

Article

Risk Allele Frequency Analysis of Single-Nucleotide Polymorphisms for Vitamin D Concentrations in Different Ethnic Group

Byung-Woo Yoon ^{1,†} , Hyun-Tae Shin ^{2,†} and Je-Hyun Seo ^{3,*} 

¹ Division of Oncology, Department of Internal Medicine, Inje University Seoul Paik Hospital, Seoul 04551, Korea; bwyoon.md@gmail.com

² Department of Dermatology, Inha University School of Medicine, Incheon 22212, Korea; hyuntae.shin@inha.ac.kr

³ Veterans Medical Research Institute, Veterans Health Service Medical Center, Seoul 05368, Korea

* Correspondence: jazmin2@naver.com; Tel.: +82-2-2225-1445; Fax: +82-2-2225-3950

† These authors contributed equally to the work.

Abstract: The prevalence of vitamin D deficiency varies from 20.8% to 61.6% among populations of different ethnicities, suggesting the existence of a genetic component. The purpose of this study was to provide insights into the genetic causes of vitamin D concentration differences among individuals of diverse ancestry. We collected 320 single-nucleotide polymorphisms (SNPs) associated with vitamin D concentrations from a genome-wide association studies catalog. Their population-level allele frequencies were derived based on the 1000 Genomes Project and Korean Reference Genome Database. We used Fisher's exact tests to assess the significance of the enrichment or depletion of the effect allele at a given SNP in the database. In addition, we calculated the SNP-based genetic risk score (GRS) and performed correlation analysis with vitamin D concentration that included latitude. European, American, and South Asian populations showed similar heatmap patterns, whereas African, East Asian, and Korean populations had distinct ones. The GRS calculated from allele frequencies of vitamin D concentration was highest among Europeans, followed by East Asians and Africans. In addition, the difference in vitamin D concentration was highly correlated with genetic factors rather than latitude effects.

Keywords: Vitamin D deficiency; global prevalence; allele frequency; single nucleotide polymorphism; East-Asians; latitude



Citation: Yoon, B.-W.; Shin, H.-T.; Seo, J.-H. Risk Allele Frequency Analysis of Single-Nucleotide Polymorphisms for Vitamin D Concentrations in Different Ethnic Group. *Genes* **2021**, *12*, 1530. <https://doi.org/10.3390/genes12101530>

Academic Editor: Courtney C. Babbitt

Received: 3 September 2021

Accepted: 24 September 2021

Published: 28 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vitamin D, a fat-soluble vitamin, plays an essential role in bone mineralization and calcium homeostasis. Its deficiency is closely related to metabolic bone disease [1] and non-skeletal conditions, such as cardiovascular, infectious, and autoimmune diseases, as well as malignancies and diabetes [2–5]. Vitamin D is produced in the skin from 7-dehydrocholesterol by UV irradiation. Serum 25-hydroxy-vitamin D (25(OH)D₃), the major circulating biomarker of vitamin D status, is converted to active vitamin D, 1,25(OH)₂D, primarily in the kidney and, to a lesser extent, in the extra-renal tissue [6]. Serum vitamin D levels are strongly influenced by numerous factors, including age, obesity, skin color, dietary intake, exposure to ultraviolet B (UVB) sunlight, geographical latitude, and dietary supplements [7]. Studies have estimated that 1 billion people worldwide have vitamin D deficiency or insufficiency [1,8] which is a significant public health concern [7].

According to a global overview, the prevalence of vitamin D deficiency or the average vitamin D concentrations varies according to ethnicity. For instance, despite the serum 25(OH)D₃ cutoff point being set at 20 ng/mL for adults, 54% of patients with African ancestry (the average 25(OH)D₃ concentrations: 21.0 ± 10.4 ng/mL) fall below this level,

compared to only 18% of Europeans ancestry (29.2 ± 10.9 ng/mL) in the United States [9]. In addition, deficiency status even differed at similar latitudes, as it ranged from 18.6% (26.0 ± 7.04 ng/mL) among Norwegians to 60.3% (18.0 ng/mL) among the Finnish [9]. With regard to East Asians, 32.1–53.9% of adults (20.4–23.4 ng/mL) in China and 53.6% in Japan (22.4 ± 7.5 ng/mL) have 25(OH)D₃ concentrations of ≤ 20 ng/mL [9]. The deficiency prevalence or the average 25(OH)D₃ concentrations is similar in Korea, and according to the Korea National Health and Nutrition Examination Survey, 47.3% of males (21.2 ± 7.5 ng/mL) and 64.5% of females (18.2 ± 7.1 ng/mL) have this deficiency [10].

The average 25(OH)D₃ concentration differences among ethnic groups point to a possible genetic component. In recent years, genome-wide association studies (GWASs) have revealed that several significant loci, including *GC*, *NADSYN1/DHCR7*, *CYP2R1*, and *CYP24A1*, are associated with 25(OH)D₃ deficiency [11–17]. These loci have been reported to function in the metabolism of vitamin D by converting into active form in the skin, liver, and kidney [2,13–17]. Previous researchers have suggested using whole-genome sequencing data of healthy subjects to identify disease phenotypes [18]. By combining data from the GWAS catalog of the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) and data from whole-genome sequencing of healthy subjects, it may be possible to identify risk-modeling of single-nucleotide polymorphisms (SNPs) related to 25(OH)D₃ levels. Our study group has previously applied SNP-related models to two ophthalmic diseases and published significant research results [19,20]. There is, however, a need to identify the risk models based on genetic and real-world data of 25(OH)D₃ levels and from populations of different ethnic groups to analyze these associations. This study aimed to identify the genetic causes and allele frequency differences related to 25(OH)D₃ concentration among populations of diverse ancestry. Moreover, it aimed to compare the composite genetic risk scores using SNPs related to 25(OH)D₃ concentrations for different ethnic groups.

2. Materials and Methods

2.1. Ethical Considerations

This study was approved by the Institutional Review Board (IRB) of the Veterans Health Service Medical Center, Korea (IRB No. 2019-07-008 and IRB No.2020-01-053). In addition, the need for informed consent was waived due to the use of de-identified data.

2.2. Comparison of Vitamin D-Related SNPs among the Global Population and East Asia

The most commonly used cut-points for serum [25(OH)D₃] levels in adults are 11–19 ng/mL, and ≤ 10 ng/mL for, deficiency, and severe deficiency, respectively. However, we used the average vitamin D concentrations for each cohort instead of the prevalence of vitamin D deficiency for two reasons: first, the data among African and South Asians are limited [9]; second, the prevalence of vitamin D deficiency and average 25(OH)D₃ concentration are related. We searched the NHGRI-EBI GWAS catalog (<https://www.ebi.ac.uk/gwas/home>, 30 December 2020) for SNPs associated with vitamin D measurements (EFO 0004631). The catalog included 13 studies and 546 associations. After eliminating repetitive SNPs and removing data not found in the 1000 Genome Projects database, 320 SNPs from the GWAS catalog were used for the analysis of allele frequencies associated with vitamin D concentrations.

The details and advantages of our method have been described elsewhere [19–21]. In brief, the population-level allele frequencies of SNPs were derived from the 1000 Genomes Project phase 3 and the Korean Reference Genome Database (KRGDB) produced by the Korea National Institute of Health in 2016. The former surveyed genetic variations among 2504 individuals from 26 worldwide populations grouped into African, East Asian, European, South Asian, and American categories based on their geographical locations and ancestry [22]. These data were downloaded (<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>, last accessed: 15 January 2020). The variant coordinates were based on the human genome assembly GRCh37. The latter included data on 1722 individuals from

the Korean population since the 1000 Genomes Project did not have this information [23]. Data on the population frequencies of the SNPs were downloaded from the web-based database (<http://152.99.75.168:9090/KRGDB/menuPages/download.jsp/>, last accessed: 15 January 2020). In order to compare the distributions of risk alleles in the Korean population, individual genotyping results from the second phase of KRGDB were obtained from 1099 individuals from the National Human Resource Bank of Korea. After statistical analysis, we performed expression quantitative trait locus (eQTL) analysis for significant SNPs using the Genotype-Tissue Expression (GTEx) portal (<https://www.gtexportal.org/>, accessed on 30 December 2020) for the significance of vitamin D SNP enrichment. Gene and transcript expression on the GTEx portal are shown in the Transcripts Per Million (TPM) unit, calculated as

$$\text{Transcripts Per Million} = \frac{n_t}{l_t} \left(\sum_{k \in T} \frac{n_k}{l_k} \right)^{-1} \times 10^6$$

where “ n_t ” refers to the number of reads for transcript/gene, the normalized transcript/gene length, and “ T ” is the set of all transcripts or genes depending on whether the quantification is at the gene level. The normalized expression (norm expression) values were calculated with edgeR (<https://gtexportal.org/home/documentationPage-#staticTextAnalysisMethods>, accessed on 30 December 2020).

2.3. Calculation of Genetic Risk Scores Using SNPs Related to Vitamin D Concentration

To compare the composite genetic risk of vitamin D deficiency, we adopted the following equation provided by Mao et al. [21]:

$$\text{Genetic risk score} = \frac{\sum_{i=1}^I X_i}{2I} \quad (1)$$

where “ I ” refers to the number of vitamin D concentration-related SNPs, and “ X_i ” to the copies of risk alleles ($X_i \in \{0,1,2\}$) the i th SNP. Thus, if a person had two copies of the risk allele at each vitamin D concentration-related SNP, their risk score was set as 1. In contrast, if a person had no copies, their risk score was 0. A person with a composite genetic risk score (GRS) of 1 has the highest possible genetic risk of higher vitamin D concentration, whereas a person with a score of 0 has the lowest. If copies of effect alleles (0/1/2) were randomly assigned to each SNP, the expected value of the risk score was set at 0.5. SNPs with frequency differences of more than 10% between the total ($n = 1722$) and second-phase ($n = 1099$) data of KRGDB were excluded from the GRS calculation. We used the average composite GRS to determine correlations with population vitamin D concentration data from similar geographical latitudes (51°) and fitting curve vitamin D concentration for its original endogenous population vs. GRS [9]. In addition, the correlation analysis of the difference of vitamin D concentrations was performed included both latitude factor and GRS factor since a previous study showed the impact of these on vitamin D concentration in patients of African and European ancestry [24].

2.4. Statistical Analyses

We used Fisher’s exact test to assess whether the effect allele at a given SNP was significantly higher or lower compared to the global population frequency in the 1000 Genomes Project database, and the p values were initially $\log_{10}$ -transformed. In the heatmap generated to visualize allele patterns in different populations, red and blue colors were used to indicate higher and lower frequencies, respectively, compared to the global average. If the effect allele was enriched in a population, then the negative $\log_{10}$ of the p -value (a positive number) was used to represent the SNP associated with that population in the heatmap. In contrast, if it was depleted, then the $\log_{10}$ of the p -value (a negative number) was used. Statistical analyses were performed using R software version 4.0.1 (R Foundation, Vienna, Austria), and statistical significance was set at $p < 0.05$.

3. Results

3.1. Vitamin D Concentration-Related SNPs in the Global Population

We collected 320 vitamin D concentration-associated SNPs from 13 GWASs using the NHGRI-EBI catalog. We determined the effect allele frequencies (EAFs) for each of the continental groups and the Korean population based on the information from the 1000 Genomes Project and KRGDB (Supplementary Table S1). The heatmap shows how significantly the effect allele was enriched or depleted across these populations (Supplementary Figure S1). In the Korean population, 106 vitamin D-related SNPs were significantly enriched, 120 were depleted, and 94 were comparable to the global EAF. The hierarchical clustering tree showed the differences among the populations, with Europeans, Americans, and South Asians in one cluster and Africans, East Asians, and Koreans in another. In addition, SNPs with significantly different frequencies among the Korean population (Log-adjusted p -value of Fisher's exact test in Koreans >100 or <-100) are summarized in Table 1 and Figure 1.

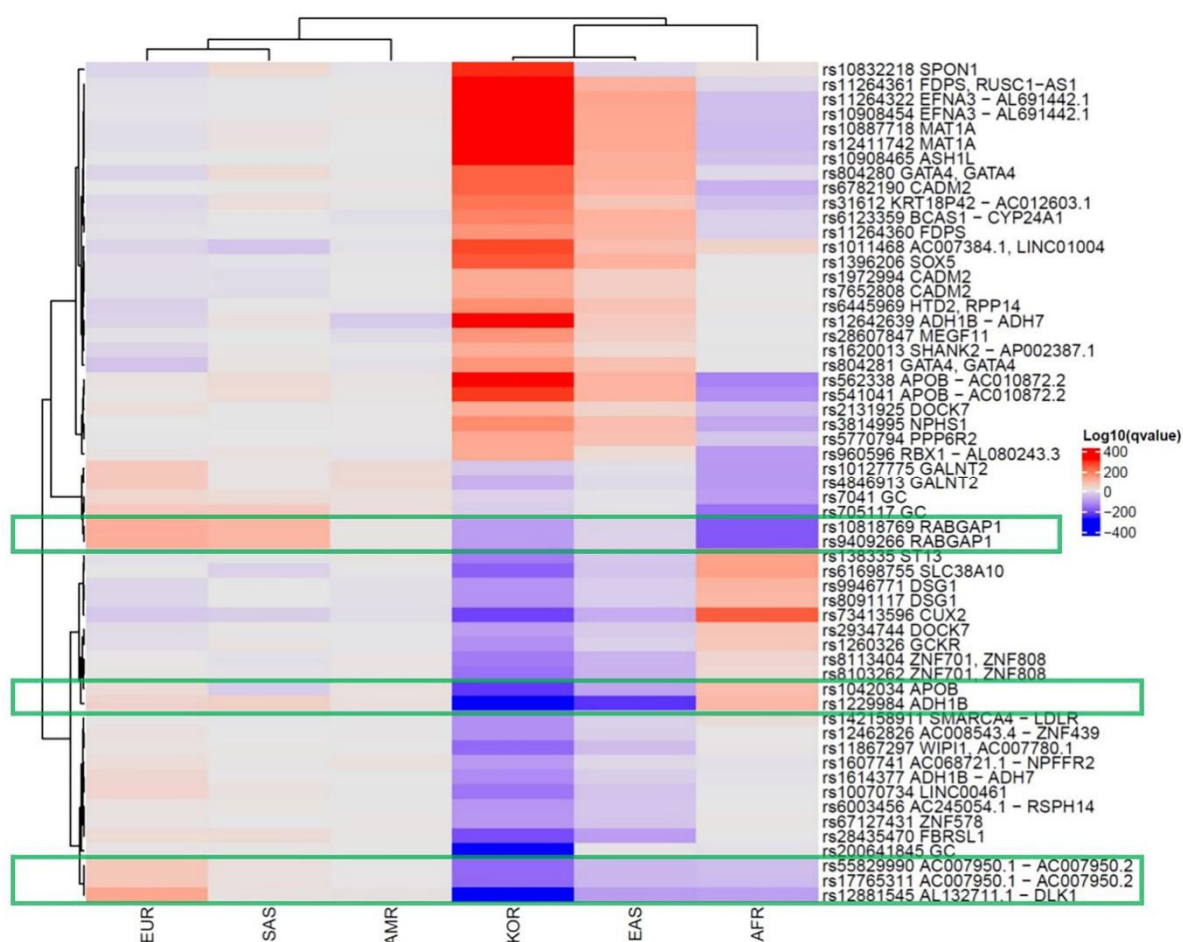


Figure 1. Significant single-nucleotide polymorphisms related to the vitamin D concentration in the global population. Each row shows an SNP, and each column is a population group. The red color indicates enrichment of the effect allele, whereas the purple color indicates the depletion. The following SNPs ((rs10818769 and rs9409266 (*RABGAP1*) in the red box, followed by rs12881545 (*DLK1*), rs10070734 (*LINC00461*), and rs17765311 (*AC007950.2*)) were highly depleted in East Asians (including Koreans) and Africans and were enriched in Europeans. (AMR: American, EUR: European, SAS: South Asian, AFR: African, EAS: East Asian, KOR: Korean.).

Table 1. Effect allele frequencies of vitamin D concentration-related single-nucleotide polymorphisms among continental groups.

SNP ID	Chr ^a	Position	MAPPED_GENE	Function	Ref Allele ^b	Alt Allele ^c	Global EAF ^d	AMR ^e EAF	AMR log10 P	AFR ^f EAF	AFR log10 P	EAS ^g EAF	EAS log10 P	SAS ^h EAF	SAS log10 P	EUR ⁱ EAF	EUR log10 P	KOR ^k EAF	KOR log10 P
rs2131925	chr1	63025942	DOCK7	intron_variant	G	T	0.56	0.61	1.61	0.31	−58.18	0.77	35.67	0.53	−0.97	0.69	13.25	0.785	102.29
rs10908454	chr1	155066416	EFNA3- AL691442.1	TF_binding_ site_variant	G	A	0.55	0.64	4.64	0.31	−53.75	0.91	116.71	0.51	−1.48	0.48	−3.99	0.919	312.61
rs11264322	chr1	155087933	EFNA3- AL691442.1	intergenic_variant	G	A	0.54	0.61	2.91	0.3	−53.97	0.91	122.04	0.5	−1.51	0.45	−6.28	0.919	312.61
rs11264360	chr1	155284586	FDPS	intron_variant	T	A	0.35	0.3	−1.79	0.2	−25.83	0.7	91.61	0.3	−2.46	0.27	−5.79	0.638	149.83
rs11264361	chr1	155289545	FDPS, RUSC1-AS1	non_coding_trans cript_exon_variant	T	G	0.35	0.3	−1.79	0.21	−22.25	0.71	96.83	0.31	−1.66	0.26	−7.22	0.801	312.61
rs10908465	chr1	155389688	ASH1L	intron_variant	C	T	0.34	0.31	−0.75	0.14	−48.92	0.71	102.27	0.31	−1.04	0.28	−3.40	0.811	312.61
rs562338	chr2	21288321	APOB- AC010872.2	intergenic_variant	A	G	0.73	0.82	6.07	0.35	−138.10	0.98	91.24	0.87	21.07	0.8	5.29	0.995	311.61
rs541041	chr2	21294975	APOB- AC010872.2	intergenic_variant	G	A	0.75	0.82	3.96	0.4	−119.94	0.99	93.79	0.88	19.24	0.8	3.04	0.995	282.11
rs6782190	chr3	85639672	CADM2	intron_variant	G	A	0.65	0.73	4.16	0.37	−73.11	0.94	91.71	0.64	−0.23	0.65	0.00	0.936	231.64
rs200641845	chr4	72620895	GC	intron_variant	T	A	0.42	0.43	0.14	0.34	−6.80	0.45	1.03	0.47	2.24	0.45	1.02	0.080	−285.57
rs1607741	chr4	72719033	AC068721.1- NPFFR2	intergenic_variant	G	C	0.58	0.72	11.29	0.49	−8.10	0.42	−19.40	0.61	0.98	0.73	18.02	0.338	−106.15
rs1614377	chr4	100279332	ADH1B-ADH7	intergenic_variant	G	A	0.15	0.17	0.61	0.095	−6.69	0.026	−33.64	0.18	1.55	0.31	28.74	0.012	−125.88
rs12642639	chr4	100301241	ADH1B-ADH7	intergenic_variant	C	A	0.38	0.14	−37.19	0.34	−2.07	0.65	54.04	0.51	12.56	0.21	−24.90	0.825	312.61
rs10070734	chr5	87940026	LINC00461	intron_variant	T	C	0.5	0.46	−1.12	0.46	−1.92	0.25	−48.15	0.61	9.11	0.71	33.25	0.217	−155.96
rs31612	chr5	108996643	KRT18P42- AC012603.1	intergenic_variant	T	C	0.32	0.28	−1.27	0.12	−51.58	0.61	63.57	0.46	15.17	0.18	−18.77	0.654	201.59
rs804280	chr8	11612698	GATA4, GATA4	intron_variant	C	A	0.73	0.69	−1.31	0.6	−18.48	0.99	103.54	0.87	21.07	0.56	−24.11	0.975	231.56
rs10818769	chr9	125719923	RABGAP1	intron_variant	C	G	0.49	0.58	4.58	0.071	−197.61	0.31	−24.91	0.83	89.25	0.85	104.26	0.260	−100.85
rs9409266	chr9	125745042	RABGAP1	intron_variant	G	A	0.49	0.58	4.58	0.07	−198.46	0.31	−24.91	0.83	89.25	0.85	104.26	0.260	−101.09
rs10887718	chr10	82042624	MAT1A	intron_variant	C	T	0.66	0.67	0.16	0.41	−58.77	0.97	114.96	0.76	8.79	0.54	−11.52	0.979	312.61
rs12411742	chr10	82042782	MAT1A	intron_variant	G	A	0.66	0.67	0.16	0.41	−58.77	0.97	114.96	0.76	8.79	0.54	−11.52	0.979	312.61
rs1620013	chr11	71089210	SHANK2- AP002387.1	intergenic_variant	C	T	0.48	0.45	−0.72	0.45	−1.21	0.66	24.42	0.53	2.16	0.3	−24.67	0.717	104.75
rs1396206	chr12	24576859	SOX5	intron_variant	A	T	0.7	0.67	−0.78	0.67	−1.37	0.97	94.88	0.61	−6.85	0.58	−12.07	0.967	245.87
rs12881545	chr14	101176212	AL132711.1- DLK1	TF_binding_ site_variant	G	C	0.26	0.38	9.25	0.029	−97.30	0.0069	−103.17	0.35	7.30	0.65	116.98	0.000	−286.57
rs17765311	chr15	63789952	AC007950.1- AC007950.2	regulatory_region_ variant	A	C	0.16	0.2	1.80	0.016	−58.56	0.001	−65.52	0.26	11.55	0.4	57.06	0.003	−174.20
rs55829990	chr15	63790642	AC007950.1- AC007950.2	intergenic_variant	T	C	0.16	0.2	1.80	0.019	−54.99	0.002	−63.59	0.26	11.55	0.4	57.06	0.003	−174.20
rs28607847	chr15	66284913	MEGF11	intron_variant	G	A	0.38	0.22	−15.36	0.36	−0.69	0.62	42.80	0.38	0.00	0.29	−6.89	0.668	149.95
rs3814995	chr19	36342212	NPHS1	missense_variant	C	T	0.29	0.34	1.85	0.056	−84.87	0.6	74.07	0.24	−2.58	0.31	0.62	0.587	162.75
rs6123359	chr20	52714706	BCAS1- CYP24A1	regulatory_ region_variant	A	G	0.21	0.1	−11.63	0.083	−28.70	0.55	96.31	0.22	0.27	0.11	−13.14	0.511	180.25
rs960596	chr22	41393520	RBX1- AL080243.3	intergenic_variant	C	T	0.27	0.2	−3.71	0.026	−107.11	0.43	21.65	0.38	10.42	0.35	5.99	0.508	108.51

p-value: adjusted Fischer's test, statistical significance was set at $p < 0.05$. a: chromosome. b: reference allele. c: alternative allele. d: effect allele frequency. e: Americans. f: Africans. g: East Asians. h: South Asians. i: Europeans. k: Koreans.

From the data, rs10818769 and rs9409266 were found to be depleted in Koreans, East Asians, and Africans but enriched in Europeans. The SNP (rs10818769, rs9409266) is located in an intronic region of the *RABGAP1* gene, which encodes guanosine triphosphatase-activating protein of RAB6A, and has alleles of C > G and G > A, respectively. The major allele was detected in 85% of Europeans and 26% of Koreans. Although the *RABGAP1* gene is known to be related to body height and birth weight, these SNPs may be related to modulation of the evolution-related gene for the ethnic component of vitamin D concentration. The box plots of eQTL of the *RABGAP1* genes related to vitamin D in skin tissues of both sun-exposed and non-exposed areas show a significantly different expression, according to the alleles of rs10818769 and rs9409266 in the GTEx data (Figure 2). Comparison of allele frequency of major vitamin D-related genes, such as *GC*, *NADSYN1/DHCR7*, *CYP2R1*, and *CYP24A1*, are summarized in Supplementary Table S2.

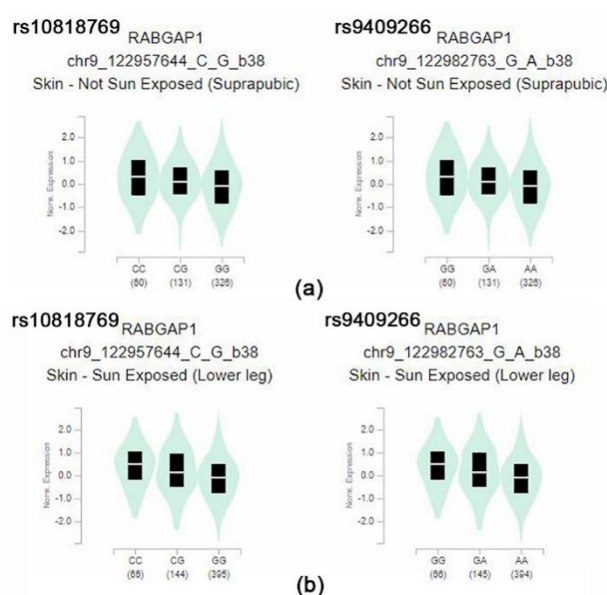


Figure 2. (a) Box plot of eQTL variant results ($p = 3.4 \times 10^{-12}$, $p = 3.4 \times 10^{-12}$, respectively): rs10818769-skin (suprapubic), rs9409266-skin (suprapubic) (b) box plot of eQTL variant results ($p = 6.4 \times 10^{-12}$, $p = 6.0 \times 10^{-12}$, respectively): rs10818769-skin (lower leg), rs9409266-skin (lower leg) These variants showed significantly different expression eQTL in their minor allele of the *RABGAP1*. Normalized expression (norm expression) was calculated in Transcript Per Million (TPM) units.

3.2. Genetic Risk Scores Calculated Using SNPs Related to Vitamin D Levels

We calculated the composite GRS based on the number of copies of effect alleles at the 320 vitamin D-associated SNPs, assuming that allelic associations from most GWAS-identified variants could be replicated in non-European populations. The GRS of vitamin D concentration was highest among Europeans, followed by Americans, South Asians, East Asians, and Africans (Figure 3).

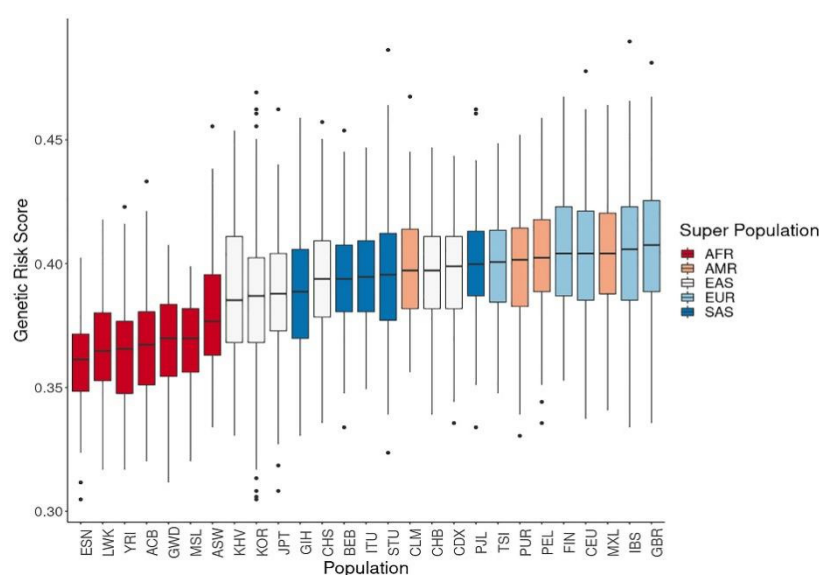


Figure 3. Genetic risk score calculations of vitamin D concentration using related single-nucleotide polymorphisms. The genetic risk scores calculated from allele frequencies for vitamin D concentration were highest in Europeans, followed by East Asians and Africans. (ACB: African Caribbean in Barbados; ASW: African ancestry in the Southwest USA; BEB: Bengali in Bangladesh; CDX: Chinese Dai in Xishuangbanna; CEU: Utah residents with Northern and Western European ancestry; CHB: Han Chinese in Beijing, China; CHS: Southern Han Chinese, China; CLM: Colombian in Medellin, Colombia; ESN: Esan in Nigeria; FIN: Finnish in Finland; GBR: British in England and Scotland; GIH: Gujarati Indian in Houston, TX, USA; GWD: Gambian in Western Division, Gambia; IBS: Iberian populations in Spain; ITU: Indian Telugu in the UK; JPT: Japanese in Tokyo, Japan; KOR: Korean in the Republic of Korea; KHV: Kinh in Ho Chi Minh City, Vietnam; LWK: Luhya in Webuye, Kenya; MSL: Mende in Sierra Leone; MXL: Mexican ancestry in Los Angeles, CA, USA; PEL: Peruvian in Lima, Peru; PUL: Punjabi in Lahore, Pakistan; PUR: Puerto Rican in Puerto Rico; STU: Sri Lankan Tamil in the UK; TSI: Toscani in Italy; YRI: Yoruba in Ibadan, Nigeria.).

A strong correlation was observed between the vitamin D concentration from several studies [25–30] and GRS with a similar geographic latitude (51° , $R^2 = 0.59$) in the grey dashed line (Figure 4). In addition, the vitamin D concentration for its original endogenous population vs. GRS fitted to the U curves of the black line (Figure 4). Correlation plot of the difference between vitamin D concentration and GRS using related SNPs or latitude Vitamin D concentration was strongly correlated with average GRS rather than latitude effect, when reviewing the vitamin D difference vs. GRS with R^2 value of 0.9996 (A), instead of latitude difference with an R^2 value of 0.6438 (B) in Figure 5.

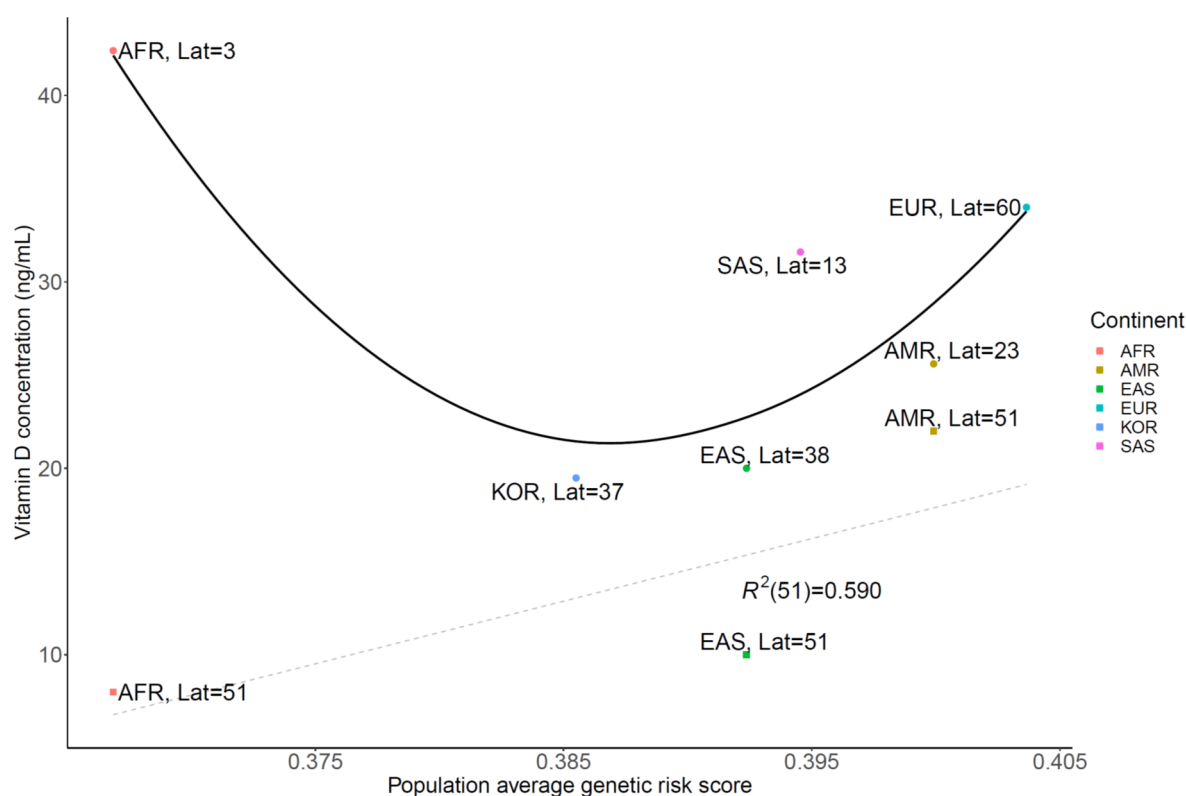


Figure 4. Correlation plot of vitamin D concentration and genetic risk score using related single-nucleotide polymorphisms, with latitude. Vitamin D concentration was strongly correlated with average genetic risk score at the same latitude ($R^2 = 0.590$) in the grey dashed line. Its original endogenous population vitamin D concentration vs genetic risk score fitted in the black line. (Lat: Latitude, AMR: American, EUR: European, SAS: South Asian, AFR: African, EAS: East Asian).

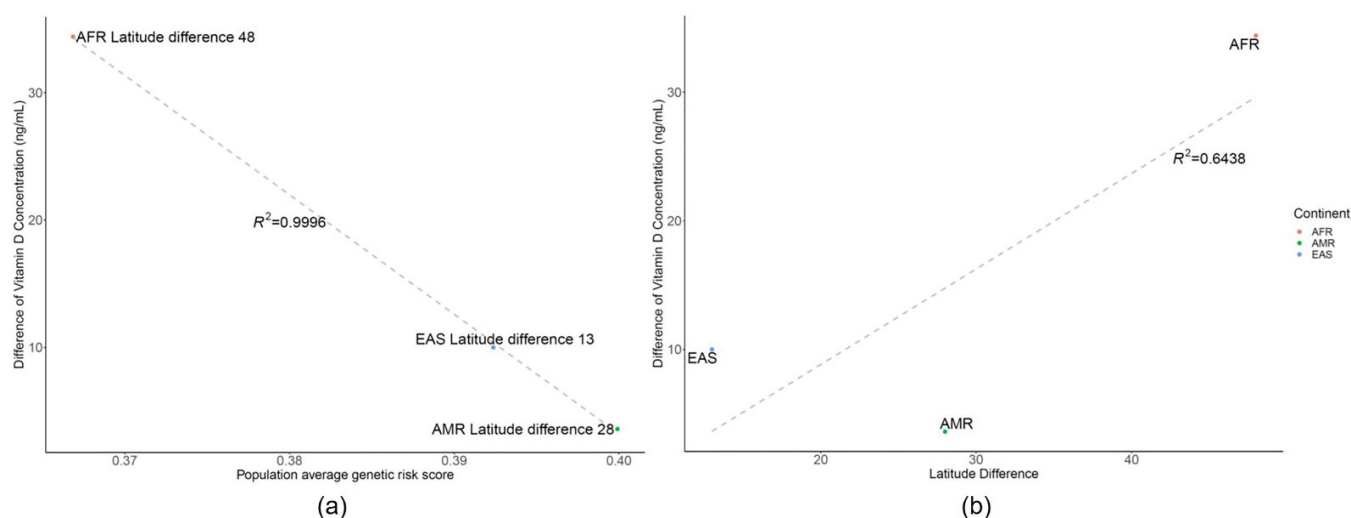


Figure 5. Correlation plot of the difference between vitamin D concentration and genetic risk score using related single-nucleotide polymorphisms or latitude. (a) The difference between vitamin D and genetic risk score, showing that a lower genetic risk score cannot compensate or generate vitamin D when this population is relocated to a higher latitude, where the regression R^2 is 0.9996. (b) Difference between vitamin D and latitude, showing that latitude difference does not impact the difference of vitamin D as much as genetic risk score, with an R^2 value of 0.6438. (AFR: African, AMR: American, EAS: East Asian).

Vitamin D concentration was strongly correlated with average genetic risk score rather than latitude effect when reviewing the vitamin D difference versus genetic risk score with an R^2 value of 0.9996, instead of latitude difference with an R^2 value of 0.6438.

4. Discussion

Vitamin D deficiency is associated with unfavorable bone conditions and chronic diseases such as cancer and diabetes [31]. Thus, in this study, we aimed to assess the different risk alleles of ethnic groups that may reflect vitamin D concentrations. We found that allele frequencies were found to differ dependent on ethnic group, and the SNP-based genetic score was shown to have a strong correlation with real-world data of vitamin D levels.

Previously conducted GWASs revealed several significant loci, including *GC*, *NADSYN1/DHCR7*, *CYP2R1*, and *CYP24A1* [11–17], that played an important role in vitamin D concentrations. Subsequently, using these significant loci (46 SNPs), Jones et al. showed variations of vitamin D levels among European, East Asian, and African populations by UVB exposure and ancestry [24]. Our study hypothesized that different allele frequencies of ethnic groups and Koreans might have significant loci for the evolution of different vitamin D concentrations regardless of environmental factors. In our study, the rs200641845 and rs7041 related *GC*, encoding vitamin D binding carrier protein, were highly depleted in Koreans and Africans. The rs3829251 and rs11233933 associated with *NADSYN1* were depleted in Africans whereas they were enriched in Koreans. This could be one piece of the evidence in relation to Koreans and Africans having a different mechanism related to low vitamin D levels, as *NADSYN1/DHCR7* is involved in UVB-induced vitamin D metabolism in the skin.

Additionally, we found some SNPs [rs10818769 (*RABGAP1*), rs9409266 (*RABGAP1*), rs12881545 (*DLK1*), rs10070734 (*LINC00461*), and rs17765311 (*AC007950.2*)] that were highly underexpressed in East Asians (including Koreans) and Africans, while they were overexpressed in Europeans. The eQTL analysis showed that rs10818769 and rs9409266 affected *RABGAP1* expression in the skin regardless of sun exposure. The pigmentation-associated allele evolution has been shown to include *SLC24A5* [32] and *RABGAP1* in a previous study [33], and *RABGAP1* was the signature gene for vitamin D deficiency and skin pigmentation. It was found to be underexpressed among East Asians (including Koreans) and Africans, while highly expressed among Europeans in our study. This gene may provide a possible link between skin pigmentation and vitamin D concentration; however, further experimental studies are needed to confirm this. This result is consistent with the nutrigenomics of vitamin D in that the main evolutionary driver of decreased skin pigmentation was the need for sufficient endogenous vitamin D production [34]. Skin color and genetic variation may explain vitamin D deficiency and adaptation to life in the latitudes [35].

The GRS was the highest among Europeans, followed by Americans, South Asians, East Asians, and Africans, and was correlated with vitamin D concentrations. This result is consistent with the estimates of 25(OH)D₃ levels <20 ng/mL that have been reported as 24% in the USA, 37% in Canada, and 40% in Europe [36,37]. European Caucasians have been shown to have lower rates of vitamin D deficiency compared with nonwhite individuals [36,38]. In addition to genetics, environmental factors such as nutrition and sunlight exposure are important determinants of vitamin D concentration, and latitude was one of the factors considered in this study. Moreover, vitamin D deficiency is common in non-western immigrants due to low sunshine exposure, pigment skin, and low calcium intake [39]. In this regard, the comparison of multi-ethnic group data in a single country would be desirable. According to the study by Van der Meer et al., the mean vitamin D concentration was 26.8 ng/mL among the Dutch and 13.2 ng/mL among Africans in the Netherlands at a latitude of 51° [40]. These results are consistent with those of another study showing that African Americans had lower levels of vitamin D than European Americans [41]. The pooled prevalence of low vitamin D status in Africa was 33.22% (26.22–43.68%) with a cutoff of serum 25(OH)D₃ concentration of <20 ng/mL and an overall mean of 26.8 ng/mL [42]. Furthermore, vitamin D concentrations were strongly correlated with GRS rather than latitude effects when examining GRS vs. vitamin D concentration differences instead of latitudinal differences. Thus, latitude factors should be

considered for vitamin D concentration assessment in genetic models, as was performed in our study.

A major strength of our study was the inclusion of the Korean whole-genome dataset of 1722 individuals that reflected the allele frequency of SNPs related to vitamin D deficiency. Moreover, we computed the risk model using a significant number of alleles ($n = 320$) related to vitamin D compared to a previous study that used only the major loci [24]. Additionally, we did not systematically organize the new vitamin D cohort and analyze the effect; instead, we compared the data from the 1000 Genomes Project with the vitamin D-related SNP data from the GWAS catalog. Despite these strengths, there are some limitations to this study. First, the GWAS catalog contained data where the risk allele was not clearly defined according to the minor allele frequency (MAF). We did not exclude these from our study because the majority of MAFs were likely to be risk alleles. Therefore, inaccurate subgroup analysis could have arisen. To address this issue, risk allele curation is necessary for the GWAS catalog based on the results of additional large population studies using cohorts in whom vitamin D was measured. Moreover, the statistical significance of EAF in the Korean population was high and should be interpreted with caution since Fisher's test can decrease the p -value as the number of subjects increases, even with the same odds ratios. Third, latitude and genomic modeling were only used for vitamin D analysis; other environmental factors (such as nutrition) were not considered. A previous study on a multi-ethnic population of Norway has shown that there are many modifiable risk factors related to 25(OH)D₃ levels [43]. Finally, we used the composite GRS instead of polygenic risk score for two reasons: the weighted-odd ratios of vitamin D concentrations varied according to the ethnic group even for the same SNP, and as there were inaccuracies of weighted-odd ratios due to insufficient study data among the African and American populations. In the future, a polygenic risk score with the effect size-weighted odd ratio should be evaluated.

5. Conclusions

Our study found a substantial population difference in terms of allele frequencies in vitamin D-related SNPs. The GRS for vitamin D concentrations was higher in Europeans compared to that found in East Asians and Africans, which were highly correlated with actual data. In addition, vitamin D concentration was strongly correlated with average GRS rather than latitude effect. From the public health perspective of vitamin D deficiency, genetic variants associated with vitamin D, as well as environmental factors (latitude, UVB exposures), should be considered. Further studies are needed to identify variant SNPs in genes such as *RABGAP1* (rs10818769, rs9409266) that reflect vitamin D deficiency in East Asians or Africans and to assess their modifiable roles for evolutionary differences.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12101530/s1>, Figure S1: Entire differences of single-nucleotide polymorphisms related to the vitamin D concentration in the global population, Table S1: Effect allele frequencies (EAFs) of vitamin D concentrations related single nucleotide polymorphisms in continental groups, including Koreans, Table S2: Effect allele frequencies (EAFs) of vitamin D concentrations related single nucleotide polymorphisms in continental groups, including Koreans.

Author Contributions: Conceptualization, B.-W.Y., H.-T.S. and J.-H.S.; methodology, B.-W.Y. and H.-T.S.; formal analysis, B.-W.Y., H.-T.S. and J.-H.S.; investigation, B.-W.Y., H.-T.S. and J.-H.S.; resources, H.-T.S.; data curation, B.-W.Y., H.-T.S. and J.-H.S.; writing—original draft preparation, B.-W.Y., H.-T.S. and J.-H.S.; writing—review and editing, B.-W.Y., H.-T.S. and J.-H.S.; visualization, B.-W.Y.; supervision, H.-T.S. and J.-H.S.; funding acquisition, J.-H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Veterans Health Service Medical Center Research Grant (grant no.: VHSMC20026).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Veterans Health Service Medical Center, Korea (IRB No. 2019-07-008 and IRB No. 2020-01-053).

Informed Consent Statement: Patient consent was waived due to retrospective data analysis and de-identify data, Institutional Review Board of Veterans Health Service Medical Center approved waiver of informed consent.

Data Availability Statement: The raw datasets generated and analyzed during the current study are not publicly available since any data providing the whole-genome sequencing data is considered to be personal property by the Korea Bioethics law. However, the raw whole-genome sequencing data for research are available at the reasonable request under the permission of the National Biobank of Korea contact at [[http://nih.go.kr/biobank/cmm/main/mainPage.do/?/](http://nih.go.kr/biobank/cmm/main/mainPage.do?/), accessed on 15 January 2020] and e-mail [biobank@korea.kr]. The allele frequency of Korea reference genome data base (KRGDB) is available [<http://152.99.75.168:9090/KRGDBDN/dnKRGinput.jsp>, accessed on 15 January 2020], files required are all three of ‘the totally merged sets’ of common variants, rare variants, and indels. The 1000genomes data is available, all the files from the following folder were downloaded, [<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>] (last accessed: 15 January 2020). The genome-wide association study (GWAS) catalog data is available in the (NHGRI-EBI, [<https://www.ebi.ac.uk/gwas/docs/file-downloads>, accessed on 15 January 2020], “All associations v1.0.2—with added ontology annotations, GWAS Catalog study accession numbers and genotyping technology”, December 2020).

Acknowledgments: This study was conducted with bioresources from the National Biobank of Korea, the Center for Disease Control and Prevention, Republic of Korea (KBN-2019-053). 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 NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 18 July 2021.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Holick, M.F. Vitamin D deficiency. *N. Engl. J. Med.* **2007**, *357*, 266–281. [[CrossRef](#)]
- Altieri, B.; Muscogiuri, G.; Barrea, L.; Mathieu, C.; Vallone, C.V.; Mascitelli, L.; Bizzaro, G.; Altieri, V.M.; Tirabassi, G.; Balercia, G.; et al. Does vitamin D play a role in autoimmune endocrine disorders? A proof of concept. *Rev. Endocr. Metab. Disord.* **2017**, *18*, 335–346. [[CrossRef](#)] [[PubMed](#)]
- Skaaby, T.; Thuesen, B.H.; Linneberg, A. Vitamin D, Cardiovascular Disease and Risk Factors. *Adv. Exp. Med. Biol.* **2017**, *996*, 221–230. [[CrossRef](#)]
- Lee, C.J.; Iyer, G.; Liu, Y.; Kalyani, R.R.; Bamba, N.; Ligon, C.B.; Varma, S.; Mathioudakis, N. 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 Complicat.* **2017**, *31*, 1115–1126. [[CrossRef](#)] [[PubMed](#)]
- Gois, P.H.F.; Ferreira, D.; Olenski, S.; Seguro, A.C. Vitamin D and Infectious Diseases: Simple Bystander or Contributing Factor? *Nutrients* **2017**, *9*, 651. [[CrossRef](#)] [[PubMed](#)]
- Bikle, D.D. Vitamin D metabolism, mechanism of action, and clinical applications. *Chem. Biol.* **2014**, *21*, 319–329. [[CrossRef](#)]
- Makariou, S.; Liberopoulos, E.N.; Elisaf, M.; Challa, A. Novel roles of vitamin D in disease: What is new in 2011? *Eur. J. Intern. Med.* **2011**, *22*, 355–362. [[CrossRef](#)]
- Bischoff-Ferrari, H.A.; Giovannucci, E.; Willett, W.C.; 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. [[CrossRef](#)]
- van Schoor, N.; Lips, P. Global Overview of Vitamin D Status. *Endocrinol. Metab. Clin. N. A.* **2017**, *46*, 845–870. [[CrossRef](#)]
- Choi, H.S.; Oh, H.J.; Choi, H.; Choi, W.H.; Kim, J.G.; Kim, K.M.; Kim, K.J.; Rhee, Y.; Lim, S.K. 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*, 643–651. [[CrossRef](#)]
- Ahn, J.; Yu, K.; Stolzenberg-Solomon, R.; Simon, K.C.; McCullough, M.L.; Gallicchio, L.; Jacobs, E.J.; Ascherio, A.; Helzlsouer, K.; Jacobs, K.B.; et al. Genome-wide association study of circulating vitamin D levels. *Hum. Mol. Genet.* **2010**, *19*, 2739–2745. [[CrossRef](#)] [[PubMed](#)]
- Wang, T.J.; Zhang, F.; Richards, J.B.; Kestenbaum, B.; van Meurs, J.B.; Berry, D.; Kiel, D.P.; Streeten, E.A.; Ohlsson, C.; Koller, D.L.; et al. Common genetic determinants of vitamin D insufficiency: A genome-wide association study. *Lancet* **2010**, *376*, 180–188. [[CrossRef](#)]

13. Malik, S.; Fu, L.; Juras, D.J.; Karmali, M.; Wong, B.Y.; Gozdzik, A.; Cole, D.E. Common variants of the vitamin D binding protein gene and adverse health outcomes. *Crit. Rev. Clin. Lab. Sci.* **2013**, *50*, 1–22. [[CrossRef](#)] [[PubMed](#)]
14. Anderson, D.; Holt, B.J.; Pennell, C.E.; Holt, P.G.; Hart, P.H.; Blackwell, J.M. Genome-wide association study of vitamin D levels in children: Replication in the Western Australian Pregnancy Cohort (Raine) study. *Genes Immun.* **2014**, *15*, 578–583. [[CrossRef](#)] [[PubMed](#)]
15. Moy, K.A.; Mondul, A.M.; Zhang, H.; Weinstein, S.J.; Wheeler, W.; Chung, C.C.; Mannisto, S.; Yu, K.; Chanock, S.J.; Albanes, D. Genome-wide association study of circulating vitamin D-binding protein. *Am. J. Clin. Nutr.* **2014**, *99*, 1424–1431. [[CrossRef](#)] [[PubMed](#)]
16. Sapkota, B.R.; Hopkins, R.; Bjornnes, A.; Ralhan, S.; Wander, G.S.; Mehra, N.K.; Singh, J.R.; Blackett, P.R.; Saxena, R.; Sanghera, D.K. 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*, 149–156. [[CrossRef](#)]
17. Wang, J.; Thingholm, L.B.; Skieceviciene, J.; Rausch, P.; Kummert, M.; Hov, J.R.; Degenhardt, F.; Heinsen, F.A.; Ruhlemann, M.C.; Szymczak, S.; et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat. Genet.* **2016**, *48*, 1396–1406. [[CrossRef](#)]
18. Pinese, M.; Lacaze, P.; Rath, E.M.; Stone, A.; Brion, M.J.; Ameur, A.; Nagpal, S.; Puttick, C.; Husson, S.; Degraeve, D.; et al. The Medical Genome Reference Bank contains whole genome and phenotype data of 2570 healthy elderly. *Nat. Commun.* **2020**, *11*, 435. [[CrossRef](#)]
19. Shin, H.T.; Yoon, B.W.; Seo, J.H. Comparison of risk allele frequencies of single nucleotide polymorphisms associated with age-related macular degeneration in different ethnic groups. *BMC Ophthalmol.* **2021**, *21*, 97. [[CrossRef](#)]
20. Shin, H.T.; Yoon, B.W.; Seo, J.H. Analysis of risk allele frequencies of single nucleotide polymorphisms related to open-angle glaucoma in different ethnic groups. *BMC Med. Genom.* **2021**, *14*, 80. [[CrossRef](#)]
21. Mao, L.; Fang, Y.; Campbell, M.; Southerland, W.M. Population differentiation in allele frequencies of obesity-associated SNPs. *BMC Genom.* **2017**, *18*, 861. [[CrossRef](#)]
22. Consortium, G.P.; Auton, A.; Brooks, L.D.; Durbin, R.M.; Garrison, E.P.; Kang, H.M.; Korbel, J.O.; Marchini, J.L.; McCarthy, S.; McVean, G.A.; et al. A global reference for human genetic variation. *Nature* **2015**, *526*, 68–74. [[CrossRef](#)]
23. Kim, J.; Weber, J.A.; Jho, S.; Jang, J.; Jun, J.; Cho, Y.S.; Kim, H.M.; Kim, H.; Kim, Y.; Chung, O.; et al. KoVariome: Korean National Standard Reference Variome database of whole genomes with comprehensive SNV, indel, CNV, and SV analyses. *Sci. Rep.* **2018**, *8*, 5677. [[CrossRef](#)] [[PubMed](#)]
24. Jones, P.; Lucock, M.; Chaplin, G.; Jablonski, N.G.; Veysey, M.; Scarlett, C.; Beckett, E. Distribution of variants in multiple vitamin D-related loci (DHCR7/NADSYN1, GC, CYP2R1, CYP11A1, CYP24A1, VDR, RXRalpha and RXRgamma) vary between European, East-Asian and Sub-Saharan African-ancestry populations. *Genes Nutr.* **2020**, *15*, 5. [[CrossRef](#)] [[PubMed](#)]
25. Choi, H.S. Vitamin d status in Korea. *Endocrinol. Metab.* **2013**, *28*, 12–16. [[CrossRef](#)] [[PubMed](#)]
26. Eloi, M.; Horvath, D.V.; Szejnfeld, V.L.; Ortega, J.C.; Rocha, D.A.; Szejnfeld, J.; Castro, C.H. Vitamin D deficiency and seasonal variation over the years in Sao Paulo, Brazil. *Osteoporos. Int.* **2016**, *27*, 3449–3456. [[CrossRef](#)]
27. Cashman, K.D.; Dowling, K.G.; Skrabakova, Z.; Kiely, M.; Lamberg-Allardt, C.; Durazo-Arvizu, R.A.; Sempos, C.T.; Koskinen, S.; Lundqvist, A.; Sundvall, J.; et al. Standardizing serum 25-hydroxyvitamin D data from four Nordic population samples using the Vitamin D Standardization Program protocols: Shedding new light on vitamin D status in Nordic individuals. *Scand. J. Clin. Lab. Investig.* **2015**, *75*, 549–561. [[CrossRef](#)]
28. Nakamura, K.; Kitamura, K.; Takachi, R.; Saito, T.; Kobayashi, R.; Oshiki, R.; Watanabe, Y.; Tsugane, S.; Sasaki, A.; Yamazaki, O. Impact of demographic, environmental, and lifestyle factors on vitamin D sufficiency in 9084 Japanese adults. *Bone* **2015**, *74*, 10–17. [[CrossRef](#)]
29. Chailurkit, L.O.; Aekplakorn, W.; Ongphiphadhanakul, B. Regional variation and determinants of vitamin D status in sunshine-abundant Thailand. *BMC Public Health* **2011**, *11*, 853. [[CrossRef](#)]
30. Luxwolda, M.F.; Kuipers, R.S.; Kema, I.P.; van der Veer, E.; Dijck-Brouwer, D.A.; Muskiet, F.A. Vitamin D status indicators in indigenous populations in East Africa. *Eur. J. Nutr.* **2013**, *52*, 1115–1125. [[CrossRef](#)]
31. Amrein, K.; Scherkl, M.; Hoffmann, M.; Neuwersch-Sommeregger, S.; Kostenberger, M.; Tmava Berisha, A.; Martucci, G.; Pilz, S.; Malle, O. Vitamin D deficiency 2.0: An update on the current status worldwide. *Eur. J. Clin. Nutr.* **2020**, *74*, 1498–1513. [[CrossRef](#)] [[PubMed](#)]
32. Batai, K.; Cui, Z.; Arora, A.; Shah-Williams, E.; Hernandez, W.; Ruden, M.; Hollowell, C.M.P.; Hooker, S.E.; Bathina, M.; Murphy, A.B.; et al. Genetic loci associated with skin pigmentation in African Americans and their effects on vitamin D deficiency. *PLoS Genet.* **2021**, *17*, e1009319. [[CrossRef](#)]
33. Ju, D.; Mathieson, I. The evolution of skin pigmentation-associated variation in West Eurasia. *Proc. Natl. Acad. Sci. USA* **2021**, *118*. [[CrossRef](#)] [[PubMed](#)]
34. Carlberg, C. Nutrigenomics of Vitamin D. *Nutrients* **2019**, *11*, 676. [[CrossRef](#)] [[PubMed](#)]
35. Hanel, A.; Carlberg, C. Skin colour and vitamin D: An update. *Exp. Dermatol.* **2020**, *29*, 864–875. [[CrossRef](#)]
36. Cashman, K.D. Vitamin D Deficiency: Defining, Prevalence, Causes, and Strategies of Addressing. *Calcif. Tissue Int.* **2020**, *106*, 14–29. [[CrossRef](#)]

37. Sarafin, K.; Durazo-Arvizu, R.; Tian, L.; Phinney, K.W.; Tai, S.; Camara, J.E.; Merkel, J.; Green, E.; Sempos, C.T.; Brooks, S.P. Standardizing 25-hydroxyvitamin D values from the Canadian Health Measures Survey. *Am. J. Clin. Nutr.* **2015**, *102*, 1044–1050. [[CrossRef](#)] [[PubMed](#)]
38. Cashman, K.D.; Dowling, K.G.; Skrabakova, Z.; Gonzalez-Gross, M.; Valtuena, J.; De Henauw, S.; Moreno, L.; Damsgaard, C.T.; Michaelsen, K.F.; Molgaard, C.; et al. Vitamin D deficiency in Europe: Pandemic? *Am. J. Clin. Nutr.* **2016**, *103*, 1033–1044. [[CrossRef](#)]
39. Lips, P.; de Jongh, R.T. Vitamin D deficiency in immigrants. *Bone Rep.* **2018**, *9*, 37–41. [[CrossRef](#)]
40. van der Meer, I.M.; Middelkoop, B.J.; Boeke, A.J.; Lips, P. Prevalence of vitamin D deficiency among Turkish, Moroccan, Indian and sub-Sahara African populations in Europe and their countries of origin: An overview. *Osteoporos. Int.* **2011**, *22*, 1009–1021. [[CrossRef](#)]
41. Batai, K.; Murphy, A.B.; Shah, E.; Ruden, M.; Newsome, J.; Agate, S.; Dixon, M.A.; Chen, H.Y.; Deane, L.A.; Hollowell, C.M.; et al. Common vitamin D pathway gene variants reveal contrasting effects on serum vitamin D levels in African Americans and European Americans. *Hum. Genet.* **2014**, *133*, 1395–1405. [[CrossRef](#)] [[PubMed](#)]
42. Mogire, R.M.; Mutua, A.; Kimita, W.; Kamau, A.; Bejon, P.; Pettifor, J.M.; Adeyemo, A.; Williams, T.N.; Atkinson, S.H. Prevalence of vitamin D deficiency in Africa: A systematic review and meta-analysis. *Lancet Glob. Health* **2020**, *8*, e134–e142. [[CrossRef](#)]
43. Petrenya, N.; Lamberg-Allardt, C.; Melhus, M.; Broderstad, A.R.; Brustad, M. Vitamin D status in a multi-ethnic population of northern Norway: The SAMINOR 2 Clinical Survey. *Public Health Nutr.* **2020**, *23*, 1186–1200. [[CrossRef](#)] [[PubMed](#)]