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

Ye An Kim¹*, Ji Won Yoon²*, Young Lee³, Hyuk Jin Choi², Jae Won Yun¹, Eunsin Bae³, Seung-Hyun Kwon³, So Eun Ahn⁴, Ah-Ra Do⁴, Heejin Jin⁵, Sungho Won⁴,⁵,⁶, 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
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 groupspecific 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 (CMA) using an Architect i2000SR system (Abbott, Singapore) for the KARE and VHSMC cohorts. Vitamin D deficiency was defined...
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 \(0.2<Homozygosity_{\text{chrX}}<0.8\); (2) missing genotype rate over 0.05; or (3) \(P\) value of heterozygosity \(<1\times10^{-5}\). SNPs were filtered if (4) the call rate from the Hardy-Weinberg equilibrium (HWE) permutation test was low \((P<1\times10^{-5})\); (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.nhlbi.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 \(1\times10^{-5}\); 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× interquartile range \((\text{IQR}_{\text{PC}})\) 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 \(5\times10^{-8}\), annotated with ANNOVAR (ANNOtate VARiation) [34] and used for regional plots.
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: \( P=2.14\times10^{-24} \); \( P=1.08\times10^{-32} \), for mega-analysis and meta-analysis, respectively), rs7041 (GC: \( P=1.72\times10^{-5} \); \( P=1.79\times10^{-11} \), respectively), rs11023332 (phosphodiesterase 3B \( \text{PDE3B} \); \( P=3.43\times10^{-11} \); \( P=3.2\times10^{-11} \), respectively), rs12803256 (actin epsilon 1, pseudogene \( \text{ACTE1P} \); 

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

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-hydroxyvitamin 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 (\(P = 2.6 \times 10^{-10}\)) (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 (\(h_{SNP}^2\)) was 5.37% when the prevalence for severe vitamin D deficiency (<10 ng/mL) was 8.0%, and the \(h_{SNP}^2\) was 8.59%, considering a prevalence of 60% for vitamin D deficiency (<20 ng/mL) [12]. The \(h_{SNP}^2\) 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-acetylglactosaminide alpha-2,6-sialyltransferase 3 (ST6GALNA3), FLJ42102, aquaporin 9 (AQP9); hepatic lipase C (LIPC), NPFFR2, 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).
In our study, variants in 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 (ACTE1P) 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 (CY2R1), 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-
Korean GWAS of Serum Vitamin D

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

<table>
<thead>
<tr>
<th>CHR</th>
<th>SNPs</th>
<th>Ref</th>
<th>Alt</th>
<th>BP</th>
<th>Location</th>
<th>Gene</th>
<th>Beta</th>
<th>SE</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs12144344</td>
<td>C</td>
<td>T</td>
<td>76373851</td>
<td>Intronic</td>
<td>ST6GALNA3</td>
<td>-0.3182</td>
<td>0.1493</td>
<td>0.033</td>
</tr>
<tr>
<td>2</td>
<td>rs1047891</td>
<td>C</td>
<td>A</td>
<td>210675783</td>
<td>Exonic</td>
<td>CPS1</td>
<td>0.4943</td>
<td>0.1479</td>
<td>8.38E-04</td>
</tr>
<tr>
<td>4</td>
<td>rs11723621</td>
<td>A</td>
<td>G</td>
<td>71749645</td>
<td>Intronic</td>
<td>GC</td>
<td>1.2797</td>
<td>0.1251</td>
<td>2.14E-24</td>
</tr>
<tr>
<td>4</td>
<td>rs7041</td>
<td>A</td>
<td>C</td>
<td>71752617</td>
<td>Exonic</td>
<td>GC</td>
<td>-0.7853</td>
<td>0.1302</td>
<td>1.72E-09</td>
</tr>
<tr>
<td>4</td>
<td>rs13107347</td>
<td>T</td>
<td>C</td>
<td>72109031</td>
<td>Intronic</td>
<td>NPFFR2</td>
<td>-0.4338</td>
<td>0.1155</td>
<td>1.74E-04</td>
</tr>
<tr>
<td>11</td>
<td>rs2060793</td>
<td>A</td>
<td>G</td>
<td>14893764</td>
<td>Intragenic</td>
<td>CYP2R1;CALCA</td>
<td>0.4052</td>
<td>0.1173</td>
<td>5.56E-04</td>
</tr>
<tr>
<td>11</td>
<td>rs12803256</td>
<td>A</td>
<td>G</td>
<td>71421822</td>
<td>3′UTR_exonic</td>
<td>FLJ42102</td>
<td>-0.6508</td>
<td>0.1184</td>
<td>4.02E-08</td>
</tr>
<tr>
<td>11</td>
<td>rs4423214</td>
<td>T</td>
<td>C</td>
<td>71642208</td>
<td>Intronic</td>
<td>NADSYN1</td>
<td>-0.5846</td>
<td>0.1179</td>
<td>7.31E-07</td>
</tr>
<tr>
<td>14</td>
<td>rs8018720</td>
<td>G</td>
<td>C</td>
<td>39086981</td>
<td>Exonic</td>
<td>SEC23A</td>
<td>0.3071</td>
<td>0.1184</td>
<td>0.010</td>
</tr>
<tr>
<td>15</td>
<td>rs261291</td>
<td>T</td>
<td>C</td>
<td>58387979</td>
<td>Intragenic</td>
<td>AQP9;LIPC</td>
<td>0.2842</td>
<td>0.1147</td>
<td>0.013</td>
</tr>
<tr>
<td>19</td>
<td>rs8133404</td>
<td>C</td>
<td>T</td>
<td>52562326</td>
<td>Intragenic</td>
<td>ZNF808;ZNF701</td>
<td>-0.7612</td>
<td>0.2487</td>
<td>0.002</td>
</tr>
<tr>
<td>20</td>
<td>rs17216707</td>
<td>T</td>
<td>C</td>
<td>54115823</td>
<td>Intragenic</td>
<td>BCAS1;CYP24A1</td>
<td>0.5312</td>
<td>0.2229</td>
<td>0.017</td>
</tr>
</tbody>
</table>

CHR, chromosome; SNP, single nucleotide polymorphism; Ref, reference allele; Alt, alternative allele; BP, base position; SE, standard error; ST6GALNA3, ST6 N-acetylgalactosaminidase alpha-2,6-sialyltransferase 3; CPS1, carbamoyl phosphate synthetase 1; 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.

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 amido-hydrolase domain-containing protein 1 (AMDH1) 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 DHC7 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 AG04, 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 (ACTE1P). The ACTE1P 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 (ACTE1P) 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 ACTE1P 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 experiments for the role of ACTE1P, 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 ACTE1P 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**


**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**

17. Malik S, Fu L, Juras DJ, Karmali M, Wong BY, Gozdzik A, et al. Common variants of the vitamin D binding protein gene...
40. Bouillon R. Genetic and racial differences in the vitamin D endocrine system. Endocrinol Metab Clin North Am 2017;