Common genetic determinants of vitamin D insufficiency: a genome-wide association study


Summary

Background Vitamin D is crucial for maintenance of musculoskeletal health, and might also have a role in extraskeletal tissues. Determinants of circulating 25-hydroxyvitamin D concentrations include sun exposure and diet, but high heritability suggests that genetic factors could also play a part. We aimed to identify common genetic variants affecting vitamin D concentrations and risk of insufficiency.

Methods We undertook a genome-wide association study of 25-hydroxyvitamin D concentrations in 33,996 individuals of European descent from 15 cohorts. Five epidemiological cohorts were designated as discovery cohorts (n=16,125), five as in-silico replication cohorts (n=9,367), and five as de-novo replication cohorts (n=8,504). 25-hydroxyvitamin D concentrations were measured by radioimmunooassay, chemiluminescent assay, ELISA, or mass spectrometry. Vitamin D insufficiency was defined as concentrations lower than 75 nmol/L or 50 nmol/L. We combined results of genome-wide analyses across cohorts using Z-score weighted meta-analysis. Genotype scores were constructed for confirmed variants.

Findings Variants at three loci reached genome-wide significance in discovery cohorts for association with 25-hydroxyvitamin D concentrations, and were confirmed in replication cohorts: 4p12 (overall p=1.9×10⁻¹⁰⁹ for rs6013897); participants with a genotype score (combining the three confirmed variants) in the highest quartile were at increased risk of having 25-hydroxyvitamin D concentrations lower than 75 nmol/L (OR 2.47, 95% CI 2.20–2.78, p=2.3×10⁻⁴⁸) or lower than 50 nmol/L (1.92, 1.70–2.16, p=1.0×10⁻²⁶) compared with those in the lowest quartile.

Interpretation Variants near genes involved in cholesterol synthesis, hydroxylation, and vitamin D transport affect vitamin D status. Genetic variation at these loci identifies individuals who have substantially raised risk of vitamin D insufficiency.

Funding Full funding sources listed at end of paper (see Acknowledgments).

Introduction Vitamin D insufficiency affects as many as half of otherwise healthy adults in developed countries.¹ The musculoskeletal consequences of inadequate vitamin D concentrations are well established, and include childhood rickets, osteomalacia, and fractures.² Growing number of other disorders have also been linked to vitamin D insufficiency, although causal associations have not yet been established in randomised trials. These extraskeletal disorders include type 1 and type 2 diabetes,³,⁴ cardiovascular disease,⁵ increased risk of falls,⁶ and cancers of the breast, colon, and prostate.⁷,⁸ Results of a 2007 meta-analysis suggested that vitamin D supplementation substantially reduced mortality.⁹

Personal, social, and cultural factors are important determinants of vitamin D availability via their effects on sun exposure and diet. Sufficient exposure to ultraviolet light or adequate intake from diet or supplements is needed to maintain vitamin D status. Concentrations of the widely accepted biomarker for vitamin D, 25-hydroxyvitamin D, are highest in the summer and lowest in the winter in northern latitudes. However, only about a quarter of the interindividual variability in 25-hydroxyvitamin D concentration is attributable to season of measurement, geographical latitude, or reported vitamin D intake.⁹,¹⁰ Results of previous twin and family studies suggest that genetic factors contribute substantially to this variability, with estimates of...
heritability as high as 53%. Although several rare mendelian disorders cause functional vitamin D insufficiency, data for the effect of common genetic variation on vitamin D status are scarce. Candidate gene studies have been done to examine the effect of specific vitamin D pathway genes, but these studies have been limited by small sample sizes and the small numbers of variants examined.15–18

The SUNLIGHT consortium (Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits) was formed in 2008. It represents a collaboration of cohorts from the UK, USA, Canada, Netherlands, Sweden, and Finland. We aimed to identify common genetic variants affecting vitamin D concentrations and risk of vitamin D insufficiency.

Methods Participants
We undertook a large, multicentre, genome-wide association study of 15 cohorts in Europe, Canada, and the USA. The discovery sample consisted of 16125 individuals of European descent drawn from five epidemiological cohorts: the Framingham Heart Study, TwinsUK, the Rotterdam Study, the 1958 British Birth Cohort (1958BC), and the Amish Family Osteoporosis Screening Study (APOSS), and 2715 additional participants from one of the discovery cohorts (1958BC).

Full descriptions of all participating cohorts are shown in the webappendix (pp 7–14). 25-hydroxyvitamin D concentrations were measured by radioimmunoassay or chemiluminescent assay (DiaSorin Inc, Stillwater, MN, USA) in the Framingham Heart Study, ELISA on a Dade-Behring BEP2000 analyser (sensitivity 0·36 1·17×10 –5 7·59×10–5 4·99×10–9 0·43 7·43×10 –5 1·12×10–6 2·67×10–9 0·40 2·94×10 –6 1·28×10–6 6·25×10–11 0·23 1·43×10 –12 7·36×10–4 8·70×10–15 0·23 8·09×10 –5 6·24×10–4 2·36×10–7 1·73×10–10 0·23 5·98×10 –13 6·39×10–4 2·54×10–15 0·23 1·56×10 –10 7·57×10–4 8·96×10–10 0·22 1·27×10 –10 2·39×10–4 2·12×10–9 0·22 9·84×10–10 6·39×10–4 2·54×10–9 0·22 8·09×10–10 6·44×10–4 3·38×10–9 0·22 8·09×10–10 6·44×10–4 3·38×10–9 0·22 8·09×10–10 6·44×10–4 3·38×10–9 0·22 8·09×10–10 6·44×10–4 3·38×10–9

Results within each locus are ordered by strength of association with 25-hydroxyvitamin D concentration. MAF=minor allele frequency.

Table 1: Single nucleotide polymorphisms identified in genome-wide association analyses for 25-hydroxyvitamin D concentrations

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Position</th>
<th>Nearest gene(s)</th>
<th>MAF</th>
<th>Combined p value for discovery samples (up to n=16 124)</th>
<th>Combined p value for replication samples (up to n=17 744)</th>
<th>Overall p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n2282679</td>
<td>4</td>
<td>72827247</td>
<td>GC</td>
<td>0·29</td>
<td>4·57×10–9</td>
<td>2·88×10–8</td>
</tr>
<tr>
<td>n2255967</td>
<td>4</td>
<td>72828526</td>
<td>GC</td>
<td>0·29</td>
<td>7·41×10–10</td>
<td>3·00×10–10</td>
</tr>
<tr>
<td>n17467825</td>
<td>4</td>
<td>72834831</td>
<td>GC</td>
<td>0·29</td>
<td>3·85×10–10</td>
<td>1·61×10–10</td>
</tr>
<tr>
<td>n1155563</td>
<td>4</td>
<td>72862352</td>
<td>GC</td>
<td>0·30</td>
<td>4·70×10–10</td>
<td>4·26×10–10</td>
</tr>
<tr>
<td>n2298850</td>
<td>4</td>
<td>72833131</td>
<td>GC</td>
<td>0·28</td>
<td>8·94×10–10</td>
<td>2·12×10–10</td>
</tr>
<tr>
<td>n7041</td>
<td>4</td>
<td>72837198</td>
<td>GC</td>
<td>0·44</td>
<td>3·74×10–9</td>
<td>1·78×10–8</td>
</tr>
<tr>
<td>n12783878</td>
<td>11</td>
<td>70845097</td>
<td>DHCR7/NADSYN1</td>
<td>0·23</td>
<td>1·27×10–9</td>
<td>2·39×10–9</td>
</tr>
<tr>
<td>n7944926</td>
<td>11</td>
<td>70843273</td>
<td>DHCR7/NADSYN1</td>
<td>0·23</td>
<td>1·56×10–9</td>
<td>7·57×10–9</td>
</tr>
<tr>
<td>n12804638</td>
<td>11</td>
<td>70848651</td>
<td>DHCR7/NADSYN1</td>
<td>0·23</td>
<td>5·98×10–10</td>
<td>6·39×10–10</td>
</tr>
<tr>
<td>n3294060</td>
<td>11</td>
<td>70856327</td>
<td>DHCR7/NADSYN1</td>
<td>0·23</td>
<td>8·09×10–10</td>
<td>6·44×10–10</td>
</tr>
<tr>
<td>n4934508</td>
<td>11</td>
<td>7088896</td>
<td>DHCR7/NADSYN1</td>
<td>0·24</td>
<td>8·98×10–10</td>
<td>6·11×10–10</td>
</tr>
<tr>
<td>n4944457</td>
<td>11</td>
<td>70845681</td>
<td>DHCR7/NADSYN1</td>
<td>0·23</td>
<td>1·43×10–9</td>
<td>7·36×10–9</td>
</tr>
<tr>
<td>n10741657</td>
<td>11</td>
<td>14871454</td>
<td>CYP2R1</td>
<td>0·40</td>
<td>3·91×10–10</td>
<td>2·09×10–10</td>
</tr>
<tr>
<td>n2606793</td>
<td>11</td>
<td>14871886</td>
<td>CYP2R1</td>
<td>0·40</td>
<td>2·69×10–10</td>
<td>2·36×10–10</td>
</tr>
<tr>
<td>n1993116</td>
<td>11</td>
<td>14866810</td>
<td>CYP2R1</td>
<td>0·40</td>
<td>2·94×10–10</td>
<td>1·28×10–10</td>
</tr>
<tr>
<td>n12794714</td>
<td>11</td>
<td>14870151</td>
<td>CYP2R1</td>
<td>0·43</td>
<td>6·24×10–10</td>
<td>8·71×10–10</td>
</tr>
<tr>
<td>n10500804</td>
<td>11</td>
<td>14866849</td>
<td>CYP2R1</td>
<td>0·43</td>
<td>7·43×10–10</td>
<td>1·12×10–9</td>
</tr>
<tr>
<td>n7116578</td>
<td>11</td>
<td>14838347</td>
<td>CYP2R1</td>
<td>0·36</td>
<td>1·17×10–10</td>
<td>7·59×10–10</td>
</tr>
</tbody>
</table>

Procedures Details of genotyping methods, quality control, and imputation procedures used in all participating cohorts are shown in the webappendix (pp 7–14). 25-hydroxyvitamin D concentrations were measured by radioimmunoassay or chemiluminescent assay (DiaSorin Inc, Stillwater, MN, USA) in the Framingham Heart Study, TwinsUK, Rotterdam Study, Health ABC, AFOS, the GOOD cohort, and CaMoS. Detection limits ranged from 4 nmol/L to 10 nmol/L. In the 1958BC samples, 25-hydroxyvitamin D was measured with automated application of the ImmunoDiagnostics Systems OCTEIA ELISA on a Dade-Behring BEP2000 analyser (sensitivity of 5·0 nmol/L; Marburg, Germany).19 In the Cardiovascular Health Study, NFBC1966, the Indiana cohort, Chingford, Hertfordshire, and APOSS, total 25-hydroxyvitamin D was measured with high-performance liquid chromatography-tandem mass
The detection limit was 50 mg/L.

### Statistical analyses

At the threshold $\alpha=5\times10^{-8}$, with a conservative discovery sample size of 14,000, our study had 80% power to detect single nucleotide polymorphisms accounting for 0.28% of the total variance in 25-hydroxyvitamin D concentrations, and 90% power to detect polymorphisms accounting for 0.32% of the total variance.

Genome-wide analyses were done within each cohort. In the Framingham Heart Study, TwinsUK, the Rotterdam Study, 1958BC, AFOS, NFBC1966, the Indiana cohort, Health ABC, and the GOOD study, linear regression models were used to generate cohort-specific residuals of 25-hydroxyvitamin D concentrations adjusted for age, sex, body-mass index (BMI), and season. Log transformation was used to reduce skewness in the distribution of 25-hydroxyvitamin D. We modelled season using categorical variables for summer (July–September), autumn (October–December), winter (January–March), and spring (April–June). One set of definitions was used for season because most cohorts were at similar latitudes, and all were in the northern hemisphere.

In cohorts that included related individuals (Framingham, TwinsUK, AFOS, Indiana Women), we assessed association between additively coded single nucleotide polymorphism genotypes and standardised 25-hydroxyvitamin D residuals using either linear mixed-effect models or the score test implemented in MERLIN (version 1.1.2). For imputed single nucleotide polymorphisms, expected number of minor alleles (ie, dose) was used in assessments of association between genotype and 25-hydroxyvitamin D residuals. In the Cardiovascular Health Study, analyses were adjusted for age, sex, and study site by inclusion of these factors as covariates in the model. In all samples, the genomic control approach was used to adjust $p$ values for potential effects of mild population stratification and to prevent inflation of type I error occurring from any departure from normality of the trait variable.

A priori, we designated the first five genome-wide association studies, all of which used immunoenzymometric assays to measure 25-hydroxyvitamin D concentrations, as discovery samples. The remaining five studies, three of which measured 25-hydroxyvitamin D by mass spectrometry and two by immunomessay, were designated as in-silico replication samples. We selected single nucleotide polymorphisms for replication if they had meta-analytic $p$ values for association with 25-hydroxyvitamin D concentrations that were lower than $5\times10^{-8}$ in the discovery samples. Additionally, we considered polymorphisms at or near six prespecified vitamin D pathway candidate genes: vitamin D receptor (VDR), 1-α-hydroxylase (CYP27B1), 25-hydroxylase (CYP2R1), 24-hydroxylase (CYP24A1), vitamin D binding protein (GC), and 27-hydroxylase and 25-hydroxylase (CYP27A1). These polymorphisms were tested in the replication samples if they met a $p$ value threshold of $10^{-3}$ in the discovery samples. Lastly, we assessed selected polymorphisms for 25-hydroxyvitamin D association in the de-novo replication samples, using the same analytic approach. We then generated combined $p$ values across the 15 studies.

We undertook the meta-analysis using a weighted Z-score-based approach, as implemented in the software METAL (version 2009-10-10). In this approach, association $p$ values were converted to signed $Z$ statistics, for which the sign showed the direction of effect with respect to a reference allele. All $Z$ scores were assigned a weight proportional to the square root of the sample size. Weighted $Z$ statistics were summed across studies to obtain a global $Z$ score and a corresponding two-sided $p$ value. We regarded $p$ values lower than $5\times10^{-8}$ as genome-wide significant.

We also assessed whether selected genetic variants from the continuous trait analyses were associated with vitamin D insufficiency in the Framingham Heart Study, TwinsUK, CaMoS, and 1958BC. We used two thresholds

### Table 2: Mean 25-hydroxyvitamin D concentrations by genotype, season, and supplementation status

<table>
<thead>
<tr>
<th>GC*</th>
<th>Framingham Heart Study (n=5655)</th>
<th>1958 British Birth Cohort (n=5552)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor homozygotes (nmol/L)</td>
<td>82.6 (0.73)</td>
<td>61.9 (0.34)</td>
</tr>
<tr>
<td>Heterozygotes (nmol/L)</td>
<td>74.8 (0.81)</td>
<td>57.0 (0.22)</td>
</tr>
<tr>
<td>Minor homozygotes (nmol/L)</td>
<td>64.6 (1.79)</td>
<td>52.8 (0.28)</td>
</tr>
<tr>
<td>Season</td>
<td>Framingham Heart Study</td>
<td>1958 British Birth Cohort</td>
</tr>
<tr>
<td>Winter (nmol/L)</td>
<td>61.6 (1.00)</td>
<td>43.2 (0.26)</td>
</tr>
<tr>
<td>Spring/autumn (nmol/L)</td>
<td>77.4 (0.68)</td>
<td>57.1 (0.30)</td>
</tr>
<tr>
<td>Summer (nmol/L)</td>
<td>95.8 (1.00)</td>
<td>71.7 (0.31)</td>
</tr>
</tbody>
</table>

| Data are mean (SE). Sample from 1958 British Birth Cohort (1958BC) consists of a combination of the genome-wide association study sample and the de novo genotyping sample (webappendix p 2). *rs7944926 in Framingham cohort, rs4588 in TwinsUK cohort.20 The detection limit was 50 mg/L.

Spectrometry. Serum concentrations of vitamin D binding protein were measured with immunonephelometric assay in the TwinsUK cohort.20 The detection limit was 50 mg/L.
for vitamin D insufficiency: 25-hydroxyvitamin D concentrations lower than 75 nmol/L (30 ng/mL) and lower than 50 nmol/L (20 ng/mL). Covariates were age, sex, season, and BMI. We combined effect estimates from the logistic regression analysis across cohorts by meta-analysis using an inverse-variance weighting approach. We also did analyses using a 25 nmol/L (10 ng/mL) threshold, to examine whether genetic variants were associated with severe vitamin D deficiency.

Additionally, we constructed a genotype score by taking a weighted average of the number of risk alleles for members of a cohort, with weights established using β coefficients from the meta-analysis. Logistic regression was used to calculate the odds of vitamin D insufficiency according to quartile of the genotype score. For this analysis, we combined data from the Framingham Heart Study, TwinsUK, and 1958BC using a multivariate approach, with β coefficients for each quartile of genotype score meta-analysed jointly, as previously described.

Role of the funding source
The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results
Characteristics of the study cohorts are summarised in the webappendix (pp 15–17). Table 1 shows the results of genome-wide association analyses. In analysis of data from the five discovery samples, single nucleotide polymorphisms at three unique loci met the prespecified threshold for genome-wide significance: 4p12, 11q12, and 11p15. The 4p12 polymorphisms were within or near the GC gene, and the results included a non-synonymous polymorphism in this gene, rs7041. The 11q12 polymorphisms were near DHCR7/NADSYN1 (7-dehydrocholesterol reductase/NAD synthetase 1) and the 11p15 polymorphisms near CYP2R1 (cytochrome P450, subfamily IIR).

Associations at all three loci were confirmed in replication samples. The polymorphism at GC with the lowest p value in discovery samples, rs2282679, had a combined p value of 2.9×10⁻⁴⁸ in in-silico replication samples, with a consistent direction of effect. Additional genotyping was not done for this polymorphism. Polymorphism rs10741657 at CYP2R1 had a p value of 3.27×10⁻²⁰. Polymorphism rs10741657 region

Figure 1: Regional linkage disequilibrium plots for single nucleotide polymorphisms at GC (A), DHCR7/NADSYN1 (B), CYP2R1 (C), and CYP24A1 (D)

for vitamin D insufficiency: 25-hydroxyvitamin D concentrations lower than 75 nmol/L (30 ng/mL) and lower than 50 nmol/L (20 ng/mL). Covariates were age, sex, season, and BMI. We combined effect estimates from the logistic regression analysis across cohorts by meta-analysis using an inverse-variance weighting approach. We also did analyses using a 25 nmol/L (10 ng/mL) threshold, to examine whether genetic variants were associated with severe vitamin D deficiency.

Additionally, we constructed a genotype score by taking a weighted average of the number of risk alleles for members of a cohort, with weights established using β coefficients from the meta-analysis. Logistic regression was used to calculate the odds of vitamin D insufficiency according to quartile of the genotype score. For this analysis, we combined data from the Framingham Heart Study, TwinsUK, and 1958BC using a multivariate approach, with β coefficients for each quartile of genotype score meta-analysed jointly, as previously described.

Role of the funding source
The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results
Characteristics of the study cohorts are summarised in the webappendix (pp 15–17). Table 1 shows the results of genome-wide association analyses. In analysis of data from the five discovery samples, single nucleotide polymorphisms at three unique loci met the prespecified threshold for genome-wide significance: 4p12, 11q12, and 11p15. The 4p12 polymorphisms were within or near the GC gene, and the results included a non-synonymous polymorphism in this gene, rs7041. The 11q12 polymorphisms were near DHCR7/NADSYN1 (7-dehydrocholesterol reductase/NAD synthetase 1) and the 11p15 polymorphisms near CYP2R1 (cytochrome P450, subfamily IIR).

Associations at all three loci were confirmed in replication samples. The polymorphism at GC with the lowest p value in discovery samples, rs2282679, had a combined p value of 2.9×10⁻⁴⁸ in in-silico replication samples, with a consistent direction of effect. Additional genotyping was not done for this polymorphism. Polymorphism rs10741657 at CYP2R1 had a p value of 3.27×10⁻²⁰.
Table 3: Variants and risk of vitamin D insufficiency

<table>
<thead>
<tr>
<th>Genotype score</th>
<th>Quartile 1</th>
<th>Quartile 2</th>
<th>Quartile 3</th>
<th>Quartile 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio (95% CI)</td>
<td>p value</td>
<td>Odds ratio (95% CI)</td>
<td>p value</td>
</tr>
<tr>
<td>Quartile 1</td>
<td>1·0 (Reference)</td>
<td>–</td>
<td>1·0 (Reference)</td>
<td>–</td>
</tr>
<tr>
<td>Quartile 2</td>
<td>1·29 (1·15–1·46)</td>
<td>1·0 (0·97–1·25)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Quartile 3</td>
<td>1·56 (1·39–1·75)</td>
<td>1·38 (1·22–1·57)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Quartile 4</td>
<td>2·47 (2·20–2·78)*</td>
<td>1·92 (1·70–2·16)*</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

For individual variants, odds ratios are per copy of the risk allele. All logistic regressions were adjusted for age, sex, body-mass index, and season. *p values for trends in odds ratios for genotype scores were 2·3×10–48 for 25-hydroxyvitamin D concentrations lower than 75 nmol/L and 1·0×10–40 for lower than 50 nmol/L.

Discussion

Vitamin D insufficiency has been implicated in many musculoskeletal and extraskeletal diseases,12 which has led to substantial interest in the determinants of vitamin D status. Our findings establish a role for common genetic variants in regulation of circulating 25-hydroxyvitamin D concentrations. The presence of harmful alleles at the three confirmed loci more than doubled the risk of vitamin D insufficiency. These findings improve our understanding of vitamin D homeostasis and could assist identification of a subgroup of the white population who are at risk of vitamin D insufficiency.

DHCR7/NADSYN1 is a novel locus for association with vitamin D status, but one with compelling biological plausibility. DHCR7 encodes the enzyme 7-dehydrocholesterol (7-DHC) reductase, which converts 7-DHC to cholesterol, thereby removing the substrate from the synthetic pathway of vitamin D₃, a precursor of 25-hydroxyvitamin D₃. Rare mutations in DHCR7 lead to Smith-Lemli-Opitz syndrome, which is characterised by reduced activity of 7-DHC reductase, accumulation of 7-DHC, low cholesterol, and many congenital abnormalities.25 Mutations in DHCR7 might also confer a competitive advantage to heterozygous carriers, because high concentrations of 7-DHC could provide protection against rickets and osteomalacia from hypovitaminosis D.26 However, few data exist for vitamin D status in individuals with Smith-Lemli-Opitz syndrome or carriers of mutations.7 The finding that common variants at DHCR7 are strongly associated with circulating 25-hydroxyvitamin D concentrations suggests that this enzyme could have a larger role in regulation of vitamin D status than has previously been recognised.

The gene at the second locus, CYP2R1, encodes a hepatic microsomal enzyme. CYP2R1 could be the enzyme underlying 25-hydroxylation of vitamin D in the liver, as reflected by the association with vitamin D binding protein (p=4·0×10–42), with the minor allele related to reduced protein concentrations.
A few investigators have reported associations (including 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D) that binds and transports vitamin D and its metabolites which is a 52–59 kDA protein synthesised in the liver, but this suggestion is uncertain because many other enzymes with 25-hydroxylase activity in vitro have been described.9 Previous clinical studies have been limited to a case report of a Nigerian man with a point mutation in CYP2R1 who had a history of rickets,28 and a previous candidate gene study in 133 individuals with type 1 diabetes.16 Because affected individuals with CYP2R1 polymorphisms have been difficult to identify, redundancy in the enzymes involved in the 25-hydroxylation step has been proposed. Thus, our finding that common variants at the CYP2R1 locus are associated with circulating 25-hydroxyvitamin D concentrations is the strongest evidence so far that CYP2R1 is the enzyme underlying the crucial first step in vitamin D metabolism.

The third gene, GC, encodes vitamin D binding protein, which is a 52–59 kDA protein synthesised in the liver that binds and transports vitamin D and its metabolites (including 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D).29 A few investigators have reported associations between non-synonymous single nucleotide polymorphisms in this gene6,12,13,16 and 25-hydroxyvitamin D concentrations. However, their studies were small (<1500 participants) and results were not replicated. The most widely studied GC variants are the non-synonymous polymorphisms rs7041 (Asp→Glu) and rs4588 (Thr→Lys). The previous nomenclature for GC haplotypes (GC1S, GC1F, and GC2) was based on specific combinations of alleles at these non-synonymous polymorphisms.13 Our data strongly confirm the association of rs7041 with circulating 25-hydroxyvitamin D. The other variant, rs4588, is not in the HapMap dataset and is thus not part of our imputed results. However, rs4588 is only 11 bp away from rs7041, and direct genotyping of rs4588 in one of our samples (TwinsUK) confirms that it is in linkage disequilibrium (r²>0.99) with several associated variants from our genome-wide association study.

We also showed that GC variants associated with low 25-hydroxyvitamin D concentrations were strongly related to reduced concentrations of vitamin D binding protein. Whether variation in the amount of circulating binding protein affects metabolism and availability of vitamin D is not well established. Concentrations of the binding protein have been postulated to affect delivery of 25-hydroxyvitamin D and activated vitamin D (1,25-dihydroxyvitamin D) to target organs, as well as clearance of vitamin D metabolites from the circulation.25,26 Alternatively, changes in quantity or function of the binding protein could be accompanied by changes in the relative proportions of free and bound 25-hydroxyvitamin D, with the free proportion being the potential rate-limiting factor for 1,25-dihydroxyvitamin D production. Further studies are needed to assess the effects of variation in serum concentrations of vitamin D binding protein.

In a screen of candidate gene variants, we noted an additional association at the locus containing CYP24A1 that was genome-wide significant in pooled analyses of the discovery and replication samples. CYP24A1 encodes 24-hydroxylase, which initiates degradation of both 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D. In previous candidate gene and linkage studies, investigators have not shown an association of variants at this locus with 25-hydroxyvitamin D concentrations, but these studies were relatively small.30,31

A high genotype score for the three variants identified in our genome-wide association study conferred roughly a two-fold increase in risk of vitamin D insufficiency (25-hydroxyvitamin D concentrations <50 nmol/L or <75 nmol/L) compared with a score in the lowest quartile, after we accounted for environmental factors. This result suggests that variation at a few genetic loci could have a clinically important effect on risk of vitamin D insufficiency. High genotype score was associated with a 1.4-fold raised risk of severe vitamin D deficiency (<25 nmol/L). Whether the reduced odds ratio for the 25 nmol/L threshold shows an increased contribution of environmental factors to the most severe forms of vitamin D deficiency is unclear, because severe deficiency was rare in our community-based cohorts.

Whether genetic predisposition modifies response to sun exposure or dietary supplementation warrants further study, especially in view of the large interindividual differences that have been reported in response to treatment with identical doses of vitamin D.33 Furthermore, these variants might provide useful genetic approaches to investigate the role of vitamin D insufficiency in several chronic diseases with which this disorder has been epidemiologically linked.

The validity of our findings is lent support by the large study sample (more than 30 000 participants combined in discovery and validation samples), consistent results across several standard assays for 25-hydroxyvitamin D, and the strong biological plausibility of genes at the principal loci. Several limitations of the study also deserve mention, however. The study was not designed...
to identify uncommon or rare variants. Resequencing at selected loci, partly on the basis of our results, could be used to identify uncommon variants with potentially large effects.

We used a multistage design to achieve maximum homogeneity of the assays used in the discovery analyses. We might have identified more genome-wide significant associations had we combined all study cohorts into one stage, but we would not have had a large replication sample. Other factors that might have contributed to reduced statistical power are second-order interactions (eg, with age) and the use of a stringent p-value threshold in the discovery stage.\textsuperscript{14} Accordingly, the absence of specific candidate genes, such as those affecting vitamin D action or skin pigmentation, from our most significant results does not exclude an effect of genetic variation at these loci on vitamin D concentrations, but their contributions might be small compared with those of the genes that we identified.

Assays used to measure 25-hydroxyvitamin D concentrations varied between cohorts. To keep potential variability introduced by cohort-specific measurement techniques to a minimum, we standardised 25-hydroxyvitamin D concentrations within cohorts and analysed this variable as a continuous trait. Furthermore, primary results were meta-analysed with a Z-score-weighted approach, which is not scale-dependent.

Specific information about dietary intake and sunlight exposure was not available from all cohorts. Such factors probably contribute to non-genetic variability in 25-hydroxyvitamin D concentrations, which would reduce the effect noted in our analyses.

The single nucleotide polymorphisms that we have identified might not be causal variants, but rather be in linkage disequilibrium with these variants. We did not examine downstream markers of vitamin D status, because 25-hydroxyvitamin D concentration is regarded as the most reliable indicator of vitamin D status. Other molecules, such as 1,25-dihydroxyvitamin D or parathyroid hormone, have greater intraindividual variability than does 25-hydroxyvitamin D and are affected by several determinants other than vitamin D status. Lastly, we studied only white individuals of European descent. Whether the genetic variants we identified affect vitamin D status in other racial or ethnic groups is unknown and warrants further study.

\textbf{Contributors}

JD, FZ, JBR, BK, JBM, DB, CO, DLK, JDC, PFO, NLG, IV, JS, MM, BK, KR, ML(2), ML(2), LJH, ADH, GZ, RJFL, and TF took part in data analysis. TJW, JBR, BK, JBM, DB, DPK, CO, MRJ, FR, DG, NJW, NKA, CC, ALH, ED, CP, NS, ML(2), TBH, AH, AGU, LP, DK, SBK, JCF, JAT, EH, TDS contributed to study design. TJW, FZ, DB, DPK, EAS, CO, DLK, MRJ, PFO, DKH, IV, MP, HIM, PA, MM, DJS, GLB, DG, NH, NW, DH, NKA, CC, WDF, GC, HMM, RJFL, DMR, AH, ED, YL, CP, HES, Lj, NS, J, TBH, JDC, LL, JAC, LP, DSS, MJE, SBK, JCF, JAT, EH, and TDS read the manuscript critically. The writing group consisted of TJW, FZ, JBR, BK, EAS, DLK, MRJ, MG, JCF, JAT, JD, EH, and TDS. ML(1)=Martin Ladouceur. ML(2)=Mattias Lorentzon.

\textbf{Conflicts of interest}

TJW has served on the scientific advisory board of Diasorin. DKH has received honoraria from Abbott Nutrition. MW has received consultancy fees, Honoraria, and speakers’ fees from Abbott and Genzyme. DMR has acted as a consultant for Novartis, Roche, Pfizer, Amgen, Shire, Merck, and Servier, has received speakers’ fees from Novartis, Roche, and Amgen, and owns stock in GlaxoSmithKline. NJW has received lecture fees from Novartis and Sanoﬁ-Aventis. All other authors declare that they have no conflicts of interest.

\textbf{Acknowledgments}

\textit{Framingham Heart Study} The Framingham Heart Study of the National Heart, Lung and Blood Institute (NHLBI) of the US National Institutes of Health (NIH) and Boston University School of Medicine is supported by the NIH/NHLBI contract N01-HC-23595. The present study received support from the American Heart Association, the US Department of Agriculture, Agricultural Research Service (under Cooperative Agreement No 58-1950-7-707), and the National Institute of Aging (AG14759). DK was supported by a grant from the National Institute of Arthritis, Musculoskeletal, and Skin Diseases and the National Institute on Aging (R01 AR/AG 41398). The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource project. This work was partly supported by a contract with Affymetrix Inc for genotyping services (Contract No N02-HL-6-4278). A portion of this research used the Linux Cluster for Genetic Analysis, which is funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Twins\textit{UK and Chingford} The study was funded by the Wellcome Trust, Arthritis Research Campaign, European Community’s Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2-2008-201865-CEFOS and Seventh Framework Programme grant 200800 Treat OA (FP7/2007-2013). ENSAGE project grant agreement HEALTH-F4-2007-201413, and the FP 5 Genome E loneliness Project (QLG2-CT-2002-01254). The study also receives support from the UK Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy’s & St Thomas’ NHS Foundation Trust in partnership with King’s College London. TDS is an NIHR senior investigator. The project also received support from a Biotechnology and Biological Sciences Research Council project grant (G20234). The authors acknowledge the funding and support of the National Eye Institute (NEI) via an NIH/Center for Inherited Disease Research (CIDR) genotyping project. We thank the staff from the Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation, quality control, and genotyping; Le Centre National de Genotypage, France, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki. Genotyping was also done by CIDR as part of an NEI/NIH project grant. The \textit{Rotterdam Study} This study was funded by the Netherlands Organization of Scientific Research (NWO) Investments (017.5.010.2005.011. 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative/NWO project 050-060-810, and the European Commission (HEALTH-F2-2008-201865-GEFOS, and HEALTH-F2-2008-00-TREAT-OA). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam. Netherlands Organization for the Health Research and Development, the Research Institute for Diseases in the Elderly, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. We thank the staff from the Rotterdam Study, and the participating general practitioners and pharmacists.
1958 British Birth Cohort The project was funded by the UK Medical Research Council (MRC) (project grant G0601653), and 25-hydroxyvitamin D assays by the BUPA Foundation. EH is funded by a UK Department of Health Public Health Career Scientist Award. Use of DNA from this cohort was funded by MRC grant G0000934 and Wellcome Trust grant 068545/2/02. This research used resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062438. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of investigators who contributed to generation of the data is available from the Wellcome Trust Case-Control Consortium website. Funding for the project was provided by the Wellcome Trust under award 076113. The MRC Centre of Epidemiology for Child Health is funded by the MRC: Great Ormond Street Hospital/University College London, Institute of Child Health receives a proportion of funding from the Department of Health’s NHIR (‘Biomedical Research Centres’ funding).

Health, Aging and Body Composition Study This research was supported by the Intramural Research Program of the NIH National Institute on Aging and National Institute on Aging on contracts N01-AG-6-2101, N01-AG-6-2103, and N01-AG-6-2106. Assessment of 25-hydroxyvitamin D concentrations was funded by a National Institute on Aging grant, R01-AG029364. The genome-wide association study was funded by a National Institute on Aging grant, R01-AG029098, and genotyping services were provided by CIDR. CIDR is fully funded through a federal contract from NIH to The Johns Hopkins University (contract number HHSN268200782096C).

The Amish Family Osteoporosis Study The Amish Family Osteoporosis Study was funded by a grant from the National Institute of Arthritis, Musculoskeletal and Skin Diseases (N01-AR-46838).

Gothenburg Osteoporosis and Obesity Determinants (GOOD) Study Financial support was received from the Swedish Research Council, the Swedish Foundation for Strategic Research, The AFA/LUF research grant in Gothenburg, the Lundberg Foundation, the Emil and Vera Cornell Foundation, the Torsten and Ragnar Söderbergs Foundation, Petrus and Augusta Hedlunds Foundation, the Västra Götaland Foundation, the Göteborg Medical Society, and the Sahlgrenska Center for Cardiovascular and Metabolic Research (CMR, no A05:188), which is supported by the Swedish Foundation for Strategic Foundation.

Study of Indiana Women This work was supported by NIH grants P01 AG-18197 and M01 RR-00750. Genotyping services were provided by CIDR. CIDR is fully funded through a federal contract from NIH to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Library of Medicine.

North Finland Birth Cohort 1966 The project was funded by the MRC components funded by a grant (project grant G0601653). Financial support was received from the Academy of Finland (project grants 104781, 120315, 1114194 and Center for Excellence in Complex Disease Genetics, University Hospital Oulu, Biocenter, University of Oulu, Finland, NHLBI grant 350H1L087679-02 through the STAMPEDE program (IRLM10813268-01), ENGAGE project and grant agreement HEALTH-F4-2007-201413, MRC (studentship grant G0500539, centre grant G000705), the Wellcome Trust (project grant GR069224), UK. DNA extractions, sample quality controls, biobank up-keeping, and aliquotting were done at the National Public Health Institute, Biomedicalc Helsinki, Finland and supported financially by the Academy of Finland and Biocenterm Helsinki.

Cardiovascular Health Study Research for this cohort that was reported in this article was supported by contract numbers N01-HC-83079 to N01-HC-85086, N01-HC-35129, N01-HC-15103, N01-HC-55222, N01-HC-75150, N01-HC-45131, grants numbers U01 HL082985, R01 HL084443, R01 HL087652, and R01 AG027002 from NHLBI, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal investigators and institutions is available at the Cardiovascular Health Study website.

DNA handling and genotyping was supported in part by National Center for Research Resources grant M01RR00069 to the Cedars-Sinai General Clinical Research Center Genotyping core and NIDDK grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

Hertfordshire This study was funded by the UK MRC and the Arthritis Research Campaign.

Allenton Prospective Osteoporosis Screening Study Funding for cohort sample collection and analysis was supported in part by grants from the European Commission (QLRT-2001-02629) and the UK Food Standards Agency. We thank the clinical research staff in the Bone and Musculoskeletal Research Programme for all their hard work, and all the women who participated in the study.

Canadian Multicentre Osteoporosis Study CaMoS is funded by the Canadian Institutes of Health Research (CIHR). JBR is a clinical investigator of the CIHR and an Osteoporosis Canada New Investigator. This work was supported by grants to JBR from the CIHR and Canadian Foundation for Innovation. We acknowledge the support and funding of the Public Health Agency of Canada to assess vitamin D levels, the expert technical support of Kurtis Sarafin and Jenn Kreiger of Health Canada (vitamin D measurements) and the input of Linda S Greene-Finestone and Ross Duncan.

Other The Cambridge Institute for Medical Research is in receipt of a Wellcome Trust Strategic Award (079895). This work was also supported by JDRF, the Wellcome Trust, and the NHIR Cambridge Biomedical Centre.

References


