Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration

Vitamin D deficiency is a candidate risk factor for a range of adverse health outcomes. In a genome-wide association study of 25 hydroxyvitamin D (25OHD) concentration in 417,580 Europeans we identify 143 independent loci in 112 1-Mb regions, providing insights into the physiology of vitamin D and implicating genes involved in lipid and lipoprotein metabolism, dermal tissue properties, and the sulphonation and glucuronidation of 25OHD. Mendelian randomization models find no robust evidence that 25OHD concentration has causal effects on candidate phenotypes (e.g. BMI, psychiatric disorders), but many phenotypes have (direct or indirect) causal effects on 25OHD concentration, clarifying the epidemiological relationship between 25OHD status and the health outcomes examined in this study.
In recent decades, there has been considerable interest in the links between vitamin D concentration and general health. While classically linked to bone disorders, there is growing evidence to suggest that suboptimal vitamin D status may be a risk factor for a much wider range of adverse health outcomes. Vitamin D, the sunshine hormone, is the precursor of a secosteroid transcription regulator that operates via a nuclear receptor, and like other steroid hormones, exerts transcriptional control over many regions of the genome across many different tissues. In environments with access to adequate sunshine, ultraviolet radiation on the skin converts a precursor of cholesterol to vitamin D₃. This is then further converted to 25-hydroxyvitamin D₃ (25OHD; used in assays of general vitamin D status), and then to the active hormone 1,25-dihydroxyvitamin D₃ (1,25OHD) in a variety of tissues. Some foods and vitamin D supplements also contribute to vitamin D levels. Definitions of vitamin D deficiency (e.g., <25 nmol L⁻¹ of 25OHD) are predominantly based on bone health—according to these definitions, vitamin D deficiency is common in many countries, regardless of latitude and economic status.

Environmental factors such as season of testing and latitude contribute substantially to the serum concentration of 25OHD (lower in winter/spring; lower at higher latitudes). A recent multivariate twin study demonstrated that approximately half of the total additive genetic variation in 25OHD may reflect genetic variation in skin colour and sun exposure behaviour. Genome-wide association studies (GWAS) have identified common single-nucleotide polymorphisms (SNPs) located in biologically plausible genes. The largest GWAS to date (N = 79,366) reported six significant loci, which include GC (the vitamin D-binding protein gene), the DHCR7/NADSYN1 region (DHCR7 is involved in a conversion of a 25OHD precursor molecule to cholesterol) and CYP2R1 and CYP24A1 genes (which encode enzymes involved in 25OHD metabolism). In total, common SNPs explain 7.5% (standard error (s.e.) 1.9%) of the variance of 25OHD.

Here, we conduct a GWAS of 25OHD based on the large UK Biobank (UKB) sample and conduct a suite of post-GWAS analyses to aid interpretation of the results. We present models that explore the genetic or causal relationship between body mass index (BMI) and 25OHD (high BMI is associated with lower 25OHD concentration in observational studies). Because we have an interest in the association between 25OHD and psychiatric disorders, we use Mendelian randomisation methods to investigate the bidirectional association between 25OHD and psychiatric disorders, as well as with a wider range of traits and diseases. In addition, we present a GWAS to identify loci associated with variance in 25OHD (i.e., variance quantitative trait locus (vQTL) analysis) which can identify putative genotype-environment interactions without prior identification of the environmental effect. We identify 143 independent loci in 112 1-Mb regions associated with 25OHD concentration, and our findings implicate genes involved in lipid metabolism, dermal tissue properties and conjugation of 25OHD. We find no robust evidence that 25OHD concentration has causal effects on candidate phenotypes. However, we show that many phenotypes have (direct or indirect) causal effects on 25OHD concentration.

Heritability and SNP-based heritability. Our UKB sample included a set of 58,738 individuals related with coefficient of relationship (r) > 0.2 to at least one other person in the set (all relatives), from whom we estimate the heritability of 25OHD to be 0.32 (s.e. = 0.01) with little evidence for inflation from shared family environment (Fig. 2). Our GWAS also estimated variance quantitative trait locus (vQTL) analysis which can identify putative genotype (or 9000 SNPs of the ~1.1 million HapMap3 panel) common SNPs (Supplementary Table 2), much lower than estimates for most complex traits. The SBayesS S parameter, which describes the effect size–MAF relationship, was estimated as −0.78 (s.e. = 0.04; Supplementary Table 3), consistent with a model of negative selection on the genetic variants associated with 25OHD levels (the magnitude of S is higher than those of most complex traits studied). Estimation of $\hat{h}^2_{SNP}$ partitioned into ten components based on five MAF bins (each median split by linkage disequilibrium score) did not provide strong evidence for an increased role for less common variants, given the s.e. of estimates (Supplementary Fig. 3). Despite a strong phenotypic association between 25OHD and BMI of −0.76 nmol L⁻¹ per BMI unit (−0.036 RINT(25OHD) standard deviation (SD) units per BMI unit, linear regression $P = 2.2 \times 10^{-16}$) and a phenotypic correlation of −0.17 (Supplementary Table 1), the estimates of heritability (both family and SNP-based) were hardly impacted when BMI was included as a covariate (Supplementary Fig. 2).

Genome-wide association study (GWAS) analysis. Given the potential for collider bias from using a heritable trait as a covariate, we conducted GWAS for 25OHD with and without BMI as a covariate. We also used mtcOJO to estimate the 25OHD SNP effects conditioning on those estimated for BMI from UKB data, a summary-data-based conditional analysis approach that was shown in simulations to be robust to collider bias when conditioning on a correlated trait. Results were comparable across the three levels of BMI adjustment (Supplementary Data 2), so we report those with no correction for BMI, using results from all three analyses when this aids interpretation of the

**Results**

25OHD phenotype. In total, 417,580 European UKB participants had both measures of vitamin D 25OHD and genome-wide genotypes ("Methods"). The distribution of 25OHD concentration, in keeping with expectation, is right skewed (Supplementary Fig. 1a), and showed the expected seasonal fluctuation (Supplementary Fig. 1b–d), with median, mean and interquartile range of 47.9, 49.6, 33.5–63.2 nmol L⁻¹ (Supplementary Table 1). Covariates of age, BMI, genotyping batch, assessment centre, month of testing, supplement intake and the first four ancestry principal components (PCs), but not sex, were all significantly associated with 25OHD (Supplementary Table 1). Month of testing accounts for 14% of the variance of 25OHD. Subsequent analyses use 25OHD after rank-based inverse-normal transformation (RINT), unless otherwise stated.
Analyses using individual-level data

**Phenotype descriptive analysis**
- Extract vitamin D from UK Biobank (data field 30890)
- Assess phenotype distribution and seasonal fluctuations
- Apply rank-based inverse-normal transformation
- Assess covariate effects
- Assess relationship with BMI

**Genotypes**
- Genotype data QC
- GWAS (with and without BMI adjustment)
- Season-stratified GWAS
- Conditional and joint (COJO) analysis
- Meta-analysis with SUNLIGHT consortium GWAS results
- vQTL GWAS
- GxE analysis with season
- Heritability – 58K UKB have family member ≥ 2nd degree

Analyses using GWAS summary statistics

**SNP-based Heritability**
- SBayesR
- SBayesS
- LDSC
  - Partitioned by functional annotation (enhancer, promoter, etc)
  - Partitioned by cell-type annotation

**Annotation**
- FUMA
- SMR
  - eQTLs from blood, skin, liver & brain

**Cross-trait analysis**
- Genetic correlations
- Mendelian Randomisation analyses
  - GSMR
  - 2-sample MR for significant MR results

---

**Fig. 1 Outline of key analytic steps described in this study.** BMI body mass index, eQTL expression quantitative trait locus, FUMA functional mapping and annotation\(^{23}\), GREML\(^{69}\) genomic relationship restricted maximum likelihood\(^{68}\), GSMR generalised summary-based MR, GWAS genome-wide association study, GxE genotype-by-environment interaction, LDSC linkage disequilibrium score regression, MR Mendelian Randomisation, SMR summary-based MR\(^{26}\), QC quality control, UKB UK Biobank, vQTL variance quantitative trait locus\(^{12}\).

---

**Fig. 2 Heritability, SNP-based heritability and variance explained in out-of-sample prediction.** Heritability (left panel) and SNP-based heritability (middle panel) estimates and the variance explained in out-of-sample prediction (right panel). Heritability and SNP-based heritability estimates are presented with 95% confidence interval. GCTA-GREML was used to estimate heritability from a UKB subset that included all pairs of individuals related with coefficient of relationship > 0.2 \((N = 58,738\) relatives). GCTA-GREML was used to estimate SNP-based heritabilities labelled GREML summer or winter using samples of ~50 K participants randomly drawn from the UKB. The SBayesR SNP-based heritability is estimated from the GWAS summary statistics \((N = 417,580)\). In out-of-sample prediction into the QIMR and the UKB replication (UKBR) samples, polygenic risk scores (PRS) calculated by the standard \(P\)-value threshold method \((P + T)\) were outperformed by using SNP effect estimates calculated from GWAS summary statistics using the SBayesR or SBayesS methods. Bars of the same colour used the same methodology (noting that SBayesR generates an estimate of SNP-based heritability as well as SNP effect sizes in prediction analysis). The numbers on top of the bars are \(-\log_{10} P\)-value of the regression of 25OHD on 25OHD PRS. COJO conditional and joint, \(rg\) genetic correlation, s.e. standard error.

---

**Table:**

| Heritability | SNP-based heritability | Out-of-sample prediction |
|-------------|------------------------|--------------------------|
| Polygenic risk prediction using different SNP sets and SNP effect size estimates. |
| Dark coloured bars in each pair are for the UKB replication sample \((N = 1632)\) and light coloured bars are for the QIMR Australian replication sample \((N = 1632)\). |

---

**Note:**
- GREML, Genomic relationship restricted maximum likelihood.
- SBayesR, SBayesS, Genetic correlation.
- LDSC, Linkage disequilibrium score regression.
- FUMA, Functional mapping and annotation.
- SMR, Summed Manhattan plot regression.
results. While there is some debate about the clinical threshold for vitamin D deficiency (25 nmol L\(^{-1}\)) or 30 nmol L\(^{-1}\) serum concentration\(^1\)\(^2\), we chose the more conservative threshold of 25 nmol L\(^{-1}\). We conducted analyses bisecting 25OHD into a binary trait (less than 25 nmol L\(^{-1}\), 25 nmol L\(^{-1}\) or greater), but the results were consistent (given the expected reduced power) with our reported results treating 25OHD as a quantitative trait (Supplementary Note 2).

A total of 8,806,780 SNPs with MAF > 0.01 were tested in the GWAS analysis. Of these, 18,864 were genome-wide significant (GWS; \(P < 5 \times 10^{-8}\)). To identify independently associated loci, we applied the GCTA–COJO method\(^3\) to the GWAS summary statistics using LD between SNPs estimated from a UKB subset ("Methods"), and identified 143 independent loci (including one on chromosome X) (Fig. 3; Supplementary Data 3) in 112 1-Mb regions. Of these, 15 loci were low-frequency variants (MAF < 0.05), and 106 regions had no previously identified associations. All six loci reported in previous vitamin D GWAS\(^8\)\(^9\)\(^10\) were replicated in our study. While recognising that the COJO method cannot distinguish between SNPs in perfect LD, we note that within the 143 COJO independent variants: (a) 14 were non-synonymous variants that alter protein coding (\(rs = 0.92, \text{ s.e.} = 0.06\)). Meta-analysis with our UKB GWAS results after imputation\(^21\) of the SUNLIGHT summary statistics (Supplementary Methods) (6,912,294 overlapping SNPs) identified 15,154 GWS variants, 150 GCTA–COJO independent SNPs (Supplementary Methods, Supplementary Data 2). Of these, 91 were within 1-Mb regions also identified in the UKB alone as GWS. Given that the meta-analysis only increased the total number of significant loci by seven, and given our preference not to include BMI as a covariate, we continued with the UKB-only results for our downstream analyses. See Supplementary Note 3 for further details.

Functional mapping and annotation of GWAS. To annotate the 25OHD GWAS, we first used the FUMA online pipeline\(^23\). Gene-set analyses showed that the top four pathways were related to glucuronidation, ascorbate and aldurate metabolism and uronic acid metabolism (Supplementary Data 4, 5). Keratinisation was the top Gene Ontology (GO) biological processes identified. Based on 53 tissue types from GTEx v6\(^24\), the top tissues for differentially expressed genes identified in the GWAS were liver, brain and skin (sun exposed, and non-sun exposed; Supplementary Data 6). Partitioned SNP-based heritability analysis\(^25\) using cell-type-specific annotations identified five cell types (hepatocytes, two types of liver cells, skin cells and blood cells) at the nominal significance level of 0.05 (Supplementary Data 7), but none remained significant after correction for multiple testing (stratified LD score regression \(P (P_{\text{LDSC}}) < 2.4 \times 10^{-4}\)). In partitioned SNP-based heritability analysis using SNP annotation to 53 functional categories\(^26\), 11 passed multiple testing significance threshold \(P (P_{\text{LDSC}}) < 9.4 \times 10^{-4}\;\text{Supplementary Data 8}\) with a mix of annotations including
transcription factor binding sites and transcription start sites (notable because vitamin D operates via a nuclear receptor, which binds to vitamin D response elements), as well as a role for repressed sites, conserved regions, enhancer and coding regions and histone modification marks.

To identify 25OHD SNP associations with statistical evidence consistent with a causal/pleiotropic association via gene expression, we used summary-data-based Mendelian randomisation (SMR)26 using the 15,504 gene probes with significant cis-eQTLs identified from whole blood eQTLGen data27. After Bonferroni correction, we found 112 significantly associated gene expression probes (PSMR < 3.2 × 10−6, i.e., 0.05/15,504, being the total number of probes tested in SMR analysis; Supplementary Data 9, Supplementary Fig. 4; full details of the SMR analyses can be found on https://cnsgenomics.com). These results are discussed in detail in Supplementary Note 4, and add weight to the hypothesis that the SMR-identified eQTL variants may be causally related to 25OHD concentrations.

Putative causal relationships with other traits. First, we investigated the relationship between 25OHD and BMI. The LDSC28 genetic correlation estimated from 25OHD and BMI GWAS summary statistics was −0.17 (s.e. = 0.03) (Supplementary Fig. 2, Supplementary Data 10). Bidirectional Mendelian randomisation16 analysis provided strong support for the hypothesis that high BMI is causal for low 25OHD (PBMI.25OHD = 0.005; PGSMR = 4.7 × 10−162; based on 1020 BMI-associated SNP instruments), with no support for a causal effect of vitamin D on BMI (P25OHD.BMI = 0.008; s.e. = 0.006; PPGSMR = 0.20; based on 210 vitamin D-associated SNPs) (these results were confirmed by other MR methods29; Supplementary Table 5). Notably, the HEIDI-outlier test in the GSMR analyses excluded 70 BMI and 67 25OHD SNP instruments, whose combination of SNP effect sizes likely reflects a pleiotropic relationship or confounding. Using the SNPs excluded by the HEIDI-outlier test, the estimated effects were bBMI.25OHD = 0.17 (s.e. = 0.0182; PGSMR = 1.2 × 10−20) and b25OHD.BMI = −0.15 (s.e. = 0.017, PGSMR = 2.7 × 10−19). Hence,
disorders are associated with behaviours that lead to reduced
BMI and low 25OHD, the biological relationship between these
traits is more complex.

Next, we estimated genetic correlations (r_g) between 25OHD
and 16 traits with GWAS summary statistics available in LD
Hub28, and we used LDSC to estimate r_g between 25OHD and 18
traits (including six psychiatric disorders) with GWAS summary
statistics that are more recent than those included in LD Hub.
Although many of the traits are highly correlated, we use a
Bonferroni correction for 764 tests as the threshold for discussion of
r_g. The LDSC regression intercepts were close to zero, suggesting no
sample overlap (Supplementary Data 10), except for LD Hub traits
derived from UKB analyses, where the intercept, as expected from theory, equates to the
phenotypic correlation. We found significant associations between
25OHD and a range of brain-related phenotypes (including
autism spectrum disorder, intelligence32, major depression, bipolar
disorder and schizophrenia; Supplementary Fig. 5).

Notably, the most significant r_g were with cognitive-associated
traits—for example, a negative correlation (r_g = −0.24, s.e. =
0.03, P_{H0,rg=0} = 1.6 × 10^{−14}) with intelligence. There was also a significant negative r_g with hours spent using a computer
(r_g = −0.22, s.e. = 0.03, P_{H0,rg=0} = 5.1 × 10^{−15}). These findings
may be mediated by an association between higher intelligence and
behaviour associated with less exposure to bright sunshine
(and thus, lower 25OHD). Of note, behaviours associated with
outdoor activity (duration of walks, duration of vigorous activity)
were positively associated with 25OHD, while phenotypes related
to chronic disability were negatively associated with 25OHD.

Next, we investigated if some of the significant genetic
correlations could be explained by causal relationships using
bidirectional GMSR models—here a more complex pattern of association emerged (Fig. 5; Supplementary Data 11). We found
no evidence for putative causal effects between 25OHD and other
traits; GMSR analyses without the HEIDI-outlier filtering step
(Fig. 5a) suggests strong pleiotropy for some traits, such as
dyslipidemia, coronary artery disease, intelligence and educa-
tional attainment. Finally, we examined the reciprocal relation-
ship—if variants associated with a range of traits were
directionally associated with 25OHD. Regardless of the use of
HEIDI filtering, and often regardless of adjustments for BMI, we
found evidence consistent with increased risk of several traits or
disorders being causal (directly or indirectly) with lower 25OHD
concentrations (Fig. 5b). This was the case for intelligence,
dyslipidemia, major depression, bipolar disorder, type 2 diabetes
and schizophrenia. The findings might suggest these traits or
disorders are associated with behaviours that lead to reduced
production of 25OHD (e.g., less outdoor activity and physical
activities). The GMSR findings were also checked with the
portfolio of MR methods implemented in the two-sample MR
(2SMR) software29 (Supplementary Data 12).

**Gene-environment interplay.** We conducted a genome-wide
vQTL analysis, as implemented in OSCA12 to identify SNPs
associated with variance in 25OHD (not RINT-transformed).
Such associations can reflect genotype-by-environment interac-
tion in the absence of measurement, or indeed knowledge, of the
interacting environmental risk factor. Using data from 318,851
unrelated individuals of European ancestry, we tested 6,098,063
variants with MAF > 0.05, and identified 4008 GWS vQTLs, of
which 25 were independent (LD r^2 < 0.01, 5-MB window), and
several were in genes with previously described links to vitamin
D-related pathways (e.g., GC, UGT2B7, SEC23A, SULT2A1,
KLK10, NADSYN1). Of the 25 independent vQTLs, 23 were also
QTLs (identified as genome-wide significant in the GWAS
analysis) while the two non-QTL loci were still associated at
P_{GWAS} < 10^{−3} (Supplementary Data 13). One was in the POR
gene, which encodes a cytochrome P450 oxidoreductase that
donates electrons from NADPH to cytochrome P450 enzymes
(encoded by CYP450 genes), which are involved in vitamin D
metabolism13,14. Variants in POR have previously been associated
with coffee intake33. The other exclusive vQTL (rs1030431) is
12,126 bp upstream of UBXN2B; the SNP is significantly
associated with gall bladder diseases and lipid metabolism traits
in the UKB24.

An environmental factor with known association with 25OHD is
the season of testing. To investigate whether the associations between
the vQTLs and the phenotypic variance of 25OHD reflected
gene–environment (GxE) interactions with season of blood draw,
we performed a GxE analysis with season (winter vs. summer). Of
6,098,063 variants tested (MAF > 0.05), 1127 had a GWS (P < 5 ×
10^{−8}) interaction with season, and 1120 (99%) were also GWS in
the vQTL analysis. From the 1127 GWS interactions, five were
independent (LD r^2 < 0.01, window 5 Mb) and were located in
regions that have well-known vitamin D-related genes in
chromosomes 11 (e.g., CYP2R1 region) and 14 (e.g., SEC23A
(Supplementary Data 14). Notably, of the 20 vQTL loci without
significant GxE with season, at least half showed no evidence at all
for GxE with season (Supplementary Fig. 6), so these variants
are candidates for GxE with other environmental factors.

**Discussion**

We have identified 143 loci associated with 25OHD concentra-
tion. Recognising that only six associated loci had been reported
to date, these discoveries provide important insights into pre-
viously unknown or poorly understood vitamin D-related path-
ways, and substantially increase our knowledge of the genetic
correlates of 25OHD compared with previous studies7 (Fig. 4).
First, the three most associated loci, all identified in previous
studies8, are noteworthy (chr4:rs1352846, chr11:rs116970203 and
chr11:rs12794714, all with association test P < 1.0 × 10^{−400},
with their minor allele reducing 25OHD). rs1352846 (MAF =
0.29 (G)) is in the GC locus8,20, which encodes a protein syn-
thesised in the liver that binds to, and transports vitamin D and
its metabolites. rs116970203 is a low-frequency variant (MAF =
0.03 (A)) located in intron 11 of the PDE3B gene. It is also a
perfect proxy for rs117913124 (LD r^2 = 1), a low-frequency
synonymous coding variant in CYP2R1, which was previously
reported to associate with 25OHD19. Another CYP2R1 syn-
onymous variant was also identified (rs12794714; MAF = 0.42 (A)).
CYP2R1 encodes a crucial hepatic enzyme involved in the
hydroxylation of vitamin D to 25OHD. Given the complexity of
the association pattern observed in chromosome 11, we con-
firmed the independence of the COJO identified variants using
individual-level data (Supplementary Table 6). In line with pre-
vious findings19, the two-way conditional analysis showed that the
effect of the low-frequency SNP (rs116970203 or
rs117913124) and common SNP (rs12794714 or rs10741657) were
largely independent.

Our findings provide convergent evidence that genes related to
lipid- and lipoprotein-related pathways influence 25OHD con-
centration. In particular, we confirm a unidirectional relationship
between SNP instruments that influence higher BMI and lower
25OHD concentration, but not the reciprocal relationship. This
relationship exists against a background of a highly inter-
related pattern of relationships between genes that influence
both 25OHD and a wide range of lipid-related metabolic phe-
notypes. There were variants within genes with well-described
functions related to lipid and lipoprotein-related pathways35 (e.g.,
PCSK9, DOCK7, CELSR2, GALNT2, ABCA1, DGAT2, CETP,
APOE, APOC1, PLA2G3). In addition, several inter-genic loci had closest upstream or downstream genes of interest to lipid and lipoprotein pathways (AKR1A, APOB, CETP, LIPG, LDLR). Variants in these genes influence overall lipid concentrations, including the concentration of 7-dehydrocholesterol in the skin. We identified a locus (chr11:rs12803256) in an uncharacterised RNA gene (FLJ42102) 11,057 base pairs upstream from DHCR7. This region has been identified in previous GWAS studies, and DHCR7 is a strong candidate gene because of its known role in the conversion of 7-dehydrocholesterol in the skin to pre-vitamin D3. We note that the broad region on Chr11 containing DHCR7 and NADSYN1 included several loci of interest according to both GCTA-COJO and SMR analyses—this complex area warrants additional research.

The GWAS uncovered a range of previously unreported findings, indicating that properties of the skin not related to pigmentation are associated with 25OHD concentration. While it is well known that individuals with darker skin tend to have lower 25OHD (related to the melanin content in the skin blocking UVB)1, our findings provide evidence that SNPs associated with genes that influence dermal development (e.g., PADI36 and integrity (e.g., FLG, FLG-AS1, POU2F3, KLK10, DSG1)37,38 are also associated with 25OHD status. It has been suggested that variants in the FLG gene may have evolved in order to optimise 25OHD production at high latitude39,40. HAL (histidine ammonia-lyase) codes for an enzyme that deaminates L-histidine to trans-uronic acid. The top SNP in this region (rs10859995) is within an intron of this gene. The gene is expressed in the skin, and is upregulated during keratinocyte differentiation41. It has been demonstrated that trans-uronic acid in the stratum corneum can absorb UVB42 and can reduce the production 25OHD43. The MAGMA gene-set analysis22 also showed that variants associated the uronic acid pathways were significantly overrepresented in our findings (Supplementary Data 5). The concentration of trans-uronic acid varies widely between individuals43,44, but is not related to skin colour/pigmentation44. It is
important to note that our sample was restricted to Europeans, and analyses included 40 ancestry PCs as covariates, four of which were strongly associated with 25OHD (Supplementary Table 1). If these PCs capture variants related to skin colour within Europeans, these variants are less likely to be identified in our analyses. FUMA analyses did not identify an over-representation of variants known to be related to skin colour in our GWAS.

Our study expands the range of enzymes implicated in the synthesis and breakdown of vitamin D-related molecules. These include genes from the hydroxysteroid 17-beta dehydrogenase family (HSD17B1, HSD3B1), a family of short-chain dehydrogenases/reductases, which are involved in steroidogenesis and steroid metabolism. CYP2R1 is a key regulator of 25OHD status, via hepatic conversion of vitamin D to 25OHD—two loci were found within this gene. Other members of this large family of enzymes associated with 25OHD concentrations include CYP7A1, CYP26A1 and CYP24A1.

We identified many variants within genes related to the modification of lipoprotein molecules (including seco-steroids, such as 25OHD and related species). Associated regions on chromosomes 2 and 4 include enzymes in the UDP-glucuronosyltransferase family, which are critical in the glucuronidation pathways. The involvement of these genes in the degradation and potential conjugate recycling of 25OHD has recently been described. We identified variants in the SULT2A1 gene, which encodes the enzyme responsible for the sulphonation of 25OHD. Our findings provide support for the hypothesis that these mechanisms influence 25OHD concentration. We identified variants in the SLCO1B1 gene, which encodes a transmembrane receptor that mediates the sodium-independent uptake of numerous endogenous compounds, including sulfated steroid molecules. It is not known if this mechanism is involved in the uptake of the sulphated 25OHD. It has been proposed that vitamin D may undergo conjugate cycling (e.g., bidirectional conversion between 25OHD and 25OHD-sulphate). A proportion of the total 25OHD may exist in the sulphated form, which could act as circulating reservoir for later de-sulphation in peripheral tissues. In addition, conjugated versions of 25OHD with glucuronide and sulfate have both been detected in bile, which suggests enterohepatic mechanisms may provide another reservoir that buffers total 25OHD reserves. The findings also have implications for how to assay total 25OHD reserves. Current extraction and assay techniques used to quantitate 25OHD are not optimised for sulphonated or glucuronidated species of 25OHD, thus total 25OHD status may not accurately reflect the contribution of these conjugated species. In addition, these mechanisms would contribute to the functional half-life of 25OHD, and thus influence vitamin D status during periods of reduced exposure to bright sunshine (e.g., during winter). Finally, variants in a range of previously unreported enzymatic pathways were also associated with 25OHD concentration (e.g., short-chain dehydrogenase/reductase, aldehyde dehydrogenase, alcohol dehydrogenase).

The large sample size afforded by the UKB sample provides a thorough description of the genetic architecture of 25OHD. The SNP-based heritability (which captures the contribution to variation between people associated with common DNA variants) estimated in the UKB was 0.13 (95% CI: 0.12–0.14), which means that all genotyped/imputed variants with MAF > 0.01 explain about 41% of the heritability estimated from close relatives (i.e., 0.13/0.32, SNP-based heritability/heritability, Fig. 2). The 143 loci represent only 112 1-Mb regions, with six of the 1-Mb regions harbouring four loci each. The full set of 143 loci was achieved by applying the COJO (conditional and joint) algorithm onto the GWAS summary statistics using the linkage disequilibrium structure to account for the correlation structure between SNPs. Two regions on chromosome 11 are particularly complex, Supplementary Data 3). Polygenic score predictor using SNP effect weights estimated in the UKB explained up to 10.5% and 5.7% of variance (after accounting for covariates) in independent samples QIMR and UKB. Aside from sampling differences, the higher variance explained in the Australian QIMR sample is in line with the higher heritability estimated from family data (QIMR 0.50 (95% CI: 0.38–0.64) vs 0.32 (95% CI: 0.30–0.34) estimated in the UKB. One explanation for the difference is that the QIMR samples were predominantly recruited at latitude 27° S; at this latitude, there is sufficient UVR to allow for vitamin D synthesis throughout the year.

We also identified 25 independent SNPs associated with variance in 25OHD—these are putative GxE loci. While five of these have strong evidence of interacting with season of measurement, at least ten are GxE candidates with yet-to-be-identified environmental risk factors, and search of published GWAS results for association with these SNPs (i.e., PheWAS34) may help with this prioritisation (Supplementary Data 13). In summer months, the 25OHD concentrations are higher, and a larger proportion of the variance could be attributed to genetic factors in summer compared with winter (SNP-based heritability of 0.19, s.e. = 0.02, vs 0.10, s.e. = 0.02, \( P_{\text{diff-cont}} = 1.5 \times 10^{-3} \)), and this reflected an increase in genetic variance rather than a decrease in residual variance (Supplementary Data 1). However, the genetic correlation from summer and winter SNP effect sizes was not significantly different from 1. Five loci were identified as significant in GxE analysis with season, and for two the direction of effect was reversed (Supplementary Data 14). The vitamin D phenotype is an interesting one to explore from the perspective of GxE as seasonal fluctuations provide a natural experiment to dissect components of the genetic architecture that influence synthesis (i.e., inflow) and excretion (i.e., outflow) of 25OHD-related pathways.

In the UKB participants, high BMI is associated with reduced 25OHD concentration, and keeping with a large body of observational epidemiology. However, we did not find statistical evidence in support of a causal role for the 25OHD level on BMI. In contrast, there was evidence for pleiotropic effects of SNPs on the two traits as well as for high BMI being causal (directly or indirectly) for low 25OHD. Genetic correlations were significant between 25OHD concentration and a range of phenotypes (Fig. 5). However, in robust directional models, we found no evidence in support of a causal role for 25OHD concentration on these traits. Of interest, we found evidence that higher intelligence and an increased risk of several psychiatric disorders may cause reduced 25OHD concentrations. With respect to intelligence, this would be consistent with previous links between intelligence and years of education leading to working indoors, and subsequent lower concentrations of 25OHD. One of our motivations for undertaking this study was to investigate the hypothesis of a causality relationship between 25OHD and psychiatric disorders. The Mendelian randomisation analyses conducted here do not support a causal role for 25OHD levels and these disorders, and hence the reported epidemiological associations could reflect confounding and/or reverse causation. Vitamin D deficiency is common in those with established psychiatric disorders, as a consequence of reduced outdoor behaviour. It is feasible that the observed association between 25OHD concentration in blood spot samples taken at birth with later-life increased risk of schizophrenia and could be confounded by outdoor behaviour of mothers, which may be correlated with the mother’s genetic liability to schizophrenia. While we find no evidence to support the hypotheses that variants associated with low 25OHD concentrations were associated with any of the selected phenotypes, we note
that there is a linearity assumption in our Mendelian randomisation analyses. In other words, if only very low concentrations of 25OHD are associated with adverse outcomes, then this non-linear exposure-risk association may not be confidently detected (see Supplementary Note 2 for further discussion). As the genetic architecture of the broad range of vitamin D-related phenotypes becomes better understood, issues related to potential threshold effects (e.g., disease-specific thresholds for clinical deficiency) should be re-examined.

We have identified 143 loci associated with 25OHD concentration, and have provided new directions for vitamin D research. In particular, our findings suggest that pathways related to sulphonation and glucuronidation warrant closer scrutiny—for example, there may be a case to measure these modified species of 25OHD and related molecules in order to better understand vitamin D status. Our studies based on Mendelian randomisation do not support hypotheses that vitamin D concentration is associated with a broad range of candidate phenotypes, in particular, psychiatric disorders. The findings provide insights into the physiology of vitamin D and the relationship between 25OHD status and health.

Methods

The UK Biobank sample. The UK Biobank (UKB) is a large population cohort with phenotype, genotypic and clinical information on more than 500,000 individuals (age range from 40 to 69 years old). Participants were registered with the National Health Service, and lived ~25 miles from one of the 22 recruitment centres across the UK9. Participants were recruited between 2006 and 2010. Informed consent was obtained by UK Biobank from all participants, and the study was approved by the North West Multicentre Research Ethics Committee. The participants of the study were not representative of the European reference cluster. Among these, there was a set including all pairs with genetic similarity being >0.9 posterior probability of belonging to the 1KGP European reference cluster.

Assessment of 25-hydroxyvitamin D concentration. Vitamin D 25OHD levels were measured in blood samples collected at two instances: the initial assessment month, assessment centre, supplement-intake information, genotyping batch and the first 40 ancestry PCs as covariates in the model (see Supplementary Methods for more details).

We identified independent associations, we conducted a conditional and joint (COJO; gcta --cojo-scl) analysis10 of the GWAS results, accounting for the correlation structure between SNPs within a 10-Mb window (COJO default parameter) and using a random subset of 20,000 unrelated Europeans from the UKB as linkage disequilibrium (LD) reference. For comparison, we used PLINK1.9 (clump --clump-snps for regional lead SNPs for genome-wide significant index variants (--clump-p 1e-8) and variants were clumped with this lead SNP if they were located less than 10 Mb (~clump-kb 10000) away from, and with r² > 0.01 (~clump-r2 0.01) with the index variant. To identify previously unreported associations, we conducted a conditional analysis on the six loci previously reported as genome-wide significant (rs2282679, rs10741657, rs17758785, rs10745742, rs8018720, rs6013897)2,10,25.

Meta-analysis. The largest GWAS for 25OHD to date, from the SUNLIGHT consortium8, used BMI as a covariate, hence we also generated UKB results including BMI in the model and used those for meta-analysis. In addition, the UKB GWAS results used for meta-analysis differed from the reported GWAS results in that 25OHD levels were natural-log transformed, and supplement intake was not included as a covariate. Before meta-analysis, we imputed the SUNLIGHT summary statistics (2,579,297 SNPs) with ImpG16. After data management, we used a sample size-based approach27 to perform the meta-analysis (Supplementary Methods) on 6,912,294 SNPs that were shared between the data sets.

Relationship between vitamin D and body mass index traits. High BMI is associated with lower concentrations of 25OHD18. For this reason, previous GWAS of 25OHD have included BMI as covariate in their analyses8. However, given that BMI is a highly heritable trait, covariate adjustment can induce collider bias15 and affect downstream analyses. To better understand the relationship between 25OHD and BMI, we estimated the phenotypic and genetic correlation between them and used generalised summary-data-based Mendelian randomisation (GSMR)18 to test for statistical evidence for putative causal effects between the two traits. We confirmed through simulation that the MR regression statistic is not biased by sample overlap (Supplementary Note 5). SNP instruments were selected with the default settings of the built-in GSMR clumping step (which is less stringent than used in our clumping protocol because GSMR accounts for residual correlation between SNP instruments). In addition, we conducted a multi-trait conditional and joint (mtCOJO) analysis16 to condition the 25OHD GWAS results on BMI GWAS summary statistics generated with the UKB17, an approach that was shown in simulations to be robust to induced collider bias when conditioning on a correlated trait16. A random subset of 20,000 unrelated individuals of European ancestry from the UKB was used as LD reference in the mtCOJO analysis.

Heritability and SNP-based heritability. Our UKB sample included a set of 58,738 individuals who were related with coefficient of relationship (r) > 0.2 to at least one other person in the set (all relatives). Among these, there was a set including all pairs with 0.4 < r < 0.6 (1st degree), and a set including all pairs with 0.3 < r < 0.3 (2nd degree). We used these sets to estimate heritability of RINT (25OHD) levels, using PLINK2 (--hard-call 0.1). To estimate SNP-based heritability, we conducted a random subset (~N = 50,000), selected so that no pair of individuals had r > 0.05. We used a model that fits a single random genetic effect with a single genomic relationship matrix (GRM) constructed from all SNPs69, and also a GREML-LDMS (GremlDm8) model that fits ten random genetic effects and hence ten GRM (~N = 50,000) matrices.--ngm). The ten GRM were constructed from SNPs annotated to five MAF (0.01–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4 and 0.4–0.5) bins each divided into two by median LD score of the SNPs within the bin. The LD score of a SNP is a measure of the common genetic variation tagged by a SNP. The sum of the estimates for each MAF-LD bin is an estimate of the total SNP-based heritability. Under a neutral model, each of the five MAF bins is expected to explain 20% of the variance. Analyses were conducted with and without BMI as a covariate, and genetic correlation between 25OHD and BMI was estimated in a bivariate GREML analysis (gcta --reml-bivar). In addition, we estimated the genetic correlation and the genetic variance explained by 25OHD levels assessed in summer and winter (see definitions in vQTL and seasonal analysis section below), using bivariate GREML. Heritability and SNP-based heritability estimated as part of the GWAS analysis using fastGWAS are also reported. Finally, we estimated SNP-based heritability by LD score regression28 (software default settings for European ancestry samples), SBayesR45 and SBayesS45 from GWAS summary statistics. From SBayesS, we also estimate the polygenicity (π) and selection (S) parameters.

Replication and out-of-sample genetic risk prediction. We used the QIMR Brisbane-based twin and family sample (N = 1,632 unrelated) for replication analyses. Samples were collected between May 1992 and January 2014, mostly from South-East Queensland (latitude 27° S). At this latitude, there is sufficient UVR to allow for vitamin D synthesis throughout the year29. Legal guardians gave written, informed consent prior to study inclusion and testing. Studies were approved by the Human Research Ethics Committee of the QIMR Berghofer Medical Research Institute. Additional details of this study are provided elsewhere (25OHD analysis...
methods and genotyping on HumanCoreExome-12v1-0_C or IlluminaHuman610-WQuad bead chip and quality control16. Genotypes were imputed using the 1000 Genomes v5 of the 1000 Genomes project (Fig1)12. The phenotype analysed was RINT(25OHD) pre-regressed on sex, age, month of collection, ten ancestry PCs and imputation batch. Association analysis was conducted in PLINK for the genome-wide significant COJO SNPs from the UKB analysis. We tested for the same sign of effect size between the UKB and the QIMR results. Next, using the same set of SNPs, we performed associated polygenic risk scores in the Brisbane cohort, we selected independently associated SNPs from the UKB cohort in order to conduct standard P-value thresholding PRS analysis, choosing a range of P-value thresholds (P < 5 × 10⁻⁵, P < 1 × 10⁻⁴, P < 0.01, P < 0.05). PRS were calculated using PLINK26 for each individual in the QIMR cohort. We also calculated polygenic scores from SNP weights estimated by COJO27, SBayesR22 and SBayesS13. The Bayesian methods better account for the complex relationship between strength of association of, and correlation between, SNP effect sizes. For each set of polygenic scores, we estimated the proportion of variance explained by the scores in linear regression. We repeated these analyses in a sample of 1632 individuals provided in the LDCS(ChB, UKR), whose PC1 value was just outside our cut-off for white European (referring to as the UKBR sample; Supplementary Methods).

Functional mapping and annotation of GWAS. We conducted a number of analyses to annotate the 25OHD GWAS results. First, we used the FUMA online pipeline to obtain gene-based, gene-set and tissue-specific annotations. Second, we used functional annotations provided in the LDSC software to partition SNP-based heritability into 53 functional categories25. Annotations included elements, such as UCSC CILT (P < 0.1) and calculating polygenic scores for each individual in the QIMR cohort. We also calculated polygenic scores from SNP weights estimated by COJO27, SBayesR22 and SBayesS13. The Bayesian methods better account for the complex relationship between strength of association of, and correlation between, SNP effect sizes. For each set of polygenic scores, we estimated the proportion of variance explained by the scores in linear regression. We repeated these analyses in a sample of 1632 individuals provided in the LDCS(ChB, UKR), whose PC1 value was just outside our cut-off for white European (referring to as the UKBR sample; Supplementary Methods).

Genetic correlations and putative causal relationships with other traits. Published epidemiological studies have provided an extensive set of hypotheses about causal relationships between vitamin D and a range of phenotypes, including psychiatric and brain-related disorders19 including intelligence19. To characterize the relationship between vitamin D and psychiatric traits, we conducted two sets of analyses. First, we used bivariate LD score regression20 to estimate the genetic correlation between vitamin D and psychiatric traits using the GWAS summary statistics generated with the three levels of BMI (N = 135,504 genome-wide significant eQTLs). In general, SNPs controlling variation in one tissue are found to control eQTLs in these tissues, we used GTEX eQTL data sets, despite other relevant tissues are the liver, skin and, given our hypotheses about the relationship between 25OH and psychiatric disorders, the brain. To help prioritise putative causal genes with expression underlying 25OHD levels, we used the summary-data-based Mendelian randomisation method (SMR)26. SMR integrates GWAS and eQTL (expression quantitative trait loci, SNPs associated with gene expression) results with the aim of identifying pleiotropic or causal associations between the trait of interest and gene expression. We used eQTLs derived by the eQTLGen consortium from gene expression in whole blood25, using the a large sample for blood eQTLs (N = 1,684), identifying 15,504 genome-wide significant eQTLs. In general, SNPs controlling variation in one tissue are found to control expression in other tissues20, hence using a large eQTL dataset is the most powerful approach. Moreover, blood is a relevant tissue for vitamin D-related gene transcription20. Other relevant tissues are the liver, skin and, given our hypotheses about the relationship between 25OH and psychiatric disorders, the brain. To capture tissue-specific eQTLs in these tissues, we used GTEx eQTL data sets, despite the fact that these data sets are much smaller than the eQTLGen sample (snn-capture tissue-specific expression and functional genomic annotations constructed using ENCODE21 and Roadmap Epigenomics Consortia dataset22). Third, we assessed the SNP-based heritability enrichment associated with different cell types. Specifically, we applied LDSC analysis to the GWAS summary statistics using scores associated with cell-type-specific expression in the LDCS(ChB, UKR).

Proxy-environment vQTL and seasonal analysis. 25OH concentration is known to be affected by season of measurement, but other environmental factors may also impact 25OH levels. We conducted a genome-wide vQTL analysis6,7,8 to identify SNPs associated with variance in 25OH. Specifically, we used the Levene’s median test implemented in OSCA8. Following the guidelines of Wang et al.7, we (1) adjusted 25OH levels for selected covariates (see below), (2) removed outliers more than 5 SD from the mean and (3) standardised the residuals to have mean 0 and variance 1. Each step was performed within one of eight groups defined based on sex (male vs. female) and supplement intake (none, other, vitamin D or missing). This approach removed both the mean effect of covariates and the mean and variance differences between gender and supplement intake groups, while retaining other distributional properties of the measure. Covariates included in the phenotype pre-regression were age at assessment, assessment month, assessment centre, genotyping batch and the first 40 PCs. To avoid spurious associations due to coincidence of low-frequency variants with phenotype outliers, this analysis was restricted to SNPs with MAF >0.05. To identify near-independent vQTLs, we clumped the vQTL GWAS results with PLINK1.9 (-clump) as above, using a 5 Mb window as recommended12.

To assess if significant vQTL associations reflected a GxE with season of testing, we conducted season-stratified GWASs and compared the results with the vQTL GWAS results. Specifically, we stratified the UKB cohort into two groups after visual inspection of the mean 25OH levels per month (Supplementary Fig. 1b). We defined two discrete time periods in order to retain the maximum sample size, but optimise comparisons between months with higher and lower mean 25OH concentrations: (a) Winter—individuals assessed Dec-Apr (N = 162,591), and (b) Summer—individuals assessed Jun-Oct (N = 177,082). Individuals with vitamin D levels assessed in the months of May and November were not included in these analyses. The two season-stratified GWASs (winter and summer) were conducted as the main GWAS (i.e., linear mixed model implemented in fastGWA, with the same covariates included in the model). In addition, we conducted a second analysis using a month-by-month interaction term with season of blood draw. For this analysis, the phenotype (25OH levels) was processed as described for the vQTL analysis.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Genome-wide association summary statistics generated with the three levels of BMI correction (i.e., with and without BMI as covariate, and conditioned on BMI) are available for download from https://csgenomics.com/content/data. Results for the UKB GWAS of BMI used for conditional analysis are also available from the same website. UK Biobank data were obtained through direct application to the UK Biobank. The SUNLIGHT data were downloaded from https://drive.google.com/drive/folders/0bYzD95cDoHhRRKRO2hH3EzVzWQ. Functional annotations to partition SNP-based heritability into tissue with genetic loci were downloaded from https://data.broadinstitute.org/alkesgroup/LDSCCORE/ eQTL data were downloaded from http://www.epglen.com/cis-eqtls.html and https://csgenomics.com/software/smrt/DataResource. GWAS summary statistics used for bidirectional SMR were downloaded from https://walters.psycm.cf.ac.uk (schizophrenia), https://csgenomics.com/content/data (type II diabetes), https://ctg.crg.cat/dds/walters/Psychiatric/GWAS/WALCARE-2019 (Alzheimer’s disease, fluid intelligence, ADHD), https://www.thessgas.org/data (educational attainment), https://www.med.uni.edu/ppg/download-results (bipolar disorder, autism spectrum disorder), http://plaza.umin.ac.jp/~yokoda/dataset/software.htm (rheumatoid arthritis), ftp://ftp.ebi.ac.uk/pub/ukbiobank/data/gwas/statistics/vanderHarstP_29212778_GCST005194 (coronary artery disease), https://www.ibdgenetics.org/downloads.html (inflammatory bowel disease). All other data are contained in the article and its supplementary information, or are available on request.

Received: 27 November 2019; Accepted: 3 March 2020; Published online: 02 April 2020.

References
1. Holick, M. F. Vitamin D deficiency. N. Engl. J. Med. 357, 266–281 (2007).
2. Institute of Medicine. Dietary Reference Intakes for Calcium and Vitamin D (National Academies Press, 2010).
3. Lips, P. Worldwide status of vitamin D nutrition. J. Steroid Biochem. Mol. Biol. 121, 297–300 (2010).
4. Karohl, C. et al. Heritability and seasonal variability of vitamin D concentrations in male twins. Am. J. Clin. Nutr. 92, 1393–1398 (2010).
5. Mills, N. T. et al. Heritability of transforming growth factor-beta 1 and tumor necrosis factor-receptor type 1 expression and vitamin D levels in healthy adolescent twins. Twin Res. Hum. Genet. 18, 28–35 (2015).
6. Mitchell, B. L. et al. Half the genetic variance in vitamin D concentration is shared with skin colour and sun exposure genes. *Behav. Genet.* **49**, 386–398 (2019).

7. Jiang, X., Kiel, D. P. & Kraft, P. The genetics of vitamin D. *Bone* **126**, 59–77 (2018).

8. Jiang, X. et al. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nat. Commun.* **9**, 260 (2018).

9. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

10. Hypponen, E. & Power, C. Hypovitaminosis D in British adults at age 45 y: nationwide cohort study of dietary and lifestyle predictors. *Ann. J. Clin. Nutr.* **860**, 868–875 (2007).

11. Eyles, D. W. et al. The association between neonatal vitamin D status and risk of schizophrenia. *Sci. Rep.* **8**, 17692 (2018).

12. Wang, H. et al. Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci. Adv.* **5**, eaaw5358 (2019).

13. Zeng, J. et al. Bayesian analysis of GWAS summary data reveals differential signatures of natural selection across human complex traits and functional genomic categories. Preprint at https://www.biorxiv.org/content/10.1101/75227v1 (2019).

14. International HapMapConsortium et al. Integrating common and rare genetic variants in diverse human populations. *Nature* **467**, 52–58 (2010).

15. Day, F. R., Loh, P. R., Scott, R. A., Ong, K. K. & Perry, J. R. A robust example of collider bias in a genetic association study. *Am. J. Hum. Genet.* **98**, 392–393 (2016).

16. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Commun.* **9**, 260 (2018).

17. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **46**, 369–375 (2014).

18. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375, 513–517 (2012).

19. 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**, 227–238 (2017).

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

21. Lee, M. J. et al. Vitamin D deficiency in northern Taiwan: a community-based cohort study. *BMJ Public Health* **19**, 337 (2019).

22. Eyles, D. W., Burne, T. H. J. & McGrath, J. J. Vitamin D effects on brain development, adult brain function and the links between low levels of vitamin D and neuropsychiatric disease. *Front. Neuroendocrinol.* **34**, 47–64 (2013).

23. Adamson, J. et al. Correlates of vitamin D in psychosomatic disorders: a comprehensive systematic review. *Psychiatry Res.* **249**, 78–85 (2017).

24. McGrath, J. J. et al. Neonatal vitamin D status and risk of schizophrenia: a population-based case-control study. *Arch. Gen. Psychiatry* **67**, 889–894 (2010).

25. Fry, D., Almond, R., Moffat, S., Gordon, M., & Singh, P. B. UK Biobank: Companion Document to Accompany Serum Biomarker Data. UK Biobank Document Showcase (2019). Available at: https://biobank. nindh.ox.ac.uk/showcase/showcase/docs/serum_biochemistry.pdf. Accessed Mar 2020.

26. Jiang, L. et al. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51**, 1749–1755 (2019).
Acknowledgements
This study was carried out under the generic approval from the NHS National Research Ethics Service and conducted using the UK Biobank resource under projects 12505 and 10214. We thank the UKB participants, project team and funders for providing this important research resource. We thank the eQTLGen consortium for providing the cis eQTL dataset based on N = 32 K participants. We thank 23andMe for the use of GWAS summary statistics for major depression that include data from 23andMe. We would like to thank the research participants and employees of 23andMe for making this work possible. The Genotype-Tissue Expression (GTEX) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NHGRI, NHLBI, NIDA, NIHM and NINDS. Funding for the QIMR sample was provided by the Australian National Health and Medical Research Council (NHMRC) and further supported by NHMRC Project Grants (108767/109709) and a John Cade Fellowship (1056929). NHMRC also support Naomi Wray (1113400, 1078901), Peter Visscher (1113400, 1076037) and Jian Yang (1113400). Jian Yang is supported by the Australian Research Council (FT180100186). John McGrath is supported by the Danish National Research Foundation (Niels Bohr Professorship, the NHMRC (John Cade Fellowship 1056929). John McGrath, Darryl Eyles and Thomas Burne are employed by The Queensland Centre for Mental Health Research which receives core funding from the Queensland Health. Darryl Eyles is supported by the NHMRC (1134724, 1124721, 1141699). Brittany Mitchell received financial support from the Queensland University of Technology.

Author contributions
J.A.R., J.J.M. and N.R.W. conceived the study and designed the analyses. J.A.R., T.L., Z.Q. and B.M. conducted the analyses. K.E.K. and J.S. performed the initial preparation and quality control of the UK Biobank data. A.X., Y.H., Z.Z., J.Z., H.W., A.A.E.V. and G.Z. provided support in analysis implementation. J.F., D.E. and T.H.J.B. helped with interpretation of identified loci. N.G.M. provided the QIMR cohort, and B.M. and G.Z. conducted the analyses based on this sample. P.M.V. and J.Y. provided advice on analyses and interpretation of the results. J.A.R., J.J.M. and N.R.W. wrote the paper with the participation of all the authors. All authors reviewed and approved the final paper.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-15421-7.

Correspondence
Correspondence and requests for materials should be addressed to N.R.W. or J.J.M.

Peer review information Nature Communications thanks Dominic Furniss, Fernando Rivadeneira and Xia Jiang for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020