following criteria were excluded: (1) sex inconsistencies ($0.2 < \text{homozygosity chr X} < 0.8$); (2) missing genotype rate over 0.05; or (3) P value of heterozygosity $< 1 \times 10^{-5}$. SNPs were filtered if (4) the call rate from the Hardy-Weinberg equilibrium (HWE) permutation test was low ($P < 1 \times 10^{-5}$); (5) there were duplicated SNPs; and (6) there was high heterogeneity of minor allele frequency (MAF) and HWE among the three cohorts. The schematic plots of quality control for analysis are shown in Supplemental Fig. S1. Imputation was conducted using the Trans-Omics for Precision Medicine genotype imputation server (<https://imputation.biodatacatalyst.nih.gov>) [30] and the Haplotype Reference Consortium release v1.1. Pre-phasing and imputation were conducted with the Eagle V 2.4 [31] and Minimac4 [32] program, respectively. After the imputation processes, imputed SNPs were removed if the R-

squared, an index for imputation accuracy, was less than 0.95; there were duplicated SNPs; the missing genotype rates were more than 0.05; P values for HWE were less than 1×10^{-5} ; or the MAFs were less than 0.05. In addition, subjects whose identity-by-state was > 0.9 and principal component (PC) score was outside the $5 \times$ interquartile range (IQR_{PC}) were removed. Finally, 7,590 subjects and their 1,695,891 SNPs were used for our analyses (Fig. 1). The detailed procedure is provided in Supplemental Fig. S1.

Statistical analyses

Baseline characteristics of the study population are presented as mean with standard deviation for continuous variables and number with proportion for categorical variables. We performed GWAS using lower cut-points, 10 and 20 ng/mL, for logistic regression analysis. Linear regression was also conducted to identify genetic variants for susceptibility, which are associated with serum 25(OH)D levels. PC scores were estimated with PLINK 1.9 and used to adjust the population substructure. Ten PC scores corresponding to the 10 largest eigenvalues, age, sex, season of blood draw, vitamin D supplement intake, history of kidney disease or liver malignancy or treatment, and difference by cohort, were included as covariates. Each analysis was performed in two ways: (1) mega-analysis and (2) meta-analysis with inverse variance base adjustments [33]. The genome-wide significance level for SNPs was set at 5×10^{-8} , annotated with ANNOVAR (ANNOtate VARIation) [34] and used for regional plots. We

used PLINK 1.9 to evaluate the linkage disequilibrium (LD) across the ancestral population with Phase 3 haplotype data of 1000 Genome Projects. We calculated SNP heritability by a liability scale according to the logistic cut-off values (≤ 10 and ≤ 20 ng/mL) or continuity for vitamin D concentrations with the GCTA software after adjusting age and sex as covariates [35,36].

Replication analysis, rare variant analysis, and expression analysis

For replication analysis, we searched the National Human Genome Research Institute-European Bioinformatics Institute GWAS catalog (<https://www.ebi.ac.uk/gwas/home>, December 2020; EFO_0004631) for SNPs, and selected 406 SNPs. If the direction of the strands and beta value coincided and the *P* value was less than 0.05, the result was defined as statistically significant. Among them, lead SNPs were selected by excluding those with the same signal. For rare variant association analysis, the quality control process was similar to that of the common variant GWAS, except that MAFs of less than 0.05 were not excluded. The rare variant association analysis included the SNP-Set (Sequence) Kernel Association Test (SKAT) and the burden test with family-based rare variant association test [37]. To investigate the association between the identified SNP and the nearby gene expression levels, we performed expression quantitative trait locus (eQTL) analysis using the Genotype-Tissue Expression (GTEx; <https://gtexportal.org>) project dataset for analysis. Moreover, we

annotated genes by constructing gene interaction networks with the STRING v.11 online platforms (<https://string-db.org/>), which are used for interactive gene networks.

RESULTS

Participant characteristics

The mean age of the study subjects was 55.81 ± 9.41 years, and 46.75% of subjects were men (Table 1). The mean serum 25(OH)D was 18.65 ± 7.68 ng/mL. The proportion of vitamin D-deficient patients (< 20 ng/mL) in the cohort was 62.32%, and that of patients with severe vitamin D deficiency (< 10 ng/mL) was 9.92%. The VHSMC cohort consisted of elderly patients and had a higher serum 25(OH)D level compared with the other cohorts. In the KARE cohort, 22 subjects (0.56%) showed the presence of chronic disease. Sixty-five percent of blood samples were collected in summer to autumn, and 6.80% of participants were taking vitamin D supplements.

GWAS of 25(OH)D: mega-analysis and meta-analysis

A Manhattan plot (Fig. 2) of the genetic signal for 25(OH)D levels showed that rs11723621 (*GC*: $P = 2.14 \times 10^{-24}$; $P = 1.08 \times 10^{-32}$, for mega-analysis and meta-analysis, respectively), rs7041 (*GC*: $P = 1.72 \times 10^{-9}$; $P = 1.79 \times 10^{-11}$, respectively), rs11023332 (phosphodiesterase 3B [*PDE3B*]: $P = 3.43 \times 10^{-11}$; $P = 3.2 \times 10^{-11}$, respectively), rs12803256 (actin epsilon 1, pseudogene [*ACTE1P*]:

Table 1. Baseline Characteristics of the Study Populations

Participant characteristic	GENIE cohort (<i>n</i> =3,185)	KARE cohort (<i>n</i> =3,942)	VHSMC cohort (<i>n</i> =463)	Total (<i>n</i> =7,590)
Age, yr	52.36±8.95	57.18±8.34	67.81±8.22	55.81±9.41
No. of male participants	1,710 (53.69)	1,888 (47.89)	179 (38.66)	3,777 (49.76)
Serum vitamin D level, ng/mL	19.64±7.25	17.25±7.09	23.85±11.25	18.65±7.68
Participant no. with vitamin D level <20 ng/mL	1,824 (57.27)	2,710 (68.75)	196 (42.33)	4,730 (62.32)
Participant no. with vitamin D level <10 ng/mL	163 (5.12)	565 (14.33)	25 (5.40)	753 (9.92)
Season of blood draw				
Spring (March–May)	625 (19.62)	1,092 (27.70)	0	1,717 (22.62)
Summer (June–August)	1,014 (31.84)	1,271 (32.24)	172 (37.15)	2,457 (32.37)
Fall (September–November)	1,001 (31.43)	1,214 (30.80)	251 (54.21)	2,466 (32.49)
Winter (December–February)	545 (17.11)	365 (9.26)	40 (8.64)	950 (12.52)
Presence of chronic disease	0	22 (0.56)	0	22 (0.29)
Vitamin D supplement	218 (6.84)	283 (7.18)	15 (3.24)	516 (6.80)

Values are expressed as mean±standard deviation or number (%). The presence of chronic disease affects vitamin D levels: renal disease, liver disease, and malignancy.

GENIE, Gene-Environmental Interaction and Phenotype; KARE, Korean Association Resource; VHSMC, Veterans Health Service Medical Center.

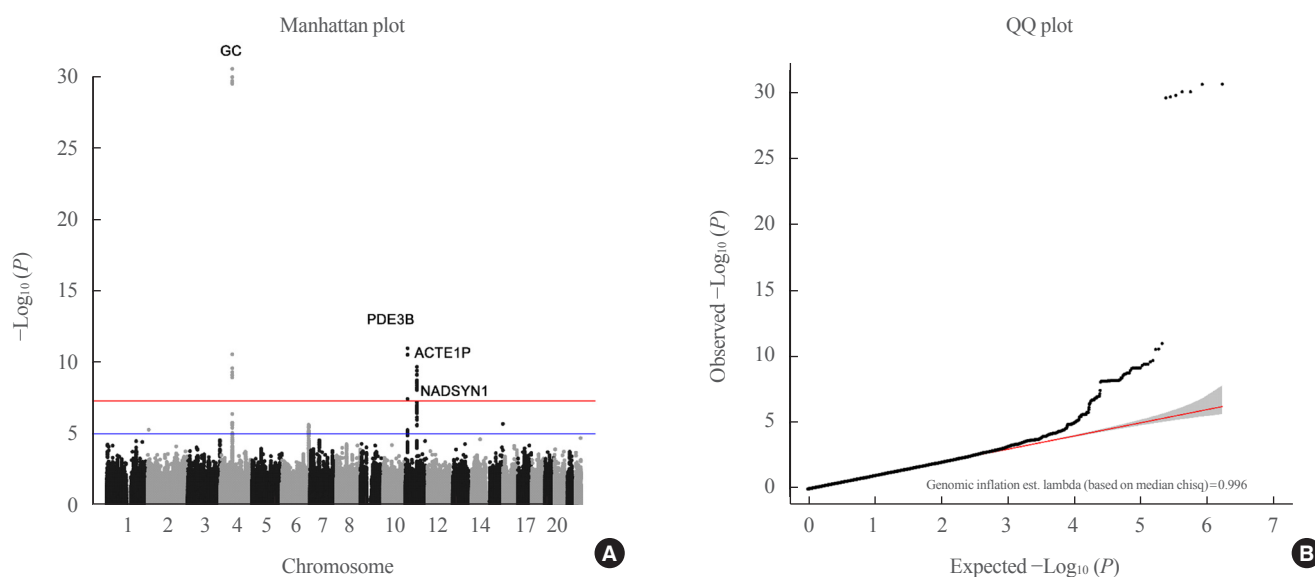


Fig. 2. Genome-wide association of circulating 25-hydroxylvitamin D concentrations by chromosome positions and $\log_{10} P$ value (Manhattan plot) and quantile-quantile plots (QQ-plot) for meta-analysis. (A) Manhattan plot for meta-analysis. (B) QQ-plot: The y-axis shows the observed $-\log_{10} P$ values, and the x-axis shows the expected $-\log_{10} P$ values for meta-analysis. GC, group-specific component; PDE3B, phosphodiesterase 3B; ACTE1P, actin epsilon 1, pseudogene; NADSYN1, glutamine-dependent NAD(+) synthetase.

$P=4.02 \times 10^{-8}$; $P=7.68 \times 10^{-11}$, respectively), and rs3831470 (*NADSYN1*: $P=2.6 \times 10^{-9}$ in meta-analysis) were significantly associated with vitamin D concentrations, as determined from mega- and meta-analyses (Table 2). In addition, rs3831470 (*NADSYN1*) showed marginally significant association in the mega-analysis ($P=9.68 \times 10^{-7}$) and significant association in the meta-analysis ($P=2.6 \times 10^{-9}$) (Table 2, Supplemental Fig. S2). rs78359207 (neuropeptide FF receptor 2 [*NPF2R2*]) and rs55715230 (parking coregulated [*PACRG*]) were suggestive loci in both mega- and meta-analyses (Table 2). In both logistic regression analyses (<10 and <20 ng/mL), only rs4588 (*GC*) showed significant association in mega- and meta-analyses (Supplemental Fig. S3).

Regional plots and SNP-based heritability estimates

The regional plot for *GC* was differently located in rs11723621 (base position [BP]: chr4:71749645, red dot) and in rs7041 (BP chr4:71752617, blue dot) (Supplemental Fig. S4A). Moreover, the regional plots for *PDE3B* (rs11023332) showed some association with the LD locus (Supplemental Fig. S4B). The regional association plot for *ACTE1P* (rs12803256) and for *NADSYN1* (rs3831470) demonstrates strong association with other loci in LD among East Asians (Supplemental Fig. S4C, D).

The analysis of heritability of vitamin D deficiency showed that the SNP-based heritability estimate (h^2_{SNP}) was 5.37% when the prevalence for severe vitamin D deficiency (<10 ng/mL)

was 8.0%, and the h^2_{SNP} was 8.59%, considering a prevalence of 60% for vitamin D deficiency (<20 ng/mL) [12]. The h^2_{SNP} of the vitamin D concentration by continuity was 7.23%, with marginal significance ($P=0.0611$) (Supplemental Fig. S5).

Replications and rare variants

Among 406 vitamin D concentration-associated SNPs from the GWAS catalog, 84 SNPs were found in our dataset, of which 12 lead SNPs showed significant association with serum vitamin D concentration (Table 3). The SNPs were located in the genes *GC*, *CYP2R1*; calcitonin-related polypeptide alpha (*CALCA*), *NADSYN1*, ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3 (*ST6GALNAC3*), *FLJ42102*, aquaporin 9 (*AQP9*); hepatic lipase C (*LIPC*), *NPF2R2*, carbamoyl phosphate synthetase 1 (*CPS1*), brain enriched myelin associated protein 1 (*BCAS1*); *CYP24A1*, zinc finger protein 808 (*ZNF808*); *ZNF701*, and SEC23 homolog A (*SEC23A*). The rare variant analysis revealed that vitamin D concentration was significantly associated with the argonaute RISC component 4 (*AGO4*) and ATP-binding cassette subfamily G member 2 (*ABCG2*) genes from the SKAT test and with the *AGO4*, roundabout guidance receptor 2 (*ROBO2*), anthrax toxin receptor 1 (*ANTXR1*), and enoyl-CoA hydratase domain containing 3 (*ECHDC3*) genes from the burden test (Supplemental Table S1).

Table 2. Single Nucleotide Polymorphisms Identified to Be Associated with Serum 25-Hydroxyvitamin D Concentrations in Genome-Wide Analyses

CHR	SNPs	BP	Alt/Ref	MAF _{all}	MAF _{gnomAD-EA}	HWE	Rsq	Genotype	Location	Gene	Mega-analysis		Meta-analysis		
											Beta	STAT	P value	Z-score ^a	P value ^a
4	rs11723621	71749645	G/A	0.29	0.2593	0.67	0.96	Imputed	Intron	GC	1.28	10.23	2.14E-24	11.90	1.08E-32
11	rs11023332	14762564	C/G	0.421	0.3566	0.49	1.00	Genotyped	Intron	PDE3B	0.77	6.64	3.43E-11	6.63	3.2E-11
4	rs7041	71752617	C/A	0.261	0.2872	1.00	0.99	Genotyped	Missense Coding_sequence Intron	GC	0.79	6.03	1.72E-09	6.72	1.79E-11
11	rs12803256	71421822	G/A	0.372	0.3885	0.12	0.95	Imputed	Non_coding_transcript	ACTEIP	0.65	5.50	4.02E-08	6.51	7.68E-11
11	rs3831470	71454898	C/A	0.387	-	0.21	0.96	Imputed	Intron	NADSYN1	0.58	4.90	9.68E-07	5.84	2.6E-09
4	rs78359207	72032699	C/T	0.319	0.2973	0.84	0.96	Imputed	Intron_genic_upstream_transcript	NPFRR2	0.59	4.78	1.74E-06	-4.77	1.9E-06
6	rs55715230	162756538	A/C	0.213	0.1849	0.66	0.97	Imputed	Intron_genic_upstream_transcript	P4CRG	0.66	4.69	2.71E-06	4.62	3.9E-06
2	rs306141	6458749	C/T	0.274	0.3318	0.83	0.98	Genotyped	-	-	0.60	4.66	3.19E-06	-4.57	4.9E-06
8	rs17382663	128090786	T/C	0.091	0.1418	0.50	0.99	Genotyped	Intron	PVT1	-0.91	-4.55	5.32E-06	-3.95	7.9E-05
14	rs58788626	70865550	T/C	0.106	0.09395	0.72	0.99	Genotyped	-	-	-0.84	-4.50	7.03E-06	-4.25	2.2E-05

CHR, chromosome; SNP, single nucleotide polymorphism; BP, base position; Alt, alternative allele; Ref, reference allele; MAF, minor allele frequency; gnomAD-EA, The Genome Aggregation Database-East Asian; HWE, Hardy-Weinberg equilibrium; Rsq, R-squared; GC, group specific component; PDE3B, phosphodiesterase 3B; ACTEIP, actin epsilon 1, pseudogene; NADSYN1, glutamine-dependent NAD(+) synthetase; NPFRR2, neuropeptide FF receptor 2; P4CRG, parking coregulated; PVT1, plasmacytoma variant translocation 1.

^aInverse variance-based meta-analysis.

Biologic annotation with eQTL analysis and network analysis

The *ACTEIP* gene is a long non-coding RNA gene, and rs12803256 (*ACTEIP*) is located upstream of *DHCR7/NADSYN1* (Supplemental Fig. S6A). The box plot of eQTL of the genes related to vitamin D metabolism, including *NADSYN1*, in skin tissues of both non-sun-exposed and sun-exposed individuals, whole blood, and liver tissue, shows significantly different expression, according to the allele rs12803256 in GTEx data (Supplemental Fig. S6B). Among subjects who were not sun-exposed, *NADSYN1* and *DHCR7* were differentially expressed depending on the rs12803256 alleles ($P=3.9\times 10^{-27}$, $P=3.5\times 10^{-11}$, respectively), and among sun-exposed subjects, *NADSYN1* and *DHCR7* were also differentially expressed depending on the rs12803256 alleles ($P=9.4\times 10^{-17}$, $P=2.4\times 10^{-8}$, respectively). This trend was also observed for whole blood and liver tissue. The interaction network for proteins related to vitamin D metabolism, which was adapted from genes from GWAS catalog data, as well as our data, revealed that the proteins NADSYN1, DHCR7, GC, CYP24A1, and CYP2R1 are involved in vitamin D catabolic pathways (Fig. 3A). The possible mechanism of the effect of the rs12803256 *DHCR7/NADSYN1* SNP on vitamin D metabolism may be postulated from the results of the eQTL analysis (Fig. 3B).

DISCUSSION

In this study, we confirmed that rs11723621 (*GC*), rs7041 (*GC*), and rs3831470 (*NADSYN1*) were related to serum vitamin D levels, with genome-wide significance in the East Asian population. We also discovered rs12803256 (*ACTEIP*) as a new genetic variant associated with serum vitamin D levels, with genome-wide significance. In previous studies, GWAS for genes related to vitamin D levels among multi-ethnic patients were conducted mostly in those of European ancestry [20,22,38,39]. From the meta-GWAS on the Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits (SUNLIGHT) consortium, which included 33,996 subjects of European descent from 15 cohorts [38], results for SNPs, such as rs2282690 (*GC*), rs3829251 (*NADSYN1/DHCR7*), and rs2060793 (*CYP2R1*), were replicated [15]. The literature on the genetic architecture underlying vitamin D levels in populations of non-European ancestry is sparse. In this respect, one advantage of this study is the analysis, discovery, and replication of results on polymorphic variations related to vitamin D levels among East Asians. In our study, variants in *GC*, *CYP2R1*, *CALCA*, *NADSYN1*, *ST6G*-

Table 3. Replications of Single Nucleotide Polymorphisms Identified to Be Associated with Serum 25-Hydroxyvitamin D Concentration in Genome-Wide Analyses

CHR	SNPs	Ref	Alt	BP	Location	Gene	Mega-analysis		
							Beta	SE	P value
1	rs12144344	C	T	76373851	Intronic	<i>ST6GALNAC3</i>	-0.3182	0.1493	0.033
2	rs1047891	C	A	210675783	Exonic	<i>CPSI</i>	0.4943	0.1479	8.38E-04
4	rs11723621	A	G	71749645	Intronic	<i>GC</i>	1.2797	0.1251	2.14E-24
4	rs7041	A	C	71752617	Exonic	<i>GC</i>	-0.7853	0.1302	1.72E-09
4	rs13107347	T	C	72109031	Intronic	<i>NPFFR2</i>	-0.4338	0.1155	1.74E-04
11	rs2060793	A	G	14893764	Intergenic	<i>CYP2R1;CALCA</i>	0.4052	0.1173	5.56E-04
11	rs12803256	A	G	71421822	ncRNA_exonic	<i>FLJ42102</i>	-0.6508	0.1184	4.02E-08
11	rs4423214	C	T	71462208	Intronic	<i>NADSYN1</i>	-0.5846	0.1179	7.31E-07
14	rs8018720	G	C	39086981	Exonic	<i>SEC23A</i>	0.3071	0.1184	0.010
15	rs261291	T	C	58387979	Intergenic	<i>AQP9;LIPC</i>	0.2842	0.1147	0.013
19	rs8113404	C	T	52562326	Intergenic	<i>ZNF808;ZNF701</i>	-0.7612	0.2487	0.002
20	rs17216707	T	C	54115823	Intergenic	<i>BCAS1;CYP24A1</i>	0.5312	0.2229	0.017

CHR, chromosome; SNP, single nucleotide polymorphism; Ref, reference allele; Alt, alternative allele; BP, base position; SE, standard error; *ST6GALNAC3*, ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3; *CPSI*, carbamoyl phosphate synthetase I; *GC*, group specific component; *NPFFR2*, neuropeptide FF receptor 2; *CYP2R1*, cytochrome P450 2R1; *CALCA*, calcitonin-related polypeptide alpha; *FLJ42102*, uncharacterized LOC399923; *NADSYN1*, glutamine-dependent NAD(+) synthetase; *SEC23A*, SEC23 homolog A; *AQP9*, aquaporin 9; *LIPC*, lipase C, hepatic type; *ZNF808*, zinc finger protein 808; *ZNF701*, zinc finger protein 701; *BCAS1*, brain enriched myelin associated protein 1; *CYP24A1*, cytochrome P450 24A1.

ALNAC3, *FLJ42102*, *AQP9;LIPC*, *NPFFR2*, *CPSI*, *BCAS1;CYP24A1*, *ZNF808;ZNF701*, and *SEC23A* were confirmed to underlie vitamin D levels in multiple cohorts including Koreans.

Serum concentrations of 25(OH)D are influenced by environmental factors; however, racial/ethnic differences and skin color are related to genetic factors [40]. In previous twin and family studies, the heritability for blood vitamin D concentration was reported as 43% to 70% [41]. A study on the effect of genetic and non-genetic factors on 25(OH)D levels in response to vitamin D supplementation revealed that the models' genetic risk factors included rs7041 and rs4588 (*GC*), rs2228570 (vitamin D receptor [*VDR*]), and rs10741657 (*CYP2R1*), which gave better predictive values [42]. In other GWASs, several significant loci, including *GC*, *NADSYN1/DHCR7*, *CYP2R1*, and *CYP24A1*, were identified to be related to the vitamin D metabolic pathway [8,43]. Here, we conducted a network analysis based on the previous genetic studies and predicted different metabolic pathways that could affect vitamin D levels. In a recent study on UK Biobank (UKBB) participants, the heritability of 25(OH)D was estimated as 0.32 and the SNP-based heritability was estimated as 0.13 [32]. In the expanded SUNLIGHT consortium study, the overall estimate of SNP heritability was 0.075 [38]. In our study, the SNP-based heritability ranged from 0.072 to 0.08.

Investigations on candidate SNPs are necessary for the discovery of new associations. In the study of the expanded SUNLIGHT consortium, new loci, such as rs10745742 in the amidohydrolase domain-containing protein 1 (*AMDHD1*) gene and rs8018720 (*SEC23A*), were reported [38]. The Trans-Ethnic Evaluation of vitamin D (TRANSCEN-D) study on low-frequency loci for African, Hispanic, European populations showed that results for rs796666294 in the kinesin family member 4B (*KIF4B*) gene and the rs1410656 in the 5-hydroxytryptamine receptor 2A (*HTR2A*) gene were replicated in the African cohort, in addition to the results on SNPs in the *GC* and *DHCR7* genes [22]. A previous study on rare variants associated with vitamin D levels using whole-genome sequencing identified two new variants, rs3819817 in the histidine ammonia-lyase (*HAL*) gene and rs2277458 in the gene for gem-associated protein 2 (*GEMIN2*) [44]. In our study, several low-frequency SNPs, such as those in *AGO4*, *ABCG2*, *ROBO2*, *ANTXR1*, and *ECHDC3* genes, were related to vitamin D deficiency in addition to SNPs in well-established genes, such as *GC*, *NADSYN1*, and *PDE3B*. A literature review suggests that the *ABCG2* gene, a member of the ATP-binding cassette transporter family, may have a causal association between the *GC* gene and uric acid level with small clinical effects on vitamin D levels [45]. A recent trans-ethnic

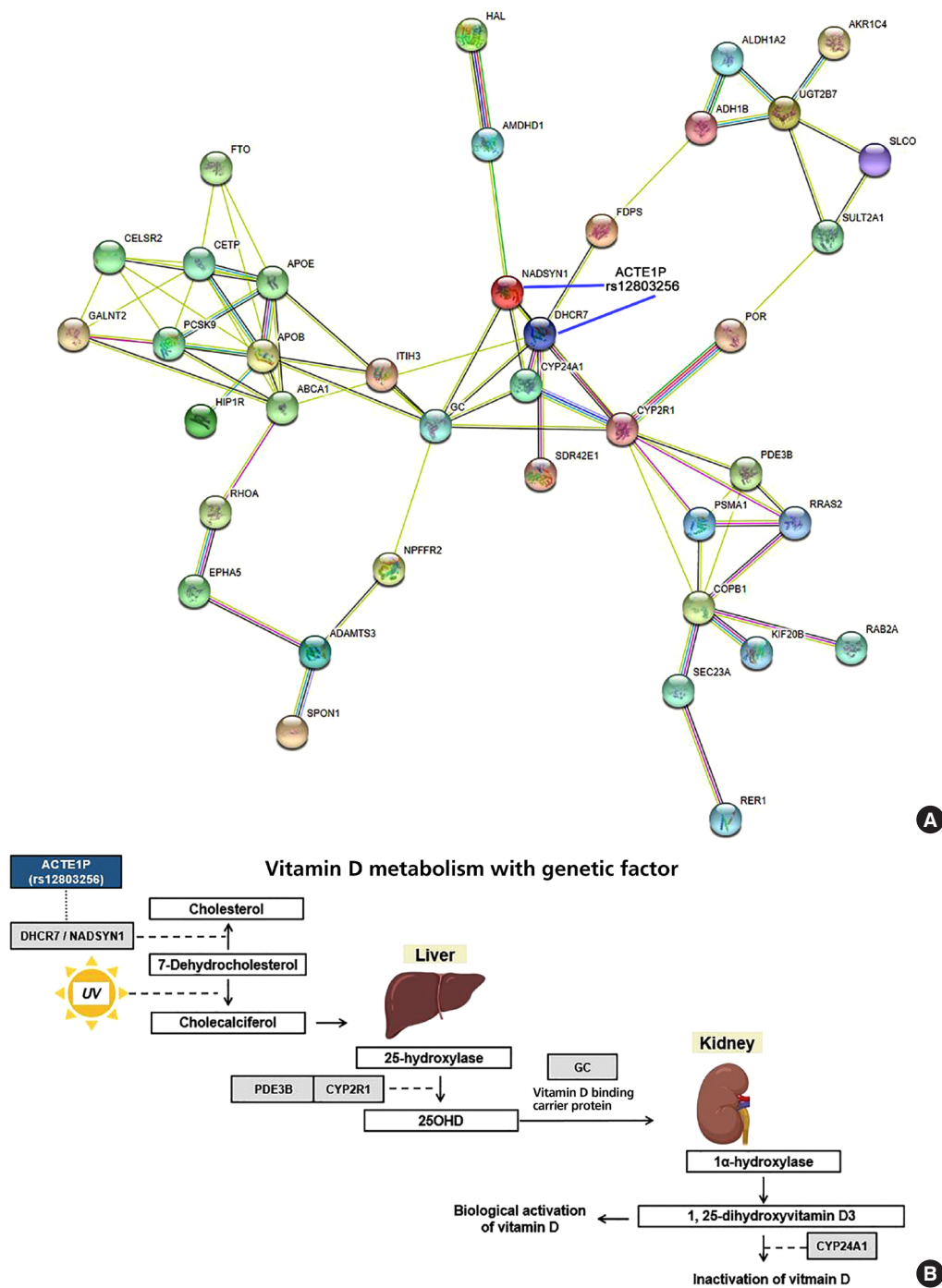


Fig. 3. Network analysis and postulated mechanisms for the effect of actin epsilon 1, pseudogene (*ACTE1P*) rs12803256 single nucleotide polymorphism (SNP) on serum 25-hydroxyvitamin D concentrations. (A) Interaction network for proteins related to vitamin D deficiency, which was adapted from the genome-wide association study (GWAS) catalog data and our data. *ACTE1P* gene is a long non-coding RNA, and rs12803256 was related with the expression of glutamine-dependent NAD(+) synthetase (*NADSYN1*) and 7-dehydrocholesterol reductase (*DHCR7*) on Genotype-Tissue Expression database (GTEx) data. Phosphodiesterase 3B (*PDE3B*) was shown to be associated with cytochrome P450 2R1 (*CYP2R1*) and SEC23 homolog A (*SEC23A*), and the results were replicated in our study. (B) A mechanism for the effect of the rs12803256 SNP of *DHCR7/NADSYN1* on vitamin D deficiency is postulated. Moreover, the *PDE3B* gene was related to the *CYP2R1* gene. *COPB1*, coat complex subunit beta 1; *RRAS2*, RAS-related protein; GC, group-specific component.

study using UKBB and BioBank Japan data showed that *ABCG2* was related to vitamin D metabolism in kidney stone disease [46], which was found to be a rare variant for vitamin D concentration in our Korean cohort.

In addition to *GC*, *NADSYN1/DHCR7* needs attention since vitamin D is activated in the skin by the action of UV rays (Fig. 3). One of the SNPs significantly associated with vitamin D levels in our functional analysis was rs12803256 (*ACTEIP*). The *ACTEIP* gene or “actin epsilon 1, pseudogene” is one of the long non-coding RNA genes for which a direct function is difficult to predict. Since rs12803256 (*ACTEIP*) is located 0.01 Mb upstream of the *DHCR7* locus, it may be related to *DHCR7* gene function. This rare finding was replicated in a recent study conducted during the same study period as our current study. In fact, Revez et al. [32] identified 143 loci associated with vitamin D levels from 8,806,780 SNPs based on UKBB; they were also interested in this region and considered that *DHCR7* and *NADSYN1* warrant additional investigation. One of our hypotheses is that the region of rs12803256 (71.42 of chromosome 11) is functionally related with *NADSYN1/DHCR7/RP11-660L16.2*, based on eQTL analysis in skin tissue. A follow-up study will be necessary to test for the possible function of *ACTEIP* as a promoter or enhancer of *DHCR7* gene expression. It is particularly noteworthy that results on recently identified loci, such as cysteine- and glycine-rich protein 1 (*CRP1*) and *FLJ42102* genes, were replicated in our study [32].

The major strengths of our study are the inclusion of a relatively large study sample of the East Asian population and it being the first study on vitamin D concentration for Koreans. We observed that *NADSYN1/DHCR7* was differentially expressed in the skin, regardless of sun exposure, depending on the rs12803256 SNP. These findings suggest that vitamin D levels may be functionally related to non-coding regions, as well as protein-coding genes. This study also has potential applications for public health and human research. Trans-ethnic analysis related to vitamin D in East Asians is a candidate topic based on our dataset consisting of Koreans for further studies. As previous studies have shown associations of vitamin D concentrations with type 2 diabetes using Mendelian randomization analysis [47], our cohort data could provide causality for vitamin D-related diseases in East Asians [23].

However, there are a few limitations to this study. First, three different cohorts were incorporated in the study, and we found significant findings by meta-analysis rather than by mega-analysis, likely because of differences in age and health status among the different cohorts. Second, our study lacks functional experi-

ments for the role of *ACTEIP*, as the effect of rs12803256 on vitamin D levels is a hypothetical postulation, although it is based on a public database. However, these findings could be complemented by experiments using the clustered regularly interspaced short palindromic repeats CRISPR-Cas9 methods in skin cell lines directly linked to vitamin D metabolism. Further studies are planned in this respect and should provide additional functional data. Third, since all the information collected was based on self-reported health surveys, there may be acquiescence or recall bias, which may have resulted in misclassification. Since this limitation can be overcome when a large number of subjects is included in the study, and owing to the fact that our results have been replicated in a recent meta-analysis from the large UKBB, this problem could be considered resolved. Fourth, we used vitamin D concentration without log-transformation because of normal distribution. However, it is expected that better results may be obtained after log-transformation in a subsequent study.

In conclusion, our study showed that the genetic factors that predispose Koreans to vitamin D deficiency include SNPs in the *GC*, *PDE3B*, *NADSYN1*, *NPFFR2*, and *ACTEIP* genes. In addition, the results for SNPs located in the *CYP2R1*, *ST6GALNAC3*, *FLJ42102*, *AQP9*, *CPS1*, *CYP24A1*, *ZNF808*, and *SEC23A* genes were replicated from previous studies on different ethnic groups. Our study showed that non-coding regions may be related to vitamin D levels. The possible mechanisms for the effect of the rs12803256 SNP on the expression of the *DHCR7/NADSYN1* gene and on vitamin D deficiency warrants further investigation.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

This study was supported by a Veterans Health Service Medical Center Research Grant (grant no.: VHSMC20026).

GENIE cohort (Gene-Environmental Interaction and phenotype): This included 7,999 Korean adults who had undergone a routine health check-up at the Seoul National University Hospital Healthcare System Gangnam Center from January 2014 to December 2014. Of them, 6,579 participants were included in the Korea Biobank array (v1.0). They were asked for their consent, and their blood samples were collected and stored for the

research. The Institutional Review Board of Seoul National University Hospital approved the storage of blood samples for genetic analysis with informed consent (IRB No. H-1103-127-357).

KARE cohort (KARE, Korean Association Resource; Ansan/Ansung study): This study was conducted with bioresources from National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (KBN-2019-054).

VHSMC cohort (Veterans Health Service Medical Center): This study was conducted thanks to bioresources from the Veterans Medical Research Institute Biobank, Republic of Korea (VBP-2020-02).

We would like to thank RexSoft for providing technical support for the Korean-Chip data quality control and supporting imputation analysis.

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

AUTHOR CONTRIBUTIONS

Conception or design: Y.A.K., J.W.Y., Y.L., H.J.C., D.J.P., C.S.S., J.H.S. Acquisition, analysis, or interpretation of data: Y.A.K., J.W.Y., Y.L., H.J.C., J.W.Y., E.B., S.H.K., S.E.A., A.R.D., H.J., S.W., J.H.S. Drafting the work or revising: Y.A.K., J.W.Y., Y.L., H.J., S.W., J.H.S. Final approval of the manuscript: Y.A.K., J.W.Y., S.W., J.H.S.

ORCID

Ye An Kim <https://orcid.org/0000-0001-8986-2632>

Ji Won Yoon <https://orcid.org/0000-0001-9003-0614>

Je Hyun Seo <https://orcid.org/0000-0003-3127-7160>

REFERENCES

- Holick MF. Vitamin D deficiency. *N Engl J Med* 2007;357:266-81.
- Scaranti M, Junior Gde C, Hoff AO. Vitamin D and cancer: does it really matter? *Curr Opin Oncol* 2016;28:205-9.
- Altieri B, Muscogiuri G, Barrea L, Mathieu C, Vallone CV, Mascitelli L, et al. Does vitamin D play a role in autoimmune endocrine disorders?: a proof of concept. *Rev Endocr Metab Disord* 2017;18:335-46.
- Skaaby T, Thuesen BH, Linneberg A. Vitamin D, cardiovascular disease and risk factors. *Adv Exp Med Biol* 2017;996:221-30.
- Lee CJ, Iyer G, Liu Y, Kalyani RR, Bamba N, Ligon CB, et al. The effect of vitamin D supplementation on glucose metabolism in type 2 diabetes mellitus: a systematic review and meta-analysis of intervention studies. *J Diabetes Complications* 2017;31:1115-26.
- Gois PH, Ferreira D, Olenski S, Seguro AC. Vitamin D and infectious diseases: simple bystander or contributing factor? *Nutrients* 2017;9:651.
- Bischoff-Ferrari HA, Giovannucci E, Willett WC, Dietrich T, Dawson-Hughes B. Estimation of optimal serum concentrations of 25-hydroxyvitamin D for multiple health outcomes. *Am J Clin Nutr* 2006;84:18-28.
- Makariou S, Liberopoulos EN, Elisaf M, Challa A. Novel roles of vitamin D in disease: what is new in 2011? *Eur J Intern Med* 2011;22:355-62.
- Looker AC, Johnson CL, Lacher DA, Pfeiffer CM, Schleicher RL, Sempos CT. Vitamin D status: United States, 2001-2006. *NCHS Data Brief* 2011;59:1-8.
- Forrest KY, Stuhldreher WL. Prevalence and correlates of vitamin D deficiency in US adults. *Nutr Res* 2011;31:48-54.
- van Schoor N, Lips P. Global overview of vitamin D status. *Endocrinol Metab Clin North Am* 2017;46:845-70.
- Choi HS, Oh HJ, Choi H, Choi WH, Kim JG, Kim KM, et al. Vitamin D insufficiency in Korea: a greater threat to younger generation: the Korea National Health and Nutrition Examination Survey (KNHANES) 2008. *J Clin Endocrinol Metab* 2011;96:643-51.
- Hunter D, De Lange M, Snieder H, MacGregor AJ, Swaminathan R, Thakker RV, et al. Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. *J Bone Miner Res* 2001;16:371-8.
- Karohl C, Su S, Kumari M, Tangpricha V, Velezdar E, Vaccarino V, et al. Heritability and seasonal variability of vitamin D concentrations in male twins. *Am J Clin Nutr* 2010;92:1393-8.
- Ahn J, Yu K, Stolzenberg-Solomon R, Simon KC, McCullough ML, Gallicchio L, et al. Genome-wide association study of circulating vitamin D levels. *Hum Mol Genet* 2010;19:2739-45.
- Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet* 2010;376:180-8.
- Malik S, Fu L, Juras DJ, Karmali M, Wong BY, Gozdzik A, et al. Common variants of the vitamin D binding protein gene

- and adverse health outcomes. *Crit Rev Clin Lab Sci* 2013; 50:1-22.
18. Anderson D, Holt BJ, Pennell CE, Holt PG, Hart PH, Blackwell JM. Genome-wide association study of vitamin D levels in children: replication in the Western Australian Pregnancy Cohort (Raine) study. *Genes Immun* 2014;15:578-83.
 19. Moy KA, Mondul AM, Zhang H, Weinstein SJ, Wheeler W, Chung CC, et al. Genome-wide association study of circulating vitamin D-binding protein. *Am J Clin Nutr* 2014;99:1424-31.
 20. Sapkota BR, Hopkins R, Bjonnes A, Ralhan S, Wander GS, Mehra NK, et al. Genome-wide association study of 25(OH) vitamin D concentrations in Punjabi Sikhs: results of the Asian Indian diabetic heart study. *J Steroid Biochem Mol Biol* 2016;158:149-56.
 21. Wang J, Thingholm LB, Skieceviciene J, Rausch P, Kummen M, Hov JR, et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat Genet* 2016;48:1396-406.
 22. Hong J, Hatchell KE, Bradfield JP, Bjonnes A, Chesi A, Lai CQ, et al. Transethnic evaluation identifies low-frequency loci associated with 25-hydroxyvitamin D concentrations. *J Clin Endocrinol Metab* 2018;103:1380-92.
 23. Lu L, Bennett DA, Millwood IY, Parish S, McCarthy MI, Mahajan A, et al. Association of vitamin D with risk of type 2 diabetes: a Mendelian randomisation study in European and Chinese adults. *PLoS Med* 2018;15:e1002566.
 24. Lee C, Choe EK, Choi JM, Hwang Y, Lee Y, Park B, et al. Health and Prevention Enhancement (H-PEACE): a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea. *BMJ Open* 2018;8:e019327.
 25. Ross AC, Manson JE, Abrams SA, Aloia JF, Brannon PM, Clinton SK, et al. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. *J Clin Endocrinol Metab* 2011;96:53-8.
 26. Holick MF, Binkley NC, Bischoff-Ferrari HA, Gordon CM, Hanley DA, Heaney RP, et al. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2011;96:1911-30.
 27. Moon S, Kim YJ, Han S, Hwang MY, Shin DM, Park MY, et al. The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. *Sci Rep* 2019;9:1382.
 28. Seo S, Park K, Lee JJ, Choi KY, Lee KH, Won S. SNP genotype calling and quality control for multi-batch-based studies. *Genes Genomics* 2019;41:927-39.
 29. Song YE, Lee S, Park K, Elston RC, Yang HJ, Won S. ONE-TOOL for the analysis of family-based big data. *Bioinformatics* 2018;34:2851-3.
 30. Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, et al. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet* 2019;15:e1008500.
 31. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* 2016;48:1443-8.
 32. Revez JA, Lin T, Qiao Z, Xue A, Holtz Y, Zhu Z, et al. Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nat Commun* 2020; 11:1647.
 33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190-1.
 34. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
 35. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 2011;88:294-305.
 36. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76-82.
 37. Choi S, Lee S, Cichon S, Nothen MM, Lange C, Park T, et al. FARVAT: a family-based rare variant association test. *Bioinformatics* 2014;30:3197-205.
 38. Jiang X, O'Reilly PF, Aschard H, Hsu YH, Richards JB, Dupuis J, et al. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nat Commun* 2018;9:260.
 39. Manousaki D, Mitchell R, Dudding T, Haworth S, Harroud A, Forgetta V, et al. Genome-wide association study for vitamin D levels reveals 69 independent loci. *Am J Hum Genet* 2020;106:327-37.
 40. Bouillon R. Genetic and racial differences in the vitamin D endocrine system. *Endocrinol Metab Clin North Am* 2017;

- 46:1119-35.
41. Jiang X, Kiel DP, Kraft P. The genetics of vitamin D. *Bone* 2019;126:59-77.
 42. Yao P, Sun L, Lu L, Ding H, Chen X, Tang L, et al. Effects of genetic and nongenetic factors on total and bioavailable 25(OH)D responses to vitamin D supplementation. *J Clin Endocrinol Metab* 2017;102:100-10.
 43. Delanghe JR, Speeckaert R, Speeckaert MM. Behind the scenes of vitamin D binding protein: more than vitamin D binding. *Best Pract Res Clin Endocrinol Metab* 2015;29:773-86.
 44. Manousaki D, Dudding T, Haworth S, Hsu YH, Liu CT, Medina-Gomez C, et al. Low-frequency synonymous coding variation in CYP2R1 has large effects on vitamin D levels and risk of multiple sclerosis. *Am J Hum Genet* 2017;101:227-38.
 45. Thakkinstian A, Anothaisintawee T, Chailurkit L, Ratana-chaiwong W, Yamwong S, Sritara P, et al. Potential causal associations between vitamin D and uric acid: bidirectional mediation analysis. *Sci Rep* 2015;5:14528.
 46. Howles SA, Wiberg A, Goldsworthy M, Bayliss AL, Gluck AK, Ng M, et al. Genetic variants of calcium and vitamin D metabolism in kidney stone disease. *Nat Commun* 2019;10:5175.
 47. Zheng JS, Luan J, Sofianopoulou E, Sharp SJ, Day FR, Imamura F, et al. The association between circulating 25-hydroxyvitamin D metabolites and type 2 diabetes in European populations: a meta-analysis and Mendelian randomisation analysis. *PLoS Med* 2020;17:e1003394.