

Comprehensive association analysis of nine candidate genes with serum 25-hydroxy vitamin D levels among healthy Caucasian subjects

Feng-Xiao Bu · Laura Armas · Joan Lappe ·
Yu Zhou · Guimin Gao · Hong-Wei Wang ·
Robert Recker · Lan-Juan Zhao

Received: 8 April 2010 / Accepted: 23 August 2010 / Published online: 1 September 2010
© Springer-Verlag 2010

Abstract Vitamin D deficiency is a common public health problem in the US. It is related to the high risk of rickets, osteoporosis and other diseases. Currently, serum 25-hydroxy vitamin D [25(OH)D] concentration is the best indicator of vitamin D status, and determination of its deficiency or sufficiency. This level has high heritability (28–80%). However, genes contributing to the wide variation in serum 25(OH)D are generally unknown. In this study, we screened nine important genes in vitamin D metabolic pathways using 49 single nucleotide polymorphism (SNP) markers in a group of 156 unrelated healthy Caucasian subjects. Significant confounding factors that may affect serum 25(OH)D variations were used as covariates for the association analyses. An association test for quantitative trait was performed to evaluate the association between candidate genes and serum 25(OH)D levels. Permutation was conducted for correcting multiple testing problems. Evidence of association was observed at SNPs

in the CYP2R1 (cytochrome P450, family 2, subfamily R, polypeptide 1) and the GC (vitamin D binding protein) gene. Next, we performed a replication study for six promising SNPs in the gene CYP2R1 and GC, using another group of 340 unrelated healthy Caucasian subjects. Association analyses were conducted in the replication cohort ($n = 340$) and the pooled cohort ($n = 496$). The CYP2R1 gene and the GC gene remain significant in the pooled cohort. The results suggest that the CYP2R1 and GC genes may contribute to the variation of serum 25(OH)D levels in healthy populations.

Introduction

Vitamin D deficiency is a common public health problem in the United States. It is reported that 40–100% elderly US and European men and women suffer from vitamin D deficiency (Holick 2007). This deficiency has been demonstrated to relate to bone health problems like rickets and osteoporosis because of reduced calcium absorption (Rovner and Miller 2008; Sahota 2000; Simonelli 2005; Wharton and Bishop 2003). Additionally, recent data indicate that vitamin D deficiency may relate with other diseases like diabetes, cardiovascular disease, renal disease, and some cancers (Borkar et al. 2009; Feskanich et al. 2004; Kilkkinen et al. 2009; Melamed et al. 2009).

Serum 25-hydroxy vitamin D [25(OH)D] concentration is the best indicator of Vitamin D status, and determination of its deficiency or sufficiency. Recent studies indicate that genetic factors may play an important role in the determination of serum 25(OH)D levels. Wjst et al. performed a genome-wide linkage scan in asthma families for serum 25(OH)D. Results show that serum 25(OH)D levels have a heritability of 80% in these families (Wjst et al. 2007).

F.-X. Bu · L. Armas · J. Lappe · Y. Zhou · R. Recker ·
L.-J. Zhao (✉)
Osteoporosis Research Center, Creighton University Medical
Center, Creighton University, 601 N 30th ST, Suite 6730,
Omaha, NE 68131, USA
e-mail: lanjuanzhao@creighton.edu

L.-J. Zhao
Department of Biomedical Sciences,
Creighton University, Omaha, NE 68131, USA

G. Gao
Department of Biostatistics,
Virginia Commonwealth University,
Richmond, VA 23298, USA

H.-W. Wang
Department of Medicine, University of Chicago,
Chicago, IL 60637, USA

In two other studies performed in 226 adolescent twins by Arguelles et al. (2009), and in 1,762 participants by Shea et al. (2009), the genetic factors accounting for the variability in serum 25(OH)D levels were estimated as 68.9 and 28.8%, respectively.

Although serum 25(OH)D has a high genetic determination, few genes have been identified. We conducted a cross-sectional study in healthy Caucasian subjects to investigate the association between variation in serum 25(OH)D levels and nine prominent candidate genes, which involve synthesizing, transporting, degrading vitamin D, or activating downstream signaling pathways for vitamin D. Next, we conducted a replication study in another independent cohort of healthy subjects to confirm possible significant results.

Methods and materials

Participants

Subjects for the discovery cohort were 156 unrelated non-Hispanic white men and women, randomly selected from three vitamin D studies in the Osteoporosis Research Center (ORC) at Creighton University. Details of the recruitment criteria have been reported in previous reports (Lappe et al. 2006). The subjects for the discovery cohort were randomly recruited from three locations: Alaska (41 subjects), Hawaii (23 subjects), and Nebraska (92 subjects). Subjects for the replication study were 340 unrelated non-Hispanic white postmenopausal women, randomly recruited from a rural area of eastern Nebraska (Lappe et al. 2006). All the subjects were generally healthy. Subjects who had diseases that may affect vitamin D metabolism were excluded. These diseases included: (1) any chronic kidney disease, (2) Paget's or other metabolic bone diseases, and (3) all known cancers. All subjects provided written informed consent, and the Institutional Review Board at Creighton University approved the project. The characteristics of the discovery and replication cohorts are presented in Table 1.

Serum samples for 25(OH)D measurements were collected at the initial visit. Total serum 25(OH)D levels, including both 25(OH)D₂ and 25(OH)D₃, were measured by radioimmunoassay (Nichols/Quest Diagnostics, San Clemente, CA, USA) in a single laboratory participating in the international quality assessment process for 25(OH)D (DEQAS) assays (Carter et al. 2004). Other important facts possibly influencing variation in serum 25(OH)D levels were also collected, such as age, date of blood collection, serum 25(OH)D measurement date, height, weight, habitual Vitamin D supplementation and other medication intake.

Candidate genes, tag SNPs selection

We selected nine candidate genes according to the following criteria: (1) evidence of significant association in previous studies; (2) biological importance in vitamin D metabolism, transportation, degradation, or downstream signaling activation. The genes selected are ALPL (alkaline phosphatase), CYP24A1 (vitamin D 24-hydroxylase), CYP27A1 (vitamin D₃ 25-hydroxylase), CYP27B1 [25(OH)D-1- α hydroxylase], CYP2R1 (cytochrome P450, family 2, subfamily R, polypeptide 1), CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4), GC (vitamin D binding protein), VDR (vitamin D receptor), PTH (parathyroid hormone). The basic characteristics of the nine genes are tabulated in Table 2.

We used the software program SNPbrowser (v4.0.1) to select tag SNPs within and around the nine candidate genes, which have minor allele frequency (MAF) >10% in the HapMap CEU population. The tag SNP selection is based on the HapMap database (release 20, January 24, 2006) with two methods: pair-wise r^2 ($r^2 \geq 0.8$) and haplotype R^2 ($R^2 \geq 0.8$) (De La Vega et al. 2006). In addition to tag SNPs, we chose other SNPs in the promoter, including 3'UTR, and exons region that indicate potential functional importance. All chosen SNPs were confirmed from NCBI (<http://www.ncbi.nlm.nih.gov/SNP/>) and HapMap (<http://www.hapmap.org>).

Table 1 Participants characterization

	Discovery cohort		Replication cohort
	Male (<i>n</i> = 66) Mean (SD)	Female (<i>n</i> = 90) Mean (SD)	Female (<i>n</i> = 340) Mean (SD)
Age (years)	44.03 (14.72)	60.69 (12.82)	65.29 (6.72)
Height (m)	1.65 (0.08)	1.65 (0.08)	1.63 (0.06)
Weight (kg)	73.53 (20.65)	74.10 (13.71)	77.71 (15.00)
Body mass index (kg/m ²)	26.85 (7.17)	27.18 (4.29)	29.18 (5.35)
Habitual VD supplement per day (IU)	142.27 (249.81)	173.52 (209.96)	204.31 (176.67)
Serum 25(OH)D (nmol/L)	73.48 (25.98)	73.58 (18.93)	73.35 (20.93)

Table 2 Basic characteristics of the nine candidate genes

Acronym	Full name	Reference for the gene's function	Region	Length (kb)	Number of exons	Selected SNPs
ALPL	Alkaline phosphatase	Bang et al. (2009)	1p36.1	69.0	12	6
CYP24A1	Vitamin D 24-hydroxylase	Sawada et al. (2004)	20q13.2	20.5	12	6
CYP27A1	Vitamin D(3) 25-hydroxylase	Gottfried et al. (2006)	2q35	32.9	9	6
CYP27B1	25(OH)D-1-alpha hydroxylase	Akiba et al. (1980)	12q13.1	4.9	9	2
CYP2R1	Cytochrome P450, family 2, subfamily R, polypeptide 1	Cheng et al. (2003)	11p15.2	14.2	5	4
CYP3A4	Cytochrome P450, family 3, subfamily A, polypeptide 4	Gupta et al. (2005)	7q21.1	27.2	13	5
GC	Vitamin D binding protein	Ahn et al. (2009)	4q13.3	42.5	13	7
VDR	Vitamin D receptor	Newton-Bishop et al. (2009)	12q13.1	63.5	11	9
PTH	Parathyroid hormone	Pita Martin Portela et al. (2009)	11p15.2	4.0	3	4

SNP genotyping

DNA was extracted from peripheral blood using the Gentra Puregene Blood kits (Qiagen Inc.), following the provided protocol. DNA samples were diluted to 20 µg/ml and shipped to KBioscience Company (<http://www.kbioscience.co.uk>) for SNP genotyping using the KASPar assay. Forty-six of the 49 selected SNPs were successfully genotyped and used for data analysis. Analysis showed that SNPs rs3782130 and rs2853562 were monomorphic in the discovery cohort; genotyping for SNP rs4646535 failed. After excluding the three SNPs, the genotype call rate for the 46 successfully genotyped SNPs is 97.8%.

Statistical data analyses in the discovery cohort

To illuminate the potential effect of covariates on serum 25(OH)D levels, a Pearson correlation was conducted (SPSS, version 13.0). Age, gender, BMI (body mass index), habitual vitamin D supplementation, and blood collecting season (December–February, March–May, June–August, September–November) were considered as potential covariates. Subsequently, the serum 25(OH)D levels were adjusted by the significant covariates using a linear regression approach (SPSS, version 13.0). The adjusted serum 25(OH)D levels were used for the data analyses.

Several statistical data analyses were conducted in the discovery cohort. The Hardy–Weinberg equilibrium (HWE) of the genotypic frequencies among subjects was examined. Association between the adjusted serum 25(OH)D level (a quantitative trait) and SNPs were tested by using the Wald test implemented in PLINK (Purcell et al. 2007).

To control family-wise error rate, the Max(T) permutation procedure in PLINK was performed for correcting multiple testing and adjusting *P* values. We conducted 10,000 permutations (by shuffling the phenotypes in each permutation) to generate 10,000 replicated datasets. The

observed Wald test statistic at each SNP was compared against the maximum of all statistics (over all SNPs) for each single replicate. Let *n* denote the number of replicates for which the maximum statistic is greater than or equal to the observed statistic at the SNP. The adjusted *P* value for the SNP was estimated as the ratio of *n*/10,000.

Replication and pooled analyses

The replication analyses were conducted in the replication cohort for significant SNPs found in the discovery cohort. Significant SNPs were genotyped in the replication cohort using the same method in the same company as the discovery cohort. Age, BMI, habitual vitamin D supplementation, and phlebotomy season were tested as potential covariates for serum 25(OH)D concentrations. Significant covariates were used for adjusting serum 25(OH)D levels. The same statistical approaches and tools were used for replication analyses to identify genetic variants important for serum 25(OH)D variation. In addition, a one-way ANOVA was performed on the three genotype groups (GG, AA, and AG) of the most significant SNP rs12794714 to compare the difference of serum 25(OH)D levels.

A pooled analysis of the groups was conducted involving all 496 subjects. The same phenotypes of the groups and the genotypes of replicated SNPs were combined as one set to perform the analysis. Significant covariates were evaluated and used for adjusting serum 25(OH)D levels. Association tests were conducted using PLINK (version 1.0.7) by the same methods, as for the discovery cohort.

Results

Table 3 lists the basic information of the 46 successfully genotyped SNPs. All of them passed the HWE test. The average MAF of the 46 SNP markers are 31.5%, ranging

Table 3 Results of single SNP association with serum 25(OH)D level in the discovery cohort ($n = 156$)

Gene	Locus	SNP	Allele	Location (bp)	Function	MAF	HWE	BETA	R^2	P	EMP1	EMP2
ALPL	Intron	rs1472563	A/G	21,840,498		0.46	0.19	-5.00	0.03	0.047	0.046	0.785
ALPL	Intron	rs12141569	T/C	21,873,002		0.46	0.87	-2.82	0.01	0.232	0.232	1.000
ALPL	Intron	rs1767429	T/C	21,889,340		0.41	0.50	-2.78	0.01	0.270	0.265	1.000
ALPL	Intron	rs10917008	T/C	21,894,307		0.26	0.53	0.05	0.00	0.984	0.983	1.000
ALPL	Exon	rs3200254	T/C	21,894,735	Synonymous	0.12	0.13	6.30	0.02	0.068	0.068	0.890
ALPL	Intron	rs1780320	C/T	21,899,491		0.33	1.00	2.56	0.01	0.307	0.305	1.000
CYP24A1	Exon	rs4809957	A/G	52,771,171	3' UTR	0.24	0.83	-1.68	0.00	0.543	0.541	1.000
CYP24A1	Exon	rs2296239	C/T	52,775,528	Synonymous	0.24	0.83	-1.68	0.00	0.541	0.540	1.000
CYP24A1	Exon	rs6068816	C/T	52,781,091	Synonymous	0.11	1.00	-0.97	0.00	0.807	0.808	1.000
CYP24A1	Intron	rs2244719	T/C	52,782,858		0.45	0.75	0.73	0.00	0.755	0.759	1.000
CYP24A1	Exon	rs2296241	G/A	52,786,219	Synonymous	0.48	0.43	0.01	0.00	0.998	0.999	1.000
CYP24A1	5' Near gene	rs2248359	C/T	52,791,518	Promoter	0.35	0.86	3.73	0.02	0.123	0.123	0.983
CYP27A1	5' Near gene	rs7603709	G/A	219,642,601	Promoter	0.47	0.08	-0.85	0.00	0.697	0.697	1.000
CYP27A1	Intron	rs17470271	A/T	219,651,995		0.40	0.87	0.08	0.00	0.973	0.974	1.000
CYP27A1	Intron	rs4674344	A/T	219,669,875		0.48	0.15	-2.02	0.01	0.361	0.362	1.000
CYP27A1	Intron	rs6709815	G/T	219,675,067		0.49	0.15	-3.55	0.02	0.103	0.101	0.965
CYP27A1	3' Near gene	rs645163	G/A	219,682,257		0.13	0.48	-5.13	0.02	0.120	0.117	0.981
CYP27B1	Intron	rs4646536*	C/T	58,157,988		0.33	1.00	-2.96	0.01	0.234	0.237	1.000
CYP27B1	5' Near gene	rs10877012	G/T	58,162,085	Promoter	0.33	1.00	-3.16	0.01	0.209	0.210	0.999
CYP2R1	Exon	rs12794714	G/A	14,913,575	Synonymous	0.43	0.05	-7.63	0.08	0.001	0.001	0.022
CYP2R1	5' Near gene	rs10741657*	G/A	14,914,878	Promoter	0.40	0.87	7.56	0.07	0.001	0.002	0.051
CYP2R1	5' Near gene	rs1562902	T/C	14,918,216	Promoter	0.47	0.63	6.08	0.04	0.011	0.011	0.311
CYP2R1	5' Near gene	rs10766197*	G/A	14,921,880	Promoter	0.45	0.63	-6.45	0.05	0.005	0.005	0.153
CYP3A4	3' Near gene	rs12333983	T/A	99,354,114		0.10	1.00	-3.18	0.00	0.421	0.426	1.000
CYP3A4	Intron	rs3735451	A/G	99,355,975		0.11	1.00	-3.30	0.00	0.396	0.399	1.000
CYP3A4	Intron	rs6956344	C/T	99,359,151		0.08	0.60	-3.90	0.01	0.378	0.378	1.000
CYP3A4	Intron	rs4646437	C/T	99,365,083		0.10	0.67	-5.10	0.01	0.180	0.182	0.998
CYP3A4	Intron	rs2246709	A/G	99,365,719		0.28	0.43	2.76	0.01	0.305	0.307	1.000
GC	Intron	rs2282679*	A/C	72,608,383		0.30	0.05	-2.47	0.01	0.305	0.300	1.000
GC	Exon	rs4588	C/A	72,618,323	Missense Thr > Lys	0.29	0.08	-1.73	0.00	0.471	0.466	1.000
GC	Exon	rs7041*	G/T	72,618,334	Missense Glu > Asp	0.50	0.87	2.94	0.01	0.203	0.204	0.999
GC	Intron	rs222020	T/C	72,636,272		0.14	0.52	8.42	0.04	0.010	0.009	0.286
GC	Intron	rs1155563*	T/C	72,643,488		0.28	0.02	-2.47	0.01	0.293	0.294	1.000
GC	Intron	rs2298849	T/C	72,648,851		0.20	0.80	6.38	0.03	0.026	0.026	0.574
GC	5' Near gene	rs16847039	C/T	72,654,267	Promoter	0.20	1.00	5.80	0.03	0.052	0.052	0.817
PTH	Exon	rs6256	C/A	13,514,053	Synonymous	0.17	0.57	1.70	0.00	0.609	0.611	1.000
PTH	Intron	rs6254	G/A	13,514,263		0.30	1.00	0.47	0.00	0.858	0.856	1.000
PTH	Intron	rs177706	A/G	13,514,505		0.37	0.86	-0.89	0.00	0.722	0.725	1.000
PTH	Intron	rs3099597	T/C	13,517,404		0.37	0.48	-1.21	0.00	0.643	0.644	1.000
VDR	Exon	rs731236	T/C	48,238,757	Synonymous	0.35	0.29	-2.63	0.01	0.270	0.274	1.000
VDR	Intron	rs7975232	A/C	48,238,837		0.48	0.63	-3.15	0.01	0.169	0.172	0.997
VDR	Intron	rs1544410	G/A	48,239,835		0.36	0.86	-2.66	0.01	0.278	0.282	1.000
VDR	Intron	rs2239186	T/C	48,269,410		0.20	0.80	-4.39	0.01	0.150	0.150	0.994
VDR	Exon	rs2228570	C/T	48,272,895	Missense Thr > Met	0.46	0.74	0.74	0.00	0.758	0.755	1.000
VDR	Intron	rs10783219	A/T	48,295,488		0.34	0.11	1.77	0.00	0.512	0.514	1.000
VDR	5' Near gene	rs11568820*	G/A	48,302,545	Promoter	0.21	1.00	-2.80	0.01	0.335	0.335	1.000

Bold numbers represent significant P values

Allele major allele/minor allele, MAF minor allele frequency, HWE P values for Hardy–Weinberg Equilibrium test, BETA regression coefficient, R^2 regression r -squared, P Wald test asymptotic P value, EMP1 empirical P value (point-wise), EMP2 adjusted empirical P value by using Max(T) in PLINK for correcting for multiple testing

* SNPs referred from previous studies: rs4646536 (Orton et al. 2008); rs10741657 (Ramos-Lopez et al. 2007) ; rs10766197 (Wjst et al. 2006); rs2282679, rs7041, rs1155563 (Ahn et al. 2009); rs11568820 (Mikhak et al. 2007)

from 10 to 50%. We adjusted serum 25(OH)D concentration using significant covariates. In the discovery cohort, BMI (correlation coefficient $r = -0.253$, $P = 0.001$) and habitual vitamin D supplementation ($r = 0.281$, $P = 0.001$) had significant effects on serum 25(OH)D levels. These two covariates account for 14% of variation in serum 25(OH)D levels. We tested the effects of potential covariates for serum 25(OH)D in the replication cohort. As in the discovery cohort, BMI ($r = -0.307$, $P = 7.7e-9$) and habitual vitamin D supplementation ($r = 0.247$, $P = 4.1e-6$), were significantly correlated with serum 25(OH)D levels in replication cohort. Age, gender, and phlebotomy season were not significant in either the discovery cohort or replication cohort.

Three P values are reported for the association results. We set the nominal significance level as 0.05 for Wald test for individual SNP. Six SNPs in the CYP2R1 and GC genes were identified to be significantly associated with serum 25(OH)D levels (Table 3). Three SNPs in the promoter of the CYP2R1 gene, rs10741657 ($P_{\text{discovery}} = 0.001$), rs1562902 ($P_{\text{discovery}} = 0.011$), and rs10766197 ($P_{\text{discovery}} = 0.005$) showed evidence of association. SNP rs12794714, which is a synonymous mutation of Ser to Ser in exon 1 of the CYP2R1 gene, reached the lowest P value among all tested SNP markers. For the GC gene, SNPs rs222020 ($P_{\text{discovery}} = 0.010$) and rs2298849 ($P_{\text{discovery}} = 0.026$)

were significantly associated with serum 25(OH)D concentrations. If we set the family-wise error rate (i.e. significance level for all SNPs) as 0.05, after adjusting the multiple testing, SNP rs12794714 of CYP2R1 gene maintained significance (adjusted $P_{\text{discovery}} = 0.022$) (Table 3).

Replication analyses were conducted on the six significant SNPs identified at the nominal significance level of 0.05 for single SNP association test. The call rate was 95% for the genotyping experiment for the six SNPs. At the nominal significance level of 0.05 for single tests, SNPs rs12794714 ($P_{\text{replication}} = 0.018$) and rs10766197 ($P_{\text{replication}} = 0.022$) in the CYP2R1 gene were confirmed to be significantly associated with the variation in serum 25(OH)D levels (Table 4). The SNP marker rs222020 ($P_{\text{replication}} = 0.037$) in the GC gene was confirmed to be marginally significant.

All six replicated SNPs were used to construct the pooled dataset of the two groups. After correcting for multiple testing, SNPs rs12794714, rs10741657 and rs10766197 of CYP2R1 gene, and rs222020 of GC gene were found to have a significant association with serum 25(OH)D levels (Table 5). The SNP rs12794714 in the CYP2R1 gene reached the lowest adjusted P value at 1.00×10^{-4} .

We compared the raw serum 25(OH)D levels in the three genotypes on the most significant SNP rs12794714 of the CYP2R1 gene using one-way ANOVA (Fig. 1). For SNP rs12794714, significant difference of average serum

Table 4 Replication results of single SNPs association analyses for the serum 25(OH)D level in the replication cohort ($n = 340$)

Gene	Locus	SNP	Allele	MAF	HWE	BETA	R^2	P	EMP1	EMP2
CYP2R1	Exon	rs12794714	G/A	0.40	0.49	-3.61	0.02	0.018	0.016	0.114
CYP2R1	Promoter	rs10741657	G/A	0.39	0.47	2.33	0.01	0.149	0.146	0.640
CYP2R1	Promoter	rs1562902	T/C	0.49	0.12	1.28	0.00	0.377	0.378	0.954
CYP2R1	Promoter	rs10766197	G/A	0.41	0.64	-3.55	0.02	0.022	0.024	0.145
GC	Intron	rs222020	T/C	0.14	0.82	4.42	0.01	0.037	0.038	0.224
GC	Intron	rs2298849	T/C	0.21	1.00	2.17	0.00	0.243	0.245	0.831

Bold numbers represent significant P values

Allele major allele/minor allele, MAF minor allele frequency, HWE P values for Hardy–Weinberg Equilibrium test, BETA regression coefficient, R^2 regression r -squared, P Wald test asymptotic P value, EMP1 empirical P value (point-wise), EMP2 adjusted empirical P value by using Max(T) in PLINK for correcting multiple testing

Table 5 Results of single SNPs association analyses for the serum 25(OH)D level in the pooled dataset of the groups ($n = 496$)

Gene	Locus	SNP	Allele	MAF	HWE	BETA	R^2	P	EMP1	EMP2
CYP2R1	Exon	rs12794714	G/A	0.41	0.07	-5.03	0.03	5.74e-5	1.00e-4	1.00e-4
CYP2R1	Promoter	rs10741657	G/A	0.39	0.69	4.12	0.02	0.002	0.002	0.010
CYP2R1	Promoter	rs1562902	T/C	0.48	0.31	2.69	0.01	0.030	0.031	0.124
CYP2R1	Promoter	rs10766197	G/A	0.43	0.51	-4.53	0.03	4.04e-4	3.00e-4	0.002
GC	Intron	rs222020	T/C	0.14	0.57	5.79	0.02	0.001	0.002	0.006
GC	Intron	rs2298849	T/C	0.20	1.00	3.59	0.01	0.021	0.020	0.090

Bold numbers represent significant P values

Allele major allele/minor allele, MAF minor allele frequency, HWE P values for Hardy–Weinberg Equilibrium test, BETA regression coefficient, R^2 regression r -squared, P Wald test asymptotic P value, EMP1 empirical P value (point-wise), EMP2 adjusted empirical P value by using Max(T) in PLINK for correcting multiple testing

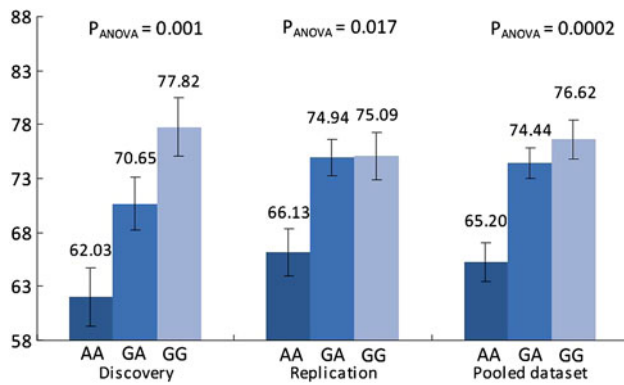


Fig. 1 The raw serum 25(OH)D means (\pm SE) in the three genotypes of SNP rs12794714 in the discovery cohort, replication cohort and pooled dataset. X axis stands for three genotypes in the two cohorts; Y axis stands for the raw serum 25(OH)D concentrations (nmol/L); Error bar stands by standard error of average serum 25(OH)D concentrations. One-way ANOVA was conducted to compare the serum 25(OH)D levels grouped by genotypes (SPSS 13.0)

25(OH)D levels was shown in the discovery cohort (AA: 62.03 ± 17.82 nmol/L, GA: 70.65 ± 21.70 nmol/L, GG: 77.82 ± 21.96 nmol/L, $P = 0.001$) and the replication cohort (AA: 66.13 ± 16.18 nmol/L, GA: 74.94 ± 19.90 nmol/L, GG: 75.09 ± 23.67 nmol/L, $P = 0.017$). By pooling two cohorts into a dataset, the lowest P value of one-way ANOVA was gained (AA: 65.20 ± 16.96 nmol/L, GA: 74.44 ± 20.83 nmol/L, GG: 76.62 ± 23.32 nmol/L, $P = 0.0002$). The results indicate that raw serum 25(OH)D levels are significantly different between GG, AA, and GA genotypes of SNP rs12794714 in both discovery and replication cohorts.

Discussion

The present study investigated the association of nine prominent candidate genes with serum 25(OH)D levels. The GC and CYP2R1 genes were shown to be associated with serum 25(OH)D levels in the discovery cohort at nominal significance level of 0.05 for single tests. Further replication analysis and pooled dataset analysis confirmed the association between the CYP2R1 and GC genes and serum 25(OH)D concentrations, suggesting that the CYP2R1 and GC genes play an important role in regulating serum 25(OH)D levels in the non-Hispanic white population.

CYP2R1 is a member of CYP2 family encoding cytochrome P450 proteins. It is a key vitamin D 25-hydroxylase which hydroxylates vitamin D at the 25-C position for 25(OH)D synthesis in the liver (Cheng et al. 2003; Shinkyo et al. 2004). Previous data show that the CYP2R1 gene is associated with several vitamin D related diseases, such as type 1 diabetes (Ramos-Lopez et al. 2007), ovarian

cancer (Downie et al. 2005), and asthma and atopy (Bosse et al. 2009).

Our study found that the CYP2R1 gene is associated with serum 25(OH)D levels. This finding is supported by previous studies. Cheng et al. reported that a patient with low circulating levels of serum 25(OH)D and classic symptoms of vitamin D deficiency had a homozygous mutation L99P in exon 2 of the CYP2R1 gene. This homozygous mutation caused inactivation of CYP2R1 (Cheng et al. 2004). Ahn et al. performed a combined meta-analysis on 4,501 subjects of European ancestry from five cohorts (Ahn et al. 2010). The significant findings were replicated in 2,221 subjects. Results show that rs2060793 and rs1993116 in the CYP2R1 gene are associated with serum 25(OH)D levels (Ahn et al. 2010). Interestingly, it is consistent with our results. SNP rs2060793 is located in the promoter region of the CYP2R1 gene. It has a high LD value with SNP rs10741657 in Caucasian population ($D' = 1$, $r^2 = 1$, HapMap Data Rel 24/phase II Nov 08). Ramos-Lopez et al. tested the association of CYP2R1 gene with variation in serum 25(OH)D concentration in 609 subjects from 203 type 1 diabetes families. The study found that SNP rs10741657 is associated with serum levels of 25(OH)D (Ramos-Lopez et al. 2007). In addition, in a recent genome-wide association study (GWAS), Wang et al. (2010) found multiple SNPs (including rs12794714 and rs10741657) in the CYP2R1 gene that are significantly associated with 25(OH)D levels in $\sim 30,000$ individuals of European descent from 15 cohorts. The SNP rs10741657 is significant in the association analyses in our pooled dataset, and the SNP rs12794714 located in exon 1 of CYP2R1 gene, is significant in both the discovery cohort and the pooled dataset in our study. In the work of Wjst et al. (2006), SNP rs10766197 in the CYP2R1 gene was significantly associated with the 25(OH)D levels in 872 participants of the German Asthma Family Study. This same SNP is also significant in the pooled dataset in our study. Both of the SNPs rs10741657 and rs10766197 are located in the promoter region of the CYP2R1 gene. All these aforementioned studies indicate that genetic variants of the CYP2R1 gene are associated with serum 25(OH)D variation.

Recent association studies revealed several important genetic variants in the GC gene for the serum 25(OH)D variation. Ahn J et al. (2009) suggested genetic markers rs12512631, rs2282679, rs7041, and rs1155563. And in a recent GWAS study for serum 25(OH)D in 4,501 persons of European ancestry, SNP rs2282679 in the GC gene was the most significant one (Ahn et al. 2010). Interestingly, SNP rs2282679 is also significantly associated with serum 25(OH)D level in the GWAS on $\sim 30,000$ individuals in the SUNLIGHT consortium (Wang et al. 2010). Our most significant finding in the GC gene is SNP rs222020. Our

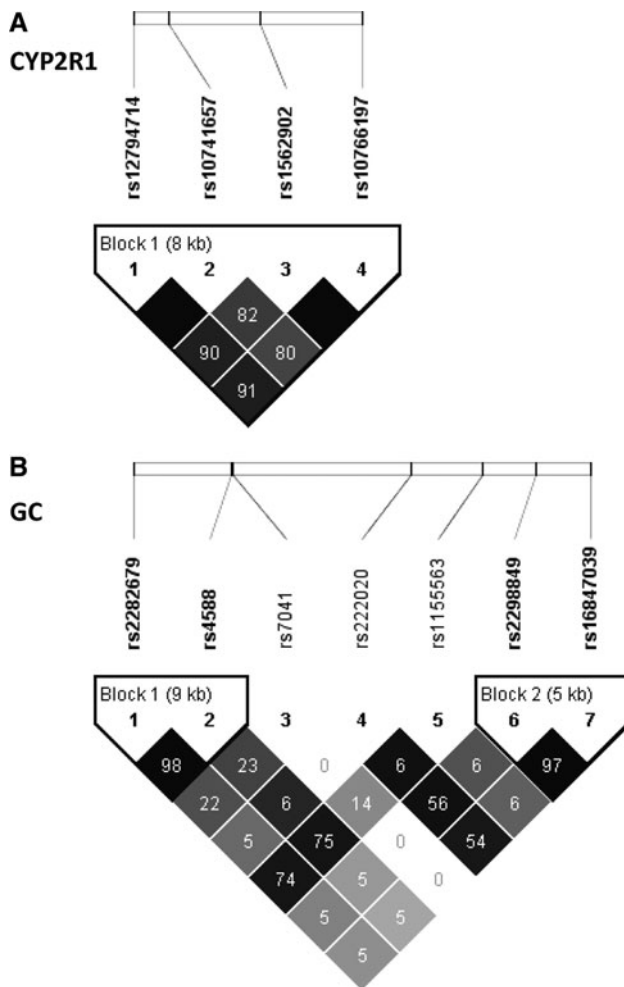


Fig. 2 LD plots with r^2 values of CYP2R1 (a) and GC (b) gene in discovery cohort. The figure is generated by Haploview. D' values were indicated by the dark depth. r^2 values multiplied by 100 were shown as number in the diamonds

LD data analysis indicates low LD between SNP rs222020 and SNP rs2282679 ($r^2 = 0.05$) (Fig. 2). Both SNP rs222020 and SNP rs2282679 are located in the intron of the gene. It is possible that an unknown functional genetic variant, which is near the two SNPs, is important for the regulation of serum 25(OH)D levels. Future sequencing studies or large-scale association studies with dense markers may reveal such genetic variance.

Compared to previous studies, our study has two strengths: (1) This study, conducted in a healthy population to analyze the genetic association with variation in serum 25(OH)D levels, eliminated potential impacts of diseases. (2) Dense markers in nine important candidate genes involved in vitamin D metabolism were selected. One limitation of the study is that the sample size for the discovery cohort is relatively small. Given the lack of knowledge

regarding genes regulating prevalent serum 25(OH)D levels, more genetic association studies are needed.

In summary, after comprehensive screening of nine functionally important vitamin D candidate genes, our study suggests that the CYP2R1 and GC genes may be important in regulating serum 25(OH)D levels in healthy Caucasian subjects. It is important to confirm these findings in other large-scale healthy populations or in other races. Further studies are needed to identify the functional genetic variants, and to characterize their functions.

Acknowledgments We thank Darin Jensen for his constructive input during the preparation of the manuscript. Investigators of this work were supported by grants from Cancer and Smoking Disease Research Bone Biology Program, the Nebraska Tobacco Settlement Biomedical Research Development Award, an NIH grant (3R01CA129488-01A2S2), and a grant from State of Nebraska Cancer and Smoking Disease Research Program (LB595).

References

- Ahn J, Albanes D, Berndt SI, Peters U, Chatterjee N, Freedman ND, Abnet CC, Huang WY, Kibel AS, Crawford ED, Weinstein SJ, Chanock SJ, Schatzkin A, Hayes RB (2009) Vitamin D-related genes, serum vitamin D concentrations and prostate cancer risk. *Carcinogenesis* 30:769–776
- Ahn J, Yu K, Stolzenberg-Solomon R, Simon KC, McCullough ML, Gallicchio L, Jacobs EJ, Ascherio A, Helzlsouer K, Jacobs KB, Li Q, Weinstein SJ, Purdue M, Virtamo J, Horst R, Wheeler W, Chanock S, Hunter DJ, Hayes RB, Kraft P, Albanes D (2010) Genome-wide association study of circulating vitamin D levels. *Hum Mol Genet* 19:2739–2745
- Akiba T, Endou H, Koseki C, Sakai F, Horiuchi N, Suda T (1980) Localization of 25-hydroxyvitamin D3-1 alpha-hydroxylase activity in the mammalian kidney. *Biochem Biophys Res Commun* 94:313–318
- Arguelles LM, Langman CB, Ariza AJ, Ali FN, Dilley K, Price H, Liu X, Zhang S, Hong X, Wang B, Xing H, Li Z, Liu X, Zhang W, Xu X, Wang X (2009) Heritability and environmental factors affecting vitamin D status in rural Chinese adolescent twins. *J Clin Endocrinol Metab* 94:3273–3281
- Bang UC, Semb S, Nordgaard-Lassen I, Jensen JE (2009) A descriptive cross-sectional study of the prevalence of 25-hydroxyvitamin D deficiency and association with bone markers in a hospitalized population. *Nutr Res* 29:671–675
- Borkar VV, Verma S, Bhalla A (2009) Low levels of vitamin D in North Indian children with newly diagnosed type 1 diabetes. *Pediatr Diabetes* 11:345–350
- Bosse Y, Lemire M, Poon AH, Daley D, He JQ, Sandford A, White JH, James AL, Musk AW, Palmer LJ, Raby BA, Weiss ST, Kozyrskyj AL, Becker A, Hudson TJ, Laprise C (2009) Asthma and genes encoding components of the vitamin D pathway. *Respir Res* 10:98
- Carter GD, Carter CR, Gunter E, Jones J, Jones G, Makin HL, Sufi S (2004) Measurement of vitamin D metabolites: an international perspective on methodology and clinical interpretation. *J Steroid Biochem Mol Biol* 89–90:467–471
- Cheng JB, Motola DL, Mangelsdorf DJ, Russell DW (2003) De-orphanization of cytochrome P450 2R1: a microsomal vitamin D 25-hydroxylase. *J Biol Chem* 278:38084–38093

- Cheng JB, Levine MA, Bell NH, Mangelsdorf DJ, Russell DW (2004) Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. *Proc Natl Acad Sci USA* 101: 7711–7715
- De La Vega FM, Isaac HI, Scafe CR (2006) A tool for selecting SNPs for association studies based on observed linkage disequilibrium patterns. *Pac Symp Biocomput* 2006:487–498
- Downie D, McFadyen MC, Rooney PH, Cruickshank ME, Parkin DE, Miller ID, Telfer C, Melvin WT, Murray GI (2005) Profiling cytochrome P450 expression in ovarian cancer: identification of prognostic markers. *Clin Cancer Res* 11:7369–7375
- Feskanich D, Ma J, Fuchs CS, Kirkner GJ, Hankinson SE, Hollis BW, Giovannucci EL (2004) Plasma vitamin D metabolites and risk of colorectal cancer in women. *Cancer Epidemiol Biomarkers Prev* 13:1502–1508
- Gottfried E, Rehli M, Hahn J, Holler E, Andreesen R, Kreutz M (2006) Monocyte-derived cells express CYP27A1 and convert vitamin D₃ into its active metabolite. *Biochem Biophys Res Commun* 349:209–213
- Gupta RP, He YA, Patrick KS, Halpert JR, Bell NH (2005) CYP3A4 is a vitamin D-24- and 25-hydroxylase: analysis of structure function by site-directed mutagenesis. *J Clin Endocrinol Metab* 90:1210–1219
- Holick MF (2007) Vitamin D deficiency. *N Engl J Med* 357:266–281
- Kilkinen A, Knekt P, Aro A, Rissanen H, Marniemi J, Heliövaara M, Impivaara O, Reunanen A (2009) Vitamin D status and the risk of cardiovascular disease death. *Am J Epidemiol* 170:1032–1039
- Lappe JM, Davies KM, Travers-Gustafson D, Heaney RP (2006) Vitamin D status in a rural postmenopausal female population. *J Am Coll Nutr* 25:395–402
- Melamed ML, Astor B, Michos ED, Hostetter TH, Powe NR, Muntner P (2009) 25-hydroxyvitamin D levels, race, and the progression of kidney disease. *J Am Soc Nephrol* 20:2631–2639
- Mikhak B, Hunter DJ, Spiegelman D, Platz EA, Hollis BW, Giovannucci E (2007) Vitamin D receptor (VDR) gene polymorphisms and haplotypes, interactions with plasma 25-hydroxyvitamin D and 1, 25-dihydroxyvitamin D, and prostate cancer risk. *Prostate* 67:911–923
- Newton-Bishop JA, Beswick S, Randerson-Moor J, Chang YM, Affleck P, Elliott F, Chan M, Leake S, Karpavicius B, Haynes S, Kukalich K, Whitaker L, Jackson S, Gerry E, Nolan C, Bertram C, Marsden J, Elder DE, Barrett JH, Bishop DT (2009) Serum 25-hydroxyvitamin D₃ levels are associated with breslow thickness at presentation and survival from melanoma. *J Clin Oncol* 27:5439–5444
- Orton SM, Morris AP, Herrera BM, Ramagopalan SV, Lincoln MR, Chao MJ, Vieth R, Sadovnick AD, Ebers GC (2008) Evidence for genetic regulation of vitamin D status in twins with multiple sclerosis. *Am J Clin Nutr* 88:441–447
- Pita Martin Portela ML, Monico A, Barahona A, Dupraz H, Sol Gonzales-Chaves MM, Zeni SN (2009) Comparative 25-OH-vitamin D level in institutionalized women older than 65 years from two cities in Spain and Argentina having a similar solar radiation index. *Nutrition* 26:283–289
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Ramos-Lopez E, Bruck P, Jansen T, Herwig J, Badenhop K (2007) CYP2R1 (vitamin D 25-hydroxylase) gene is associated with susceptibility to type 1 diabetes and vitamin D levels in Germans. *Diabetes Metab Res Rev* 23:631–636
- Rovner AJ, Miller RS (2008) Vitamin D deficiency and insufficiency in children with osteopenia or osteoporosis. *Pediatrics* 122:907–908 (author reply 908–909)
- Sahota O (2000) Osteoporosis and the role of vitamin D and calcium—vitamin D deficiency, vitamin D insufficiency and vitamin D sufficiency. *Age Ageing* 29:301–304
- Sawada N, Kusudo T, Sakaki T, Hatakeyama S, Hanada M, Abe D, Kamao M, Okano T, Ohta M, Inouye K (2004) Novel metabolism of 1 alpha, 25-dihydroxyvitamin D₃ with C24–C25 bond cleavage catalyzed by human CYP24A1. *Biochemistry* 43:4530–4537
- Shea MK, Benjamin EJ, Dupuis J, Massaro JM, Jacques PF, D'Agostino RB Sr, Ordovas JM, O'Donnell CJ, Dawson-Hughes B, Vasani RS, Booth SL (2009) Genetic and non-genetic correlates of vitamins K and D. *Eur J Clin Nutr* 63:458–464
- Shinkyo R, Sakaki T, Kamakura M, Ohta M, Inouye K (2004) Metabolism of vitamin D by human microsomal CYP2R1. *Biochem Biophys Res Commun* 324:451–457
- Simonelli C (2005) The role of vitamin D deficiency in osteoporosis and fractures. *Minn Med* 88:34–36
- Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, Kiel DP, Streeten EA, Ohlsson C, Koller DL, Peltonen L, Cooper JD, O'Reilly PF, Houston DK, Glazer NL, Vandenput L, Peacock M, Shi J, Rivadeneira F, McCarthy MI, Anelli P, de Boer IH, Mangino M, Kato B, Smyth DJ, Booth SL, Jacques PF, Burke GL, Goodarzi M, Cheung CL, Wolf M, Rice K, Goltzman D, Hidirolou N, Ladouceur M, Wareham NJ, Hocking LJ, Hart D, Arden NK, Cooper C, Malik S, Fraser WD, Hartikainen AL, Zhai G, Macdonald HM, Forouhi NG, Loos RJ, Reid DM, Hakim A, Dennison E, Liu Y, Power C, Stevens HE, Jaana L, Vasani RS, Soranzo N, Bojunga J, Psaty BM, Lorentzon M, Foroud T, Harris TB, Hofman A, Jansson JO, Cauley JA, Uitterlinden AG, Gibson Q, Jarvelin MR, Karasik D, Siscovick DS, Econs MJ, Kritchevsky SB, Florez JC, Todd JA, Dupuis J, Hypponen E, Spector TD (2010) Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet* 376:180–188
- Wharton B, Bishop N (2003) Rickets. *Lancet* 362:1389–1400
- Wjst M, Altmüller J, Faus-Kessler T, Braig C, Bahnweg M, Andre E (2006) Asthma families show transmission disequilibrium of gene variants in the vitamin D metabolism and signalling pathway. *Respir Res* 7:60
- Wjst M, Altmüller J, Braig C, Bahnweg M, Andre E (2007) A genome-wide linkage scan for 25-OH-D(3) and 1, 25-(OH)₂-D₃ serum levels in asthma families. *J Steroid Biochem Mol Biol* 103:799–802