Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels

Xia Jiang et al.#

Vitamin D is a steroid hormone precursor that is associated with a range of human traits and diseases. Previous GWAS of serum 25-hydroxyvitamin D concentrations have identified four genome-wide significant loci (GC, NADSYN1/DHCR7, CYP2R1, CYP24A1). In this study, we expand the previous SUNLIGHT Consortium GWAS discovery sample size from 16,125 to 79,366 (all European descent). This larger GWAS yields two additional loci harboring genome-wide significant variants ($P = 4.7 \times 10^{-9}$ at rs8018720 in SEC23A, and $P = 1.9 \times 10^{-14}$ at rs10745742 in AMDHD1). The overall estimate of heritability of 25-hydroxyvitamin D serum concentrations attributable to GWAS common SNPs is 7.5%, with statistically significant loci explaining 38% of this total. Further investigation identifies signal enrichment in immune and hematopoietic tissues, and clustering with autoimmune diseases in cell-type-specific analysis. Larger studies are required to identify additional common SNPs, and to explore the role of rare or structural variants and gene–gene interactions in the heritability of circulating 25-hydroxyvitamin D levels.

Correspondence and requests for materials should be addressed to E.H. (email: elina.hypponen@unisa.edu.au) or to P.K. (email: pkraft@hsph.harvard.edu) or to D.P.K. (email: kiel@hsl.harvard.edu)

#A full list of authors and their affiliations appears at the end of the paper
**Results**

**Study description and GWAS.** This study represents an expansion of our previous SUNLIGHT consortium GWAS. Here, we combine the 5 discovery cohorts and in-silico replication cohorts from this study, and augment these with 21 additional cohorts that have joined the SUNLIGHT consortium since 2010 (study characteristics are described in Supplementary Table 1, Supplementary Note 1). In contrast to the previous meta-analysis which involved discovery, in-silico and de-novo genotyping stages, we performed a first stage discovery meta-analysis on a total of up to 79,366 individuals and replicated novel findings in two independent separate in-silico data sets (40,562 individuals collected by EPIC and 2195 individuals collected by SOCCS). To assess and control for population stratification, we examined QQ-plots and genomic control in conditional and joint analysis, but not for the NADSYN1/DHCR7, CYP24A1, and SEC23A loci where only one primary associated SNP was identified (Supplementary Table 2).

**SNP by dietary vitamin D intake interaction.** In addition to performing the marginal effect meta-analysis using all samples, we also tested a model with SNPs and dietary vitamin D intake as main effects and a term for their interaction in a subset of samples. Diet questionnaires, including vitamin D intake, were available for a subset of 13 cohorts and an additional 2 cohorts that were not included in the overall meta-GWAS analysis (a total of 15 cohorts, N=41,981). We performed a GWAS explicitly allowing for an interaction between vitamin D intake and SNP genotypes, in which dietary vitamin D was coded as a continuous variable. We performed two tests: (i) a 1 degree-of-freedom interaction test between each SNP and vitamin D intake, and (ii) a 2 degree-of-freedom joint test of main genetic and interaction effects. However, for comparison purposes, we also performed, (iii) a standard test of marginal genetic effect after adjusting for vitamin D intake in the same sub-samples (Supplementary Fig. 3, and Supplementary Fig. 4). The marginal genetic effect analyses confirmed existing association signals at GC (lead SNP rs2282679, in complete linkage disequilibrium with the lead SNP rs3759567 identified from meta-GWAS using all individuals), CYP2R1, NADSYN1/DHCR7 (lead SNP rs4944062, in complete linkage disequilibrium with the lead SNP rs12785878 identified from meta-GWAS using all individuals) and CYP24A1, as well as the novel association at AMHD1 (P = 5.7×10^{-9}, Supplementary Table 3). The joint analysis also identified the five genes above, but with less significant P-values. For instance, the association at AMHD1 achieved only suggestive genome-wide significance (P = 1.2×10^{-7}, Table 2). The interaction test did not identify any variants at genome-wide significance level. Among the 5 SNPs significant in marginal effect tests, the lead SNP in CYP2R1 showed nominal significance for interaction with dietary vitamin D intake (rs10741657, P = 0.028), but no interactions were observed for the other SNPs (rs2282679, P = 0.45; rs4944062, P = 0.74; rs10745742, P = 0.64; rs17216707, P = 0.46). Repeating the analysis using a tertile coding for vitamin D instead of a continuous coding did not qualitatively change the results.

**SNP-heritability of 25-hydroxyvitamin D.** We further evaluated the SNP-heritability, defined as the heritability explained by GWAS SNPs of 25-hydroxyvitamin D, using LD score regression
The overall observed heritability of 25-hydroxyvitamin D estimated by using all common SNPs (2,579,296 after QC) was 7.54% (standard error (SE): 1.88%). After excluding genome-wide significant SNPs (533 SNPs with $P \leq 5 \times 10^{-8}$) from six loci and all SNPs within ±500 kb of those loci, the heritability decreased to 4.70% (SE: 0.72%). The estimate further decreased to 1.73% (SE: 0.32%) after excluding all SNPs that reached nominal significance (156,675 SNPs with $P \leq 0.05$). These results indicate that common variants tagged by GWAS chips explain a modest fraction of overall variability in circulating 25-hydroxyvitamin D.

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

| Gene             | SNP        | Chromosome: Position | Effect/reference allele | Allele frequency | Meta-GWAS estimates |
|------------------|------------|----------------------|-------------------------|-----------------|---------------------|
|                  |            |                      |                         |                 | Effect (Beta)       |
|                  |            |                      |                         |                 | Standard Error      |
|                  |            |                      |                         |                 | $P$-value           |
| First stage discovery meta-GWAS ($N = 79,366$) | GC         | rs3755967 4:72828262  | T/C                  | 0.28            | -0.089 | 0.0023  | 4.74E-343 |
|                  | NADSYN1/ DHCR7 | rs12785878 11:70845097 | T/G                  | 0.75            | 0.036  | 0.0022  | 3.80E-62  |
|                  | CYP2R1     | rs10741657 11:14871454 | A/G                  | 0.4             | 0.031  | 0.0022  | 2.05E-46  |
|                  | CYP24A1    | rs17216707 20:52165769 | T/C                  | 0.79            | 0.026  | 0.0027  | 8.14E-23  |
|                  | AMDHD1     | rs10745742 12:94882660 | T/C                  | 0.4             | 0.017  | 0.0022  | 1.88E-14  |
|                  | SEC23A     | rs8018720 14:38625936 | C/G                  | 0.82            | -0.017 | 0.0029  | 4.72E-09  |
| Replication data set 1: samples collected by EPIC ($N = 40,562$) | AMDHD1    | rs10745742 12:94882660 | T/C                  | 0.41            | 0.041  | 0.0071  | 1.21E-08  |
|                  | SEC23A     | rs8018720 14:38625936 | C/G                  | 0.83            | -0.032 | 0.0093  | 5.24E-04  |
| Replication data set 2: additional control samples collected by SOCCS ($N = 2195$) | AMDHD1    | rs10745742 12:94882660 | T/C                  | 0.37            | 0.045  | 0.021   | 0.03     |
|                  | SEC23A     | rs8018720 14:38625936 | C/G                  | 0.81            | -0.051 | 0.026   | 0.04     |
| Pooled analysis (discovery meta-GWAS + replication 1 + replication 2) ($N = 122,123$) | AMDHD1    | rs10745742 12:94882660 | T/C                  | 0.39            | 0.019  | 0.002   | 2.10E-20  |
|                  | SEC23A     | rs8018720 14:38625936 | C/G                  | 0.82            | -0.019 | 0.0027  | 1.11E-11  |

In the pooled analysis, $P_{\text{heterogeneity}} = 0.003$ for AMDHD1 rs10745742; $P_{\text{heterogeneity}} = 0.14$ for SEC23A rs8018720.

### Fig. 1 Genome-wide association of circulating 25-hydroxyvitamin D graphed by chromosome positions and −log10 $P$-value (Manhattan plot) and quantile-quantile plot of all SNPs from the meta-analysis (QQ-plot). a Manhattan plot: The $P$-values were obtained from the single stage fixed-effects inverse variance weighted meta-analysis. The Y axis shows $−\log_{10} P$-values, and the X axis shows chromosome positions. Horizontal gray dash line represents the thresholds of $P = 5 \times 10^{-8}$ for genome-wide significance. Known loci were colored coded as red, and novel loci were color coded as green. b QQ-plot: The Y axis shows observed $−\log_{10} P$-values, and the X axis shows the expected $−\log_{10} P$-values. Each SNP is plotted as a black dot, and the dash line indicates null hypothesis of no true association. Deviation from the expected $P$-value distribution is evident only in the tail area, with a lambda of 0.99, suggesting that population stratification was adequately controlled.
Table 2 Results from the SNP-by-dietary vitamin D intake interaction analysis

| Gene           | SNP            | Chromosome: Position | Effect/Reference Allele | Allele Frequency | SNP-by-dietary vitamin D intake Interaction analysis |
|---------------|----------------|----------------------|-------------------------|------------------|---------------------------------------------------|
|               |                |                      |                         |                  | Main Genetic Effect | Interaction Effect | P-value for interaction | P-value for joint test |
| First stage discovery meta-GWAS (N=79,366) |               |                      |                         |                  | Effect (Beta_G) | Standard Error | Effect (Beta_Int) | Standard Error |                                         |
| GC            | rs3755967      | 4:72828262           | T/C                     | 0.28              | −0.082           | 0.0042       | −2.01E-05     | 1.67E-05     | 0.23          | 2.92E-171                          |
|               | rs2282679*     | 11:70845097          | T/G                     | 0.28              | 0.085            | 0.004        | 1.20E-05     | 1.60E-05     | 0.45          | 1.40E-187                          |
| NADSYN1/DHCR7 | rs12785878     |                      | T/G                     | 0.75              | 0.033            | 0.0039       | 6.61E-06     | 1.62E-05     | 0.68          | 3.52E-29                           |
|               | rs4944062*     | 11:70864942          | T/G                     | 0.75              | 0.034            | 0.004        | 5.30E-06     | 1.60E-05     | 0.74          | 1.90E-31                           |
| CYP2R1        | rs10741657     | 11:14871454          | A/G                     | 0.4               | 0.03             | 0.0035       | 3.21E-05     | 1.46E-05     | 0.028         | 2.23E-38                           |
| CYP2RA1       | rs17216707     | 20:52165769          | T/C                     | 0.79              | 0.025            | 0.0048       | 1.39E-05     | 1.88E-05     | 0.46          | 1.32E-14                           |
| AMHD1         | rs10745742     | 12:94882660          | T/C                     | 0.4               | 0.016            | 0.0036       | −7.05E-06    | 1.49E-05     | 0.64          | 1.20E-07                           |
| SEC23A        | rs8018720      | 14:38625936          | C/G                     | 0.82              | −0.013           | 0.0051       | −2.40E-05    | 2.06E-05     | 0.24          | 1.94E-05                           |

* Top SNPs identified in the SNP-by-dietary vitamin D intake interaction analysis, performed in a subset of individuals. For GC and NADSYN1/DHCR7, the top SNPs identified through the marginal effect regression meta-analysis using all individuals were in high linkage disequilibrium with the top SNPs identified through the SNP-by-dietary vitamin D intake interaction analysis using a subset of individuals (rs for rs3755967 and rs2282679; 1:0; rs for rs12785878 and rs4944062; 1:0). Beta_G indicates the main effect of the SNP. Beta_Int indicates the interaction effect of SNP-by-dietary vitamin D intake.
HDL (Supplementary Table 6). However, with only six 25-hydroxyvitamin D associated SNPs included in the analysis, we consider an overall null directional correlation as our main finding, and further well-designed large-scale Mendelian randomization analyses are warranted.

Finally, we analyzed the 220 cell-type-specific annotations in each of the 37 traits and compared the cell-type-specific enrichments for 25-hydroxyvitamin D to the enrichments for these traits. The enrichment pattern for 25-hydroxyvitamin D differed notably from the patterns for psychiatric diseases and metabolic related traits. Psychiatric diseases showed enrichment for histone marks specific to CNS cell types, and metabolic diseases showed enrichment for gastrointestinal cell types, while these annotations were depressed in 25-hydroxyvitamin D. Conversely, 25-hydroxyvitamin D showed similar patterns with autoimmune inflammatory diseases, where multiple immune cell types were enriched. We consistently observed that 25-hydroxyvitamin D was clustered with autoimmune diseases (Supplementary Fig. 5).

**Discussion**

Vitamin D inadequacy has been linked to many diseases such as cancer, autoimmune disorder and cardiovascular conditions in addition to musculoskeletal diseases, which has led to substantial interest in the determinants of vitamin D status, especially its genetic components. We have performed a large 25-hydroxyvitamin D meta-GWAS involving 31 studies with a total of 79,366 individuals. Our results recapitulated several previously reported findings. First of all, we confirmed the role for common genetic variants in regulation of circulating 25-hydroxyvitamin D concentrations. Our study validated three loci, GC, NADSYN1/DHCR7, and CYP2R1, all were established 25-hydroxyvitamin D risk loci identified through two earlier GWASes. In addition, we were able to confirm the association of a locus containing CYP24A1 with 25-hydroxyvitamin D concentrations using our large sample size, which highlights the importance of this protein in the degradation of vitamin D molecule, by catalyzing hydroxylation reactions at the side chain of 1,25-dihydroxyvitamin D, the physiologically active form (hormonal form) of vitamin D. Significant finding at this locus was only shown in the pooled analyses involving both discovery and replication samples in an earlier GWAS.

We extended previously reported findings by identifying two additional new loci. SEC23A (Sec23 Homolog A, coat protein complex II (COPPII) component) encodes a member of SEC23 subfamily. In eukaryotic cells, secreted proteins are synthesized in the endoplasmic reticulum (ER), packaged into COPPII-coated vesicles, and traffic to the Golgi apparatus. As part of COPPII complex, SEC23 plays a role in promoting ER-Golgi protein trafficking. SEC23A mutations have been reported to cause craniofacial abnormalities, craniofacial dysplasia, a disease characterized by craniofacial and skeletal malformation such as delay in closure of fontanelles, sutural cataracts and facial dysmorphism, due to defective collagen secretion. The second novel locus is AMDHD1 (amidohydrolase domain containing 1). This gene encodes an enzyme involved in the histidine, lysine, phenylalana nine, tyrosine, proline and tryptophan catabolic pathway. Mutations in AMDHD1 are found to be associated with atypical lipomatous tumor, a cancer of connective tissues that resemble fat cells.

Our SNP-heritability results suggest that 25-hydroxyvitamin D has a modest overall heritability due to common genome-wide SNPs of 7.5%, and that an appreciable proportion (2.84% out of 7.5%, i.e., 38%) of this total could be explained by known genetic regions identified through GWAS. Our findings are in line with a previous published report (by Hiraki et al.) which estimated the variance in circulating 25-hydroxyvitamin D explained by SNPs in a total of 5,575 individuals. According to that report, by employing a linear mixed model fitting the additive genetic matrix created from all genotyped and imputed SNPs, the proportion of variance explained was 8.9%; by employing a polygenic score approach comprised of the then GWAS-discovered SNPs (GC, CYP2R1, DHCR7/NADSYN1), the proportion of variance explained was 5%. Both of these estimates were close to ours. In Hiraki et al., the known 25-hydroxyvitamin D associated environmental factors such as age, BMI, season of blood drawn, vitamin D dietary intake, vitamin D supplement intake, region of residence and ethnicity, explained ~18% of the observed variance. Our results, in agreement with these findings, suggest that although there appears to be some polygenic signals outside of the identified regions, the remaining common effects may be small. There also may be low frequency variants with larger effects that were not investigated here. For example, while this paper was under review, a related study identified low-frequency (MAF = 2.5%) synonymous coding variant rs117913124_A at CYP2R1 conferring a large effect on 25-hydroxyvitamin D levels, which was four times greater in magnitude and independent of a previously described association for a common variant (rs10741657) near CYP2R1.

Results of twin and familial studies have revealed a substantial genetic basis in the variability of circulating 25-hydroxyvitamin D levels, with estimates of heritability reaching as high as 86%, 90%, 17%, 19%. These estimates, however, seem to be influenced by environmental conditions. For example, in a study conducted by Orton et al. with 40 monozygotic and 59 dizygotic twin pairs, bloods were collected at the end of winter and a heritability of 77% was reported. Similarly, the study conducted by Karohl et al. with 310 monozygotic and 200 dizygotic male twins observed a heritability of 70% during winter, whereas in summer, serum 25-hydroxyvitamin D concentrations appeared to be

| Table 3 Heritability enrichment of ten grouped cell types |
|---------------------------------------------------------|
| **Category** | Proportion of SNPs (%) | Proportion of \( \lambda^2 \) (%) | Enrichment (standard errors) | \( P \)-value |
|-------------|------------------------|-------------------------------|-------------------------------|--------|
| Kidney      | 4.26                   | 27.27                         | 6.4 (2.44)                    | 0.027  |
| Liver       | 7.22                   | 37.68                         | 5.22 (1.55)                   | 0.01   |
| Gastrointestinal | 16.77               | 72.88                         | 4.35 (0.97)                   | 0.0017 |
| Immune and hematopoietic | 23.34           | 100.17                        | 4.29 (0.76)                   | 2.20E-05 |
| Central nervous system | 14.88         | 54.09                         | 3.64 (0.87)                   | 0.0039 |
| Cardiovascular | 11.11               | 35.74                         | 3.22 (1.26)                   | 0.078  |
| Connective tissue/bone | 11.5            | 35.65                         | 3.1 (1.04)                    | 0.037  |
| Adrenal/pancreas | 9.36               | 26.17                         | 2.8 (1.31)                    | 0.18   |
| Other       | 20.27                  | 56.68                         | 2.8 (0.98)                    | 0.076  |
| Skeletal Muscle | 10.38            | 14.29                         | 1.38 (1.25)                   | 0.76   |

Black bold font indicates significant \( P \)-values after multiple corrections (\( P < 0.05/10 \)).
Fig. 3 Genetic correlations between 25-hydroxyvitamin D and 37 traits. We collected GWAS summary statistics of 37 diseases and traits spanning a wide range of phenotypes (autoimmune inflammatory diseases, psychiatric disorders, metabolic traits, and anthropometric index) from publicly available resources, and estimated their shared genetic similarities with serum 25-hydroxyvitamin D levels. We plotted the genetic correlation together with 95% confidence intervals using a blue square and gray horizontal lines. Red vertical line indicates no genetic correlation (r = 0). Statistical significance was defined as P-value <0.05. None of the pairwise correlations passed Bonferroni corrections.

| Traits                                      | Genetic correlation (95%CI) | p-value | Significance |
|---------------------------------------------|----------------------------|---------|--------------|
| Autoimmune/inflammatory diseases            |                            |         |              |
| Age-related macular degeneration            | −0.02 (−0.1, 0.06)         | 0.57    |              |
| Celiac Disease                              | −0.02 (−0.17, 0.15)        | 0.89    |              |
| Crohn’s Disease                             | −0.07 (−0.19, 0.05)        | 0.25    |              |
| Lupus                                       | 0 (−0.15, 0.14)            | 0.98    |              |
| Multiple Sclerosis                          | −0.17 (−0.45, 0.1)         | 0.21    |              |
| Primary biliary cirrhosis                   | −0.18 (−0.33, −0.03)       | 0.019   | *            |
| Rheumatoid Arthritis                        | 0.05 (−0.08, 0.19)         | 0.44    |              |
| Ulcerative Collitis                         | 0.04 (−0.1, 0.18)          | 0.59    |              |
| UKBiobank Asthma                            | −0.01 (−0.11, 0.1)         | 0.9     |              |
| UKBiobank Eczema                            | −0.01 (−0.12, 0.1)         | 0.91    |              |
| Inflammatory Bowel Disease                  | −0.02 (−0.13, 0.09)        | 0.71    |              |
| Alzheimer’s                                 | −0.08 (−0.28, 0.11)        | 0.39    |              |
| Psychiatric disorder/traits                 |                            |         |              |
| Anorexia                                    | −0.03 (−0.15, 0.1)         | 0.7     |              |
| Autism                                      | 0.03 (−0.11, 0.17)         | 0.64    |              |
| Bipolar Disorder                            | 0.04 (−0.11, 0.18)         | 0.63    |              |
| Depressive Syndrome                         | 0.04 (−0.08, 0.17)         | 0.49    |              |
| Neuroticism                                  | 0.02 (−0.07, 0.11)         | 0.61    |              |
| Schizophrenia                               | 0 (−0.08, 0.08)            | 1       |              |
| Subject Well Being                          | 0.14 (−0.1, 0.38)          | 0.25    |              |
| Metabolism related traits                   |                            |         |              |
| Coronary Artery Disease                     | −0.03 (−0.16, 0.1)         | 0.66    |              |
| Fasting Glucose                             | 0 (−0.12, 0.12)            | 0.98    |              |
| HDL                                         | 0.07 (−0.04, 0.19)         | 0.21    |              |
| LDL                                         | 0.06 (−0.07, 0.18)         | 0.36    |              |
| Triglycerides                               | −0.06 (−0.16, 0.03)        | 0.18    |              |
| Type 2 Diabetes                             | −0.08 (−0.22, 0.06)        | 0.26    |              |
| UKBiobank Hypertension                      | 0.02 (−0.05, 0.1)          | 0.54    |              |
| Others                                      |                            |         |              |
| Ever Smoked                                 | −0.17 (−0.32, −0.03)       | 0.019   | *            |
| UKBiobank Age at Menarche                   | −0.07 (−0.02, 0.15)        | 0.13    |              |
| UKBiobank Age at Menopause                  | −0.07 (−0.19, 0.05)        | 0.25    |              |
| UKBiobank BMI                               | −0.03 (−0.11, 0.05)        | 0.45    |              |
| UKBiobank Diastolic                         | 0.02 (−0.06, 0.09)         | 0.63    |              |
| UKBiobank FEV1/FVC                          | −0.01 (−0.08, 0.06)        | 0.69    |              |
| UKBiobank FVC                               | 0.14 (0.04, 0.23)          | 0.0036  | *            |
| UKBiobank Waist-Hip Ratio                   | −0.02 (−0.09, 0.05)        | 0.64    |              |
| UKBiobank Height                            | 0.06 (−0.01, 0.12)         | 0.078   |              |
| UKBiobank Systolic                          | −0.01 (−0.08, 0.06)        | 0.8     |              |
| UKBiobank Waist-Hip Ratio                   | −0.1 (−0.2, 0)             | 0.042   | *            |

entirely determined by non-genetic factors (heritability: 0%)⁹. Comparable estimates were also identified in a slightly larger study conducted by Mills et al. (winter: 90% vs. summer: 56%)¹⁰. Consistent with season dependency, sex differences were also observed (males: 86% vs. females: 17%)¹¹. While these estimates should be treated with caution due to small samples and related imprecision, they confirm the substantial variation in 25-hydroxyvitamin D levels by season (as shown previously¹²) and illustrate that heritability estimates derived from a homogenous source may be highly inflated. In a relevantly well-powered twin study with a total of ~2100 female twins, the heritability of 25-hydroxyvitamin D was calculated to be 40%, indicating a larger proportion of variance explained by non-genetic factors. Heritability estimates obtained using GWAS SNPs have typically been found to be approximately half of those from classical twin studies⁹,¹⁰, but our estimate of 7.5%, calculated using common genome-wide SNPs, is far lower than reported heritability from twin and family based studies. In addition to potentially inflated estimates from twin studies, the difference may reflect the proportion of heritability explained by rare SNPs or structural variants that were not included in our data, and the potential gene-gene interactions that remain to be identified. The combination of our samples from all seasons is also likely to decrease the probability of finding genetic variants, and hence deflate heritability estimates.

Through partitioning the SNP-heritability of serum 25-hydroxyvitamin D levels, we observed a significant enrichment in immune and hematopoietic tissues; likewise, the cell-type-specific analysis revealed clustering of 25-hydroxyvitamin D and autoimmune diseases, indicating that these traits share a majority
of common cell types. The link between vitamin D deficiency and increased risk for autoimmune inflammatory diseases has long been recognized by epidemiological investigations. Although the underlying mechanisms remain unclear, it is now evident that vitamin D is involved in many biological processes that regulate both innate and adaptive immune responses, through ligand-receptor binding, activation, interaction with response elements in the promoter regions of different genes, and eventually lead to functional changes in a wide variety of immune cells including Th1, Th2, Th17, T regulatory and natural killer T cells. The shared cell type enrichments between vitamin D and autoimmune diseases observed in our study, further suggest that vitamin D not only affects autoimmune diseases through its direct effect (as a ligand), but also through their shared genetic etiology. Thus, individuals with vitamin D deficiency may be more susceptible to these disorders, both because of environmental and genetic influences.

Our genome-wide interaction analysis between genetic variants and dietary intake of vitamin D did not identify new signals. All significant associations observed in the joint test of main genetic and interaction effects were of equal or higher significance level (i.e., lower P-values) in the GWAS of marginal genetic effect performed in the same individuals, indicating no major contribution of interaction effects at these loci. Indeed, only one of the top 5 loci from the overall marginal GWAS showed nominally significant interaction effect, and none passed Bonferroni corrections. While smaller gene-diet interaction effects remain to be discovered, our results provide some evidence against large interactions between common SNPs and dietary vitamin D intake. Still, one cannot completely rule out the possibility of interaction, but only conclude that genetic effects appear stable within vitamin D intake range in the populations studied. Indeed, as for any gene-environment interaction tests, statistical power is highly dependent on the variance of exposure in the samples analyzed, and interactions would remain unobserved if the exposure is homogeneous among individuals. Also, we were not able to capture vitamin D supplementation adequately to include this in the dietary intake variable, and were not able to estimate sunlight exposure as a source of vitamin D production in the skin. Serum 25-hydroxyvitamin D concentrations are mainly determined by modifiable environmental factors, and contrary to estimates from previous twin studies, our large-scale analyses suggest a SNP-heritability rate that is relatively modest in magnitude when considering common variants. Our study also showed that common genetic variants are unlikely to have a strong modifying effect on increases in 25-hydroxyvitamin D following typical dietary intakes, suggesting that consideration of genetic background is not required when determining population based vitamin D intake recommendations. However, our results support the role of vitamin D in immunological diseases as we observed from cell-type-specific analysis for clustering of vitamin D and autoimmune diseases, and the evidence for signal enrichment for immune and hematopoietic tissues. These findings are in line with previous Mendelian Randomization studies which found a putative causal association between vitamin D and autoimmune diseases such as multiple sclerosis and type 1 diabetes. The additional genetic instruments identified by our results could also be used in future Mendelian Randomization analyses of the association between vitamin D and complex traits.

Methods

Study cohorts. We expanded our previous SUNLIGHT consortium GWAS, and undertook a large, multicenter, genome-wide association study of 31 cohorts in Europe, Canada and USA. Our first stage discovery meta-analysis consisted of 79,366 samples of European descent drawn from 31 epidemiological cohorts. Among those 31 cohorts, ten were used as discovery and in-silico replication samples in our previous GWAS publication (the 1958 British Birth Cohort (1958BC), the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Gothenburg Osteoporosis and Obesity Determinants study (GOOD), the health, Aging, and Body Composition study (Health ABC), the Indiana Women cohort, the North Finland Birth Cohort 1966 (NFBC1966), the Old Order Amish study (OAA), the Rotterdam Study (RS), and the Twins UK), and an additional 21 cohorts were included for the current analysis (the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC), the Atherosclerosis Risk in Communities Study (ARIC), the Atherosclerotic Risk of the Prevention Of Osteoporotic Fractures (PROOF), the Epidemiology of Diabetes Interventions and Complications (EDIC), the Case-Control Study for Metabolic Syndrome (GemMetaS), the Helsinki Birth Cohort Study (HBCS), the Helicobacter Pylori Professional Follow-up Study (HPFS, nested a coronary heart disease case-control study), the Invecchiare in Chianti Study (InChianti), the Cooperative Health Research in the region Augsburg (KORA), the Leiden Longevity Study (LLS), the Ludwigshafen Risk and Cardiovascular Health Study (LURIC), the Multi-Ethnic Study of Atherosclerosis (MESA), the Nijmegen Biomedical Study (NBS), the Nurses’ Health Study (NHS, nested a breast cancer case-control study, and a type 2 diabetes case-control study), the Orkney Complex Disease Study (ORCADES), the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO), the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER), the Study of Health in Pomerania (SHIP), the Scottish Colorectal Cancer Study (SOCCS), the Cardiovascular Risk in Young Finns Study (YFS), and more samples from the RS (RSI, RSII, and RSIII)). Full descriptions of all participating cohorts, details of genotyping platforms used, number of SNPs, and the measurements of serum 25-hydroxyvitamin D concentrations in each cohort are shown in Supplementary Table 1 and Supplementary Note 1. Written informed consent was obtained from all participants in the included studies, and the study protocols were reviewed and approved by local institutional review boards.

Power calculation. Our large sample size provided good statistical power for association analysis. At the genome-wide significance threshold of 5×10^{-8}, with a discovery sample size of 75,000, our study had 85% power to detect a genetic variant (single nucleotide polymorphism, SNP) accounting for 0.06% of the total variance in serum 25-hydroxyvitamin D concentrations, and 99% power to detect a variant that explained 0.1% of the total variance. We also had power to detect gene-environment interaction effects even smaller than the observed marginal effects. In the case where a SNP has no marginal effect on circulating 25-Hydroxyvitamin D concentrations (and so could not have been discovered via the marginal GWAS), we had 80% power to detect an interaction that explained 0.07% of the total variance in 25-hydroxyvitamin D concentrations.

Association analysis. Genome-wide analyses were performed within each cohort according to a uniform analysis plan. We fit additive genetic models using linear regression on natural-log transformed 25-hydroxyvitamin D, and adjusted the models for month of sample collection (12 categories), age, sex, and body mass index, and principal components capturing genetic ancestry. Further adjustments included cohort-specific variables, such as geographical location and assay batch, where relevant. For participating studies with a case-control design, we analyzed cases and controls separately. We performed a fixed-effects inverse variance weighted meta-analysis across the contributing cohorts, as implemented in the software METAL, with control for population structure within each cohort and quality control thresholds of minor allele frequency (MAF) >0.05, imputation info score >0.8, Hardy–Weinberg equilibrium (HWE) >1×10^{-6}, and a minimum of two studies and 10,000 individuals contributing to each reported SNP-phenotype association. We regarded P-values <5×10^{-8} as genome-wide significant.

Replication study. We replicated the identified novel loci in two independent data sets for which genotype data were available: the European Prospective Investigation into Cancer and Nutrition (EPIC) study with 40,562 individuals across two nested case-control studies (EPIC-InterAct and EPIC-CVD) and the cohort-wide EPIC-Norfolk study (Supplementary Note 1); and a cohort of 2195 individuals (all controls) additionally collected as part of the SOCCS that were not included in our discovery stage. As for the phenotype, EPIC-InterAct and the EPIC-CVD study was assayed for plasma 25-hydroxyvitamin D, and SOCCS individuals were assayed for total 25-hydroxyvitamin D. We performed the association analysis in a similar manner, adjusted for age, sex, time of sample collection, and study center where relevant. We regarded P-value <0.05 in the replication samples, and P-value <5×10^{-4} in the pooled analysis as successful replication.

Conditional analysis. After identifying the primary associated variant at each locus selected according to the strength of its association, we further tested whether there were any other SNPs significantly associated with 25-hydroxyvitamin D after accounting for the effect of lead SNP. We thus performed a stepwise selection procedure for those chromosomes where a significant variant was previously identified. We started with the most significantly associated SNP, scanning through the remaining additional reported SNPs using a stepwise procedure, one at a time, based on their conditional P-values. Finally, we fit all selected SNP into one model to estimate their joint effects.
We used GCTA-COJO software to accommodate our summary level GWAS data27, and the Cancer Genetic Marker of Susceptibility (CGEMS) GWAS with 2287 individuals of European descent and 2,543,887 genotyped and imputed (HapMap22) SNPs as reference panel.

SNP-by-diet interaction. We performed a genome-wide association screening of circulating 25-hydroxyvitamin D while accounting for potential interaction effect between SNP and dietary vitamin D intake. The tests incorporating diet interaction were based on the following model:

\[ \ln(25(OH)D) = \beta_0 + \beta_1 x + \beta_2 x E + \beta_3 x G + \beta_4 x E + \beta_5 x Z \]

where \( G \) is a SNP that was coded additively, \( x \) is the raw vitamin D intake, measured on a continuous scale. The parameters \( \beta_1, \beta_2, \beta_3 \), and \( \beta_4 \) are the intercept, main effect of SNP, the main effect of dietary vitamin D intake, and the interaction effect between \( G \) and \( E \). The model also included the same covariates \( Z \) as for the marginal effect screening: the effects of which were captured in the parameter \( \beta_5 \). We considered both a standard 1 degree-of-freedom test of interaction effect (i.e., null hypothesis of \( \beta_3 = 0 \)), and a joint 2 degree-of-freedom test of main genetic effect plus gene-by-diet interaction (i.e., null hypothesis of \( \beta_1 = 0 \) and \( \beta_2 = 0 \)). For comparison purposes, we also considered a model adjusting for vitamin D intake but not modeling interaction (i.e., not including the \( \beta_4 \times G \times E \) term) using the same subset of individuals.

Vitamin D intake was available for 15 cohorts on a total of 41,981 individuals. It included both the population based cohorts (ARIC, 1958BC, B-PROOF, FHS, Health ABC, MESA, NFBC, RS, RS III, and YFS as part of the overall Meta-GWAS, plus two additional cohorts, the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS), and the Uppsala Longitudinal Study of Adult Men (ULSAM), genotyped on custom array that were not included in the overall meta-GWAS but were included in this SNP-by-diet interaction analysis), and case-control studies (HPFS (HPFS_CHD), NHS (NHS_BRCA, NHS_T2D), and SOCCS). For the latter studies, all analyses were performed separately in cases and controls. The aforementioned interaction model was applied to each of the included cohorts, and study-specific results were meta-analyzed using inverse-variance weighted sum of effect estimates as implemented in METAL.26 For the 2 degree-of-freedom test we used joint framework described in two previous published papers28,29. Quality control filtering was performed on each study before meta-analysis. Only SNPs with imputation info score > 0.8, MAF > 0.05, HWE > 1×10⁻⁸, and a minimum of total sample size in the meta-analysis > 10,000 were retained.

Linkage Disequilibrium score regression. We performed linkage disequilibrium score regression (LDsc) analysis to estimate the SNP-heritability of serum 25-hydroxyvitamin D concentrations21.31. This method is based on a validated relationship between LD score and \( \chi^2 \)-statistic:

\[ E \left( \chi^2 \right) = N \frac{\hat{h}^2}{M} \]

where \( \hat{E} \left( \chi^2 \right) \) denotes the expected \( \chi^2 \)-statistic for the association between outcome and SNP \( j \), \( N \) is the study sample size available for SNP \( j \), \( M \) is the total number of SNPs and \( \hat{h}^2 \) denotes the LD score of SNP \( j \) defined as \( \hat{h}^2 = \sum_i \lambda_{ij}^2 \). LDsc was performed using only summary genome-wide imputed genotypes, and is computationally cost effective at large sample sizes. We used the summary statistics from 25-hydroxyvitamin D meta-GWAS results, with SNPs available in at least 2 studies and a sample size of at least 10,000. We first analyzed the SNP-heritability by using (1) SNPs across the entire genome that passed quality control; (2) SNPs excluding the top associations (SNPs reaching genome-wide significance, \( P < 5×10^{-8} \)), as well as all SNPs within \( ≤ 500 \) kb of the top hits in the region; and (3) SNPs excluding the nominally significant associations (\( P ≤ 0.05 \)). We subsequently partitioned the heritability through three different models: (1) a full baseline model including the 24 publicly available main annotations that are not specific to any cell type, the 500-bp windows around each annotation, as well as 100-bp windows around chromatin immunoprecipitation and sequencing peaks (CHIP-seq) when appropriate. This resulted in a total of 52 overlapping functional categories in the full baseline model: (2) a cell-type-specific model including 10 cell type groups: adrenal and pancreas, central nervous system (CNS), cardiovascular, connective and bone, gastrointestinal, immune and hematopoietic, kidney, liver, skeletal muscle, and other; and (3) a cell-type-specific model including 220 cell-type-specific annotations for the four histone marks with putative enhancer or promoter functions, H3K4me1, H3K4me3, H3K9ac, and H3K27ac.

Details of the 24 publicly available annotations, the 220 cell-type-specific annotations, as well as the 10 cell type groups were described by Finucane et al.31. Briefly, the 24 annotations included coding, UTR (3’UTR and 5’UTR), promoter and intronic regions, acquired from the UCSC Genome Browser22 and post-processed by Gusse et al.22, the three histone marks (mono-methylation (H3K4me1) of histone H3 at lysine 4, tri-methylation (H3K4me3) of histone H3 at lysine 4 and acetylation of histone H3 at lysine 9 (H3K9ac) processed by Trynka et al.34,36 and two versions of acetylation of histone H3 at lysine 27 (H3K27ac, one version processed by Hnisz et al.35, another version used by the Psychiatric Genomics Consortium (PGC)38, open chromatin, as reflected by DNase I hypersensitivity sites (DHs and fetal DHs)33, obtained as a combination of Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics data, and processed by Trynka et al.36, combined chromHMM and Segway predictions obtained from Hoffman et al.37, which leverage on many annotations to partition the genome into seven underlying chromatin states (the CCCTC-binding factor (CTCF), promoter-flanking, transcribed region, transcription start site (TSS), strong enhancer, weak enhancer, and the repressed region); regions that are conserved in mammals, provided by Lindblad-Toh et al.38 and post-processed by Ward and Kellis39; super-enhancers, which are large groups of putative enhancers with high levels of activity, provided by Hnisz et al.39; FANTOM5 enhancers mapped by using cap analysis of gene expression in the FANTOM5 panel of samples, obtained from Andersson et al.32; digital genomic footprint (DFG) and transcription factor binding site (TFBS) annotations downloaded from ENCODE35 and post-processed by Gusse et al.35. We included 300-bp windows around each of the 24 main annotations in the baseline model, and 100-bp windows around ChIP-seq when appropriate, to prevent upward bias of estimates generated by enrichment in the nearby regions.

In addition to the baseline model using 24 main annotations, we also performed cell-type-specific analyses using annotations of the four histone marks (H3K4me1, H3K4me3, H3K9ac and H3K27ac). Each cell-type-specific annotation corresponds to a histone mark in a single cell type (for example, H3K27ac in adipose nucleic tissues), and there was a total of 220 such annotations. We further subdivided these 220 cell-type-specific annotations into 10 categories by aggregating the cell-type-specific annotations within each group (for example, SNPs related with any of the four histone modifications in any hematopoietic and immune cells were considered as one big category). When generating the cell-type-specific models, we added each annotation individually (one at a time) to the baseline model, creating separate models to control for overlap with the genomic functional elements in the full baseline model but not overlap with the other cell types.

We additionally assembled the summary statistics from GWAS of 37 traits or diseases performed in individuals of European descent, which are publicly available34,45 or applied from the UK Biobank. These studies span a wide range of phenotypes, from anthropometric indices such as height, weight, BMI, to mental disorders (for example depressive syndrome and schizophrenia) to autoimmune and inflammatory diseases (for example rheumatoid arthritis and celiac diseases). We considered the pairwise genetic correlation (i.e., genetic correlation from cross-trait heritability) between 25-hydroxyvitamin D and each of the 37 traits. We further conducted the same cell-type-specific analysis for each trait, and plotted beta-coefficient z-score matrix, constructed from the total 220 annotations by 37 traits, into four heat-maps based on their four histone marks.

Finally, in addition to the genetic correlation analysis which reflects shared genetic factors across different traits but does not inform direction, we also attempted to identify directions of such correlation using an algorithm proposed by Pickrell et al.36. The method adopts a similar intuition as the Mendelian Randomization approach, where, if a trait \( X \) influences trait \( Y \), then SNPs associating with \( X \) should also influence \( Y \), and the SNP-specific effect sizes for the two traits should be correlated. Further, since \( Y \) does not influence \( X \), but could be influenced by mechanisms independent of \( X \), genetic variants that influence \( Y \) do not necessarily influence \( X \). Based on this intuition, the method proposes two "causal" models and two "non-causal" models, and calculates the relative likelihood ratio of the best non-causal model compared to the best causal model. We determined significant SNPs for each given trait by selected genome-wide significant (\( P < 5×10^{-8} \)) SNPs and pruned the numbers based on their LD-pattern in the European populations in Phase1 of 1000 Genome Project. We scanned through all pairs of 25-hydroxyvitamin D and traits to identify directional correlations. We consider pairs of traits with likelihood ratio non-causal vs causal < 0.05 as having evidence of directional correlations.

Data availability. The GWAS summary statistics on serum circulating vitamin D concentrations is available at dbGap https://drstring.google.com/drive/folders/ 0BzYDtCo_doHJRFRKR0ltZHZWZjQ; all relevant data are available from the authors upon request.

Received: 21 July 2017 Accepted: 15 December 2017
Published online: 17 January 2018

References

1. Mokry, L. E. et al. Vitamin D and risk of multiple sclerosis: a mendelian randomization study. PLoS Med. 12, e1001866 (2015).
2. Rhead, B. et al. Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. Nat Genet. 2, e1001866 (2015).
3. Martinez, A. R. et al. Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data. Br. Med. J. 356, 16583 (2017).
4. Pilz, S., Verheyen, N., Gruberl, M. R., Tomaschitz, A. & Mahr, W. Vitamin D and cardiovascular disease prevention. Nat. Rev. Cardiol. 13, 404–417 (2016).

5. Garland, C. F. et al. Role of vitamin D in cancer prevention. Am. J. Public Health 96, 252–261 (2006).

6. Fernandes de Abreu, D. A., Eyles, D. & Feron, F. Vitamin D, a neuro-immunomodulator: implications for neurodegenerative and autoimmune diseases. Psychoneuroendocrinology 34 Suppl 1, S265–277 (2009).

7. Lagunova, Z., Porojnicu, A. C., Lindberg, F., Hesseberg, S. & Moan, J. The dependency of vitamin D status on body mass index, gender, age and season. Anticancer Res. 29, 3713–3720 (2009).

8. Wang, T. J. et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet Lond. Engl. 376, 180–188 (2010).

9. Karolh, C. et al. Heritability and seasonal variability of vitamin D concentrations in male twins. Am. J. Clin. Nutr. 92, 1393–1398 (2010).

10. Orton, S.-M. et al. Evidence for genetic regulation of vitamin D status in twins with multiple sclerosis. Am. J. Clin. Nutr. 88, 441–447 (2008).

11. Ahn, J. et al. Genome-wide association study of circulating vitamin D levels. Hum. Mol. Genet. 19, 2739–2745 (2010).

12. Hiraki, L. T. et al. Exploring the genetic architecture of circulating 25-hydroxyvitamin D. Genet. Epidemiol. 37, 92–98 (2013).

13. Boyadjiev, S. et al. Cranio-lenticulo-sutural dysplasia associated with defects in collagen secretion. Clin. Genet. 80, 169–176 (2011).

14. Boyadjiev, S. A. et al. Cranio-lenticulo-sutural dysplasia is caused by a SEC23A mutation leading to abnormal endoplasmic reticulum-to-Golgi trafficking. Nat. Genet. 38, 1192–1197 (2006).

15. Myung, J. K. et al. Well-differentiated liposarcoma of the oesophagus: clinicopathological, immunohistochemical and array CGH analysis. Pathol. Oncol. Res. 11, 415–420 (2011).

16. Manousaki, D. et al. Low-frequency synonymous coding variation in CYP2R1 has large effects on vitamin D levels and risk of multiple sclerosis. Am. J. Hum. Genet. 101, 227–237 (2017).

17. Arguelles, L. M. et al. Heritability and environmental factors affecting vitamin D status in rural Chinese adolescent twins. J. Clin. Endocrinol. Metab. 94, 3273–3281 (2009).

18. Mills, N. T. et al. Heritability of transforming growth factor-β1 and tumor necrosis factor-receptor type 1 expression and vitamin D levels in healthy adolescent twins. Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud. 18, 28–35 (2015).

19. Livshits, G., Karasik, D. & Seibel, M. J. Statistical genetic analysis of plasma levels of vitamin D: familial study. Am. J. Hum. Genet. 63, 429–439 (1998).

20. Yu, H.-J., Kwon, M.-J., Woo, H.-Y. & Park, H. Analysis of 25-hydroxyvitamin D status according to age, gender, and seasonal variation. J. Clin. Lab. Anal. 30, 905–911 (2016).

21. Hunter, D. et al. Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res. 16, 371–378 (2001).

22. Agmon-Levin, N., Theodor, E., Segal, R. M. & Shoenfeld, Y. Vitamin D in systemic and organ-specific autoimmune diseases. Clin. Rev. Allergy Immunol. 45, 256–266 (2013).

23. Yang, C.-Y., Leung, P. S. C., Adamopoulos, I. E. & Gershwin, M. E. The implication of vitamin D and autoimmunity: a comprehensive review. Clin. Rev. Allergy Immunol. 45, 217–226 (2013).

24. Aschard, H. A perspective on interaction effects in genetic association studies. Genet. Epidemiol. 40, 678–688 (2016).

25. Cooper, J. D. et al. Inherited variation in vitamin D genes is associated with predisposition to autoimmune disease type 1 diabetes. Diabetes 60, 1624–1631 (2011).

26. Wilker, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinforma. Oxf. Engl. 26, 2190–2191 (2010).

27. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 44, 337–341 (2012).

28. Aschard, H., Hancock, D. B., London, S. J. & Kraft, P. Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. Hum. Hered. 70, 292–300 (2010).

29. Manning, A. K. et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP×environment regression coefficients. Genet. Epidemiol. 35, 11–18 (2011).

30. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genomewide association studies. Nat. Genet. 47, 291–295 (2015).

31. Finucane, H. K. et al. Partitioning heritability by functional annotation using genomewide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).

32. Bulik-Sullivan, B. K. et al. The human genome browser at UCSC. Genome Res. 12, 996–1002 (2002).

33. Gusev, A. et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Human. Genet. 95, 535–552 (2014).
Xia Jiang1,2, Paul F. O’Reilly3, Hugues Aschard1,4, Yi-Hsiang Hsu5,6,7, J. Brent Richards8, Josée Dupuis9,10, Erik Ingelsson11,12, David Karasik5, Stefan Pilz13, Diane Berry14, Bryan Kestenbaum15, Jusheng Zheng16, Jianan Luan16, Eleni Sofianopoulou17, Elizabeth A. Streeten18, Demetrius Albanes19, Pamela L. Lutsey20, Lu Yao20, Weihong Tang20, Michael J. Econs21, Henri Wallaschofski22,23, Henry Völzke23,24, Ang Zhou25, Chris Power14, Mark I. McCarthy26,27,28, Erin D. Michos29,30, Eric Boerwinkle31, Stephanie J. Weinstein19, Neil D. Friedman19, Wen-Yi Huang22, Natasja M. Van Schoor33, Nathalie van der Velde34,35, Lisette C.P.G.M.de Groot36, Anke Enneman34, L. Adrienne Cupples9,10, Sarah L. Booth37, Ramachandran S. Vasan10, Ching-Ti Liu9, Yanhua Zhou9, Samuli Ripatti38, Claes Ohlsson39, Liesbeth Vandenput39, Mattias Lorentzon40, Johan G. Eriksson41,42, M. Kyla Shea37, Denise K. Houston43, Stephen B. Kritchovsky43, Yongmei Liu44, Kurt K. Lohman45, Luigi Ferrucci46, Munro Peacock47, Christian Gieger47, Marian Beekman48, Eline Slagboom48, Joris Deelen48,49, Diana van Heemst50, Marcus E. Kleber51, Winfried März51,52,53, Ian H. de Boer54, Alexcis C. Wood55, Jerome I. Rotter56, Stephen S. Rich57,58, Caissonian Robinson-Cohen59, Martin den Heijer60, Marjo-Riitta Jarvelin61,62,63,64, Alana Cavadino14,65, Peter K. Joshi16, James F. Wilson66,67, Caroline Hayward67, Lars Lind12, Karl Michaëllson68, Stella Trompet60,69, M. Carola Zillikens60, Andre G. Uitterlinden34,60, Fernando Rivadeneira34,60, Linda Broer60, Lina Zgaga70, Harry Campbell66,71, Evropi Thodoratos66,71, Susan M. Farrington71, Maria Timofeeva72, Malcolm G. Dunlop71, Ana M. Valdes73,74, Emmi Tikkanen74, Terho Lehtimäki75,76, Leo-Pekka Lyytikäinen75,76, Mika Kähönen77,78, Olli T. Raitakari79,80, Vera Mikkilä81, M. Arfan Ikram34, Navede Sattar82, J. Wouter Jukema69,83, Nicholas J. Wareham16, Claudia Langenberg16, Nita G. Forouhi16, Thomas E. Gundersen84, Kay-Tee Khaw17, Adam S. Butterworth17, John Danesh17, Timothy Spector72, Thomas J. Wang18, Elina Hyppönen14,25, Peter Kraft1 & Douglas P. Kiel5,6,7

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-02662-2.
Division of Cardiovascular Medicine, Vanderbilt Heart and Vascular Institute, 2220 Pierce Avenue 383 Preston Research Building, Nashville, TN 37232-6300, USA. Xia Jiang, Paul F. O’Reilly, Hugues Aschard and Yi-Hsiang Hsu contributed equally to this work. Thomas J. Wang, Elina Hyppönen, Peter Kraft and Douglas P. Kiel jointly supervised this work.