

***OSBPL11* is an African-specific locus associated with 25-hydroxyvitamin D concentrations and cardiometabolic health**

Reagan M Mogire^{1,2§}, John Muthii Muriuki¹, Ruth Fiona Bayimenye³, Alexander J Mentzer^{4,5}, Amanda Chong⁴, Mateus H Gouveia², Amy R. Bentley², Gavin Band⁴, Pauline King'ori¹, Ruth Mitchell⁶, Emily L Webb⁷, Dhriti Sengupta⁸, Lynette Ekunwe⁹, Francis M Ndungu¹, Alireza Morovat¹⁰, Alex W Macharia¹, Clare L. Cutland¹¹, Gibran Hemani⁶, Sodiomon B Sirima¹², Michèle Ramsay^{8,13}, Camila A. Figueiredo¹⁴, Andrew M. Prentice¹⁵, Shabir A Madhi¹⁶, Laura M. Raffield¹⁷, Manjinder S. Sandhu¹⁸, Philip Bejon^{1,19}, George Davey Smith⁶, Alison M Elliott²⁰, Thomas N Williams^{1,21}, Charles Rotimi², Christina R. Bourne²², Anthony Burgett³, Adebawale Adeyemo^{2†}, and Sarah H Atkinson^{1,19,2†§}

Authors' affiliations

¹Centre for Geographic Medicine Research-Coast, KEMRI-Wellcome Trust Research Programme, Kenya Medical Research Institute (KEMRI), Kilifi, Kenya.

²Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, 12 South Drive, MSC 5635, Bethesda, Maryland 20891-5635, USA.

³Department of Pharmaceutical Sciences, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73117 USA

⁴Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, UK.

⁵Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK.

⁶Medical Research Council (MRC) Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

⁷MRC International Statistics and Epidemiology Group, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.

⁸Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2193, South Africa

⁹Department of Medicine, University of Mississippi Medical Center (UMMC), Jackson, MS 39213, USA

¹⁰Department of Clinical Biochemistry, Oxford University Hospitals, Oxford, UK.

¹¹African Leadership in Vaccinology Expertise (Alive), Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

¹²Centre de Recherche Action en Sante (GRAS), 06 BP 10248, Ouagadougou 06, Burkina Faso.

¹³Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

¹⁴Departamento de Ciências da Biointeração, Instituto de Ciências da Saúde, Universidade Federal da Bahia, 40110-100, Salvador, Bahia, Brazil

¹⁵MRC Unit The Gambia at London School of Hygiene and Tropical Medicine, Banjul, The Gambia.

¹⁶South African Medical Research Council: Vaccines and Infectious Diseases Analytical Research Unit (VIDA), Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

¹⁷Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

¹⁸Department of Epidemiology & Biostatistics, School of Public Health, Imperial College London, London, UK

¹⁹Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK.

²⁰Medical Research Council/Uganda Virus Research Institute and London School of Hygiene & Tropical Medicine Uganda Research Unit, Entebbe, Uganda

²¹Institute of Global Health Innovation, Department of Surgery and Cancer, Imperial College, London, UK.

²²Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019, United States

²³Department of Paediatrics, University of Oxford, Oxford, UK.

§ Corresponding authors.

†These authors contributed equally to this work.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract [150 words max]

Vitamin D deficiency is prevalent in Africa, but its genetic determinants are understudied. We report a genome-wide analysis of 25-hydroxyvitamin D (25(OH)D) concentrations in 3670 children from five countries across Africa with replication in four diaspora African ancestry populations (n=21,610). We identify a previously unreported locus at genome-wide significance in West African populations: *OSBPL11* (Oxysterol Binding Protein Like 11, lead variant, rs2979356, $p=8.01 \times 10^{-9}$). *In vitro* and molecular docking assays showed that OSBPL11 is a vitamin D binding protein likely involved in the intracellular binding of vitamin D metabolites. *OSBPL11* knockdown mice have increased fat, reduced triglycerides and improved glucose tolerance, and rs2979356 was associated with cardiometabolic health in adults of African ancestry. We also identify previously reported loci determining vitamin D status. Our study advances understanding of vitamin D genetics in Africa and indicates a novel function for *OSBPL11* in intracellular binding and transport of vitamin D metabolites.

Key words

Vitamin D, 25-hydroxyvitamin D, *OSBPL11*, GWAS, Africa, oxysterol

Introduction

[Main text – up to 4,000 words, excluding abstract, methods, references and figure legends]

Nearly one-third of individuals living in Africa have vitamin D deficiency (25(OH)D of <50 nmol/L), which has been associated with many communicable and non-communicable diseases^{1,2}. Vitamin D is essential for calcium and phosphate homeostasis, skeletal health, and immune function, and its deficiency has been linked to bone disorders, increased risk of infection, metabolic syndrome, and cardiovascular disease². However, the genetics of vitamin D status in Africa remains understudied. Vitamin D genome-wide association studies (GWAS) have been conducted almost entirely in populations of European ancestry^{3,4}, while existing knowledge of the genetics of vitamin D status in continental Africans is limited to a few gene-based studies^{5,6}. Several vitamin D GWASs have been conducted in individuals of African ancestry in the USA and UK⁷⁻⁹, however diaspora African populations have approximately 20% European admixture, and many have different lifestyles and diets compared to continental Africans¹⁰.

Due to their greater genetic diversity, continental African populations provide a unique platform for investigating the genetic influences on vitamin D status¹¹. This genetic diversity increases the likelihood of identifying novel vitamin D loci, which may potentially improve our understanding of vitamin D metabolism and physiology. Furthermore, African populations have relatively smaller linkage disequilibrium blocks, which may improve fine mapping of known loci and facilitate the identification of causal variants and their functional implications in relation to vitamin D status¹². These unique African genetic characteristics highlight the importance of studying diverse populations to gain a comprehensive understanding of the genetic architecture of vitamin D status.

To our knowledge this is the first GWAS of vitamin D concentrations in continental African populations and includes data from 3,670 children living in Kenya, Uganda, Burkina Faso, The Gambia, and South Africa. We identify a previously unreported genome-wide significant locus associated with vitamin D concentrations, *Oxysterol Binding Protein Like 11* (*OSBPL11*). Molecular docking and *in vitro* binding assays show that OSBPL11 binds to vitamin D metabolites with comparable conformations and binding energies as common oxysterols, suggesting that OSBPL11 may have a role in intracellular binding and trafficking of vitamin D metabolites. *OSBPL11* knockdown mice have altered glucose tolerance, body fat and triglyceride levels, and the lead *OSBPL11* SNP was associated with cardiometabolic health in African ancestry individuals in continental and diaspora populations.

Results

Characteristics of discovery sample

In total, 3670 children with a median age of 23.9 months (interquartile range (IQR) 12.7, 26.0) from Kenya, Uganda, Burkina Faso, The Gambia, and South Africa, had 25(OH)D measurements and genome-wide genotypes (Fig. 1A). Median 25(OH)D concentration was 77.3 nmol/L (IQR 63.5, 93.2). The distribution of 25(OH)D concentrations was right skewed as expected and was rank-based inverse-normal transformed to normalise the distribution (Supplementary Fig. 1). Median body mass index was low and the prevalence of malaria parasitemia and inflammation was high across the cohorts (Supplementary Table 1). The principal component analysis (PCA) plot of genotypes showed clustering of study participants by country and geographical region (Fig. 1B).

Discovery genetic association analysis

A total of 18,536,481 variants with minor allele frequency (MAF) > 0.01 were included in the GWAS analyses. Of these, three loci comprising of 61 variants reached genome-wide significance in a meta-analysis of all discovery cohorts ($P < 5 \times 10^{-8}$). The distribution of the association statistics for the genome-wide association analysis is shown in Fig. 1C. The association analyses were conducted by study site and adjusted for age, sex, and season, and incorporated a Genomic Relationship Matrix (GRM) to account for cryptic relatedness. There was little evidence of inflation of the association analyses ($\lambda = 0.99 - 1.01$, Supplementary Fig. 2A). The results of the genome-wide association analyses for the independent lead variants are summarized in Table 1.

Three distinct loci, mapped to genes involved in vitamin D metabolism, reached genome-wide significance: the gene for vitamin D binding protein (group-specific component; *GC*, lead SNP rs1352846), vitamin D 25-hydroxylase/ phosphodiesterase 3B/ proteasome 20S Subunit Alpha 1 (*CYP2R1/PDE3B/PSMA1*, lead SNP rs12290926) and 7-dehydrocholesterol reductase/ NAD Synthetase 1 (*DHCR7/NADSYN1*, lead SNP rs7950991) (Fig 1D, Table 1). All the three loci reached genomewide significance in the overall analysis, but only *GC* and *DHCR7/NADSYN1* reached genome-wide significance in the meta-analyses GWAS that included East African, but not West African populations (Supplementary Fig. 2). The widespread distribution of signals across broad genomic regions within the *DHCR7/NADSYN1* locus may suggest selective pressure at this site in continental African populations (Supplementary Fig. 3). The distribution of the association statistics for the individual discovery and replication cohorts are displayed in Supplementary Fig. 4. SNP based heritability for the common variants was 17.6% (standard deviation 8.4%). We found no additional independent loci that reached genome-wide significance in the meta-analysis after performing conditional and joint analysis (GCTA-COJO) on the GWAS summary statistics or after conditioning the individual cohort association analyses for the lead variants in each of the three loci.

Functional mapping and annotation

Gene-based association analyses showed that the genes that reached genome-wide significance were *GC*, *CYP2R1*, *PDE3B*, *DHCR7*, and *NADSYN1* (Supplementary Fig. 5A, Supplementary Dataset 1). MAGMA gene-set analyses indicated that the top pathways associated with 25(OH)D concentrations were related to intracellular transport to the plasma membrane, vitamin D binding, intracellular signalling, and cell differentiation (Supplementary Dataset 2). Analysis of gene expression data of 53 tissue types from GTEx

v8 revealed enriched gene expression in oesophageal mucosa, skin, vagina, and salivary gland tissue (Supplementary Fig. 5B, Supplementary Dataset 3). To predict the most likely causal variants for the three known loci, we performed fine-mapping using FINEMAP (<http://christianbenner.com/>). *GC*, *CYP2R1/PDE3B/PSMA1*, and *DHCR7/NADSYN1* were estimated to have 26, 24, and 11 possible causal variants, respectively, with posterior probabilities ranging from 0.54 to 1 (Supplementary Table 2). The variants that were predicted to be most likely to be causal had individual posterior probabilities ranging from 0.40 to 1 (Supplementary Dataset 4). Out of the 60 variants present in the credible sets of the three loci, 26 variants had annotations in PolyPhen (<http://genetics.bwh.harvard.edu/pph2/>), one was synonymous, and the rest were intronic.

Meta-analysis with diaspora African populations and transethnic meta-analyses

We performed an African ancestry meta-analysis of vitamin D concentrations that included our discovery GWAS and the Social Change Asthma and Allergy (SCAALA) study in African Brazilians (n=753), African American adults from the Jackson Heart Study (JHS, n=4798) and the Multi-Ethnic Study of Atherosclerosis (MESA, n=1677), adults of African ancestry in the UK Biobank (UKB, n=8151), and precomputed vitamin D GWAS summary data from African ancestry participants in All of Us Research Program (AoS, n=7400)¹³ (Supplementary Table 1). This African ancestry meta-analysis did not reveal any additional genome-wide significant loci (Supplementary Fig. 2B). The three genome-wide significant loci in the discovery GWAS were variably replicated (defined as same direction of effect and $p < 0.05$) across the African ancestry cohorts, although direction of effect remained consistent across cohorts (Supplementary Table 3). These loci were also the signals that were most significantly associated with vitamin D status in Europeans. The lead *DHCR7/NADSYN1* variant, rs7950991 was missing in the European GWAS. However, rs12284909, in high linkage disequilibrium with rs7950991 in our discovery cohorts ($r^2 = 0.98$), was genome-wide significant in Europeans (Supplementary Table 3).

Replication of known 25(OH)D variants in continental Africans

We assessed the transferability of previously reported vitamin D loci in our discovery GWAS. Of 69 independent variants that were genome-wide significant in a recent European vitamin D GWAS³, 66 were present in our discovery GWAS, although with substantial heterogeneity in allele frequency (Supplementary Table 4). Six variants in the *CYP2R1*, *GC*, *SEC23 Homolog A (SEC23A)*, pseudogene *RP13-379L11.3*, and Component of Oligomeric Golgi Complex 5 (*COG5*) genes showed evidence of replication (consistent direction of effect and $P < 0.05$). The strongest signals were observed for the *CYP2R1* and *GC* genes. There was little evidence of replication in the remaining 60 variants.

A novel West African locus in geographical region-specific association analyses

Due to the evidence of population stratification in our discovery cohorts (Fig. 1B), we conducted meta-analyses by geographical region to identify region-specific loci. In a meta-analysis GWAS of West Africans (Burkinabes and Gambians), we identified a novel locus that reached genome-wide significance linked to oxysterol binding protein like 11 (*OSBPL11*). The T allele of variant rs1352846, upstream of *OSBPL11*, was associated with higher circulating 25(OH)D levels ($\beta = 0.391$, $SE = 0.0678$, $P = 8.01 \times 10^{-9}$, Fig. 2A). Among West African children, median 25(OH)D levels varied significantly by rs2979356 genotype; 72.9 nmol/L for CC homozygotes, 80.9 nmol/L for the TC heterozygotes, and 90.6 nmol/L for TT homozygotes, demonstrating a strong allele effect (Fig. 2B). *OSBPL11* encodes a

member of the oxysterol-binding protein (OSBP) family, a group of intracellular lipid receptors¹⁴. Regional LocusZoom plots for *OSBPL11* show that the top variants are in high LD and are distributed across a wide genomic region in West African populations (Fig. 2C and Supplementary Fig. 6A, B). FINEMAP predicted a credible set of 8 causal variants ($k=8$) with a posterior probability of 1 (Supplementary Table 2) and the individual variants in the credible set had posterior probabilities ranging from 0.33 to 1 (Supplementary Dataset 4).

Replication and annotation of *OSBPL11*

In addition to Burkinabe and Gambian children, the lead variant in the *OSBPL11* loci, rs2979356, replicated in Ugandan and Brazilian children of African ancestry, but not in adults of African ancestry in the US and the UK (Fig. 2D). Notably, the rs2979356 T allele was the minor allele in African ancestry populations, with a frequency ranging from 0.11 to 0.31, but the major allele in European ancestry populations, where its frequency was 0.58 (Fig. 2D). The nearest signal in the European GWAS is located 219 kilobases away and is marked by the lead variant rs6438900. This variant is mapped to the *MRPL3* (Mitochondrial Ribosomal Protein L3) gene based on genomic proximity (Supplementary Fig. 6). However, rs6438900 and rs2979356 are in weak linkage disequilibrium in Europeans ($R^2=0.0059$) or Gambians ($R^2=0.0$) (LDpair, see Web Resources), suggesting that these signals are distinct. Both rs6438900 and rs2979356 are significant expression quantitative trait loci (eQTLs) for *OSBPL11* in the Open Targets Genetics database, but only rs2979356 is a significant eQTL for *OSBPL11* in the African Functional Genomic Resource and the JHS eQTL Database (see Web Resources). However, the rs2979356 is not an eQTL lead variant or part of an established eQTL credible set in the Open Targets Genetics database. We found no evidence of epistasis between these variants in the Burkinabe and Gambian cohorts ($P > 0.05$), suggesting no combined effect on the phenotype.

In vitro and molecular docking show vitamin D metabolites are ligands of *OSBPL11*

Due to the strong similarity in molecular structure of vitamin D metabolites and oxysterols, we hypothesized that they might share binding proteins such as *OSBPL11* (Fig. 3A). However, *OSBPL11* protein structures and *in vitro* ligand binding studies have not previously been reported. The AlphaFold model of the *OSBPL11* ligand binding domain shows broad similarities to existing OSBP and *OSBPL2* protein structures^{15,16} and similarly adopts a canonical incomplete beta barrel capped by a likely flexible “lid” loop under which ligands are accommodated (arrow) (Fig. 3B and Supplementary Fig. 7). Docking scores support a high potential to interact with both oxysterols and vitamin D compounds. The side-chain hydroxylated oxysterols were all shown to have only one top scoring docked pose, indicating more constrained binding orientations compared to the vitamin D compounds (Fig. 3C). Importantly, both 25-OHC and calcitriol interact with low nanomolar affinity (inhibition constants [K_i] of 32 nM and 64 nM, respectively) in 293T lysate overexpressing full length *OSBPL11*-myc-His, using a well-established ³H-25-OHC competition binding assay (Fig. 3D)¹⁷⁻¹⁹. These results, which are the first reported *in vitro* ligand binding results of *OSBPL11*, clearly support *OSBPL11* as a vitamin D binding protein and imply that *OSBPL11* may be involved in the intracellular transport of vitamin D metabolites (Fig. 3E).

In silico functional follow up of *OSBPL11*

OSBPL11 is expressed in many tissues and cell types (Supplementary Fig. 8A and B) and eQTL summary statistics for rs2979356 indicate that it is a significant eQTL for multiple genes across various tissues, suggesting its widespread regulatory impact on gene expression

(Supplementary Dataset 5). *OSBPL11* is most highly associated with increased expression in adipocytes, whole blood and monocytes (Supplementary Fig. 9A), and differential expression is also associated with various glucose metabolism and obesity related clinical phenotypes and cardiovascular diseases and obesity (Supplementary Fig. 9B and C, Supplementary Dataset 6). Single cell atlas of human and mouse white adipose tissue reveal that *OSBPL11* is highly expressed in adipocytes, macrophages, endothelial, and circulating mononuclear cells (Supplementary Fig. 10), and rs2979356 is located in super enhancers, chromatin regulatory elements and DNA methylation sites in adipose and other tissues (Haploreg, see Web Resources for links). PheWAS data linked rs2979356 to several health traits, including *OSBPL11* levels (Supplementary Dataset 7, <https://gwas.mrcieu.ac.uk/phewas>). The STRING database shows that *OSBPL11* binds to 25-hydroxycholesterol and is involved in fatty acid metabolism and regulating adipocyte triglyceride storage (<https://string-db.org/>).

Given the suggestive evidence of selection pressure in the region surrounding *OSBPL11* and the role of *OSBP* in modulating inflammatory responses²⁰ we investigated whether this might be driven by protection from severe malaria or tuberculosis (TB), but found little evidence that the rs2979356 variant affects risk of severe malaria ($\beta(\text{SE}) = -0.0396 (0.032)$, $P=0.21$)²¹ or tuberculosis infection (odds ratio (SE) = 0.993 (0.054), $P=0.91$)²² in continental Africans. As of December 2024, there were no data available for rs2979356 in the NHGRI-EBI GWAS catalog, PheGenI, dbGaP, ClinVar, PopHumanScan or in EGA (see Web Resources for links).

***In vivo* mouse studies on *OSBPL11* knockouts**

Heterozygous knockout of *OSBPL11* in mice leads to improved glucose tolerance and increased total body fat with lower circulating triglycerides compared to wild types ($P<0.0001$; International Mouse Phenotyping Consortium—see Web Resources; Supplementary Fig. 11A and B). Other measured phenotypes in heterozygous knockouts showing suggestive evidence ($P<0.05$), include increased fat mass ($P=0.0017$), reduced lean/body weight ratio ($P=0.0029$), and increased bone mineral density ($P=0.0087$). Homozygous *OSBPL11* knockout mice experience preweaning lethality suggesting that *OSBPL11* is essential for early postnatal survival and possibly plays a critical role in developmental processes.

Effect of lead *OSBPL11* variant on cardiometabolic traits in human studies

Considering the association between *OSBPL11* and glucose tolerance, adiposity and triglycerides in mice, we analyzed the effect of rs2979356, the lead variant in this locus, on cardiometabolic traits in continental and diaspora African ancestry populations. We found that rs2979356 was associated with reduced BMI and fat mass in the AADM cohort, as well as with reduced triglycerides in the GLGC and AMP-CMDKP studies and reduced LDL cholesterol in the AWI-Gen study (Table 2). Rs2979356 was also associated with higher risk of hypertension and stroke in the AMP-CMDKP study, and with stroke alone in African ancestry adults in the UKBB. Findings varied between cohorts which might reflect sample size, ancestry or haplotype distributions among populations. We found little evidence of association between rs2979356 and HDL or total cholesterol levels or with fasting glucose or insulin levels (Supplementary Table 5).

Discussion

The high genetic heritability and environmental susceptibility of vitamin D status underscore the importance of studying it across diverse global populations. This is especially relevant in sub-Saharan Africa due to immense genetic diversity, varied lifestyles and environmental conditions, together with a high risk of communicable and non-communicable diseases. In the present study, we report a genome-wide analysis of circulating 25(OH)D concentrations in 3670 children from five countries across Africa. Our study identified three known genome-wide significant loci in genes directly involved in vitamin D metabolism in overall meta-analysis. We also found a previously unreported locus (*OSBPL11*) for 25(OH)D concentrations. The lead *OSBPL11* variant had a large effect on 25(OH)D concentrations in West African children and was also replicated in Ugandan and Brazilian children of African ancestry. Our *in vitro* binding assays revealed that calcitriol binds to OSBPL11 with similar conformation and binding energies as its known ligands, the common oxysterols. Knockdown of *OSBPL11* in mice improved glucose tolerance, increased total body fat and lowered circulating triglyceride levels. The lead variant in the *OSBPL11* loci was also associated with triglyceride levels, LDL cholesterol, fat mass, BMI, hypertension and stroke in African ancestry populations.

The genome-wide significant signals identified in our overall meta-analysis are located within previously established vitamin D metabolizing genes. The *GC* gene encodes the vitamin D binding protein, responsible for the binding and circulation of 80-90% of vitamin D and its metabolites²³. The *GC* locus is the sole locus that reached genome-wide significance in recent GWASs of African American participants^{7, 24}, and was one of the strongest signals in GWASs in European^{3,4}, Hispanic²⁵, Asian²⁶ and mixed ancestry^{8,27} populations, indicating it is a multi-ancestry signal. Epidemiological studies in sub-Saharan Africa show strong associations between various DBP isoforms and vitamin D status^{5,6}. Gene-based association analysis revealed that the *CYP2R1* locus had the strongest effect on 25(OH)D concentrations of all genes, likely due to its encoding of the enzyme 25-hydroxylase, which converts vitamin D to 25-hydroxyvitamin D, the biomarker generally used in GWAS of vitamin D status²⁸. The *DHCR7* gene encodes 7-dehydrocholesterol reductase which diverts 7-dehydrocholesterol towards cholesterol synthesis and away from vitamin D synthesis in the skin²⁸. However, the locus was not genome-wide significant in the replication cohort of diaspora African-ancestry populations, possibly because these populations rely less on sun exposure for vitamin D synthesis and more on vitamin D supplementation and food fortification. Although the three loci are replicated in previous European GWASs, we found that only a small proportion (fewer than 10%) of variants that influence 25(OH)D concentrations in European populations replicate in continental African populations. These differences may be due to genetic heterogeneity due to different causal variants (which is common in complex traits), differences in genetic linkage and LD structure, gene-environment interactions, allele frequency differences or the limitations of sample size of our study²⁹.

Our study identified a novel locus, *OSBPL11*, which was associated with higher 25(OH)D concentrations in West African children with replication in Ugandan and Brazilian children of African ancestry. This effect was not observed in Kenyan or South African children, or adults of African and European ancestry (despite the allele being at a higher frequency in the latter), highlighting both subcontinental heterogeneity within African populations and inter-ancestry differences across the study cohorts. A plausible explanation is that the *OSBPL11* locus exhibits different linkage disequilibrium (LD) patterns across Africa and in African versus European ancestry groups leading to ancestry-specific effects. Furthermore,

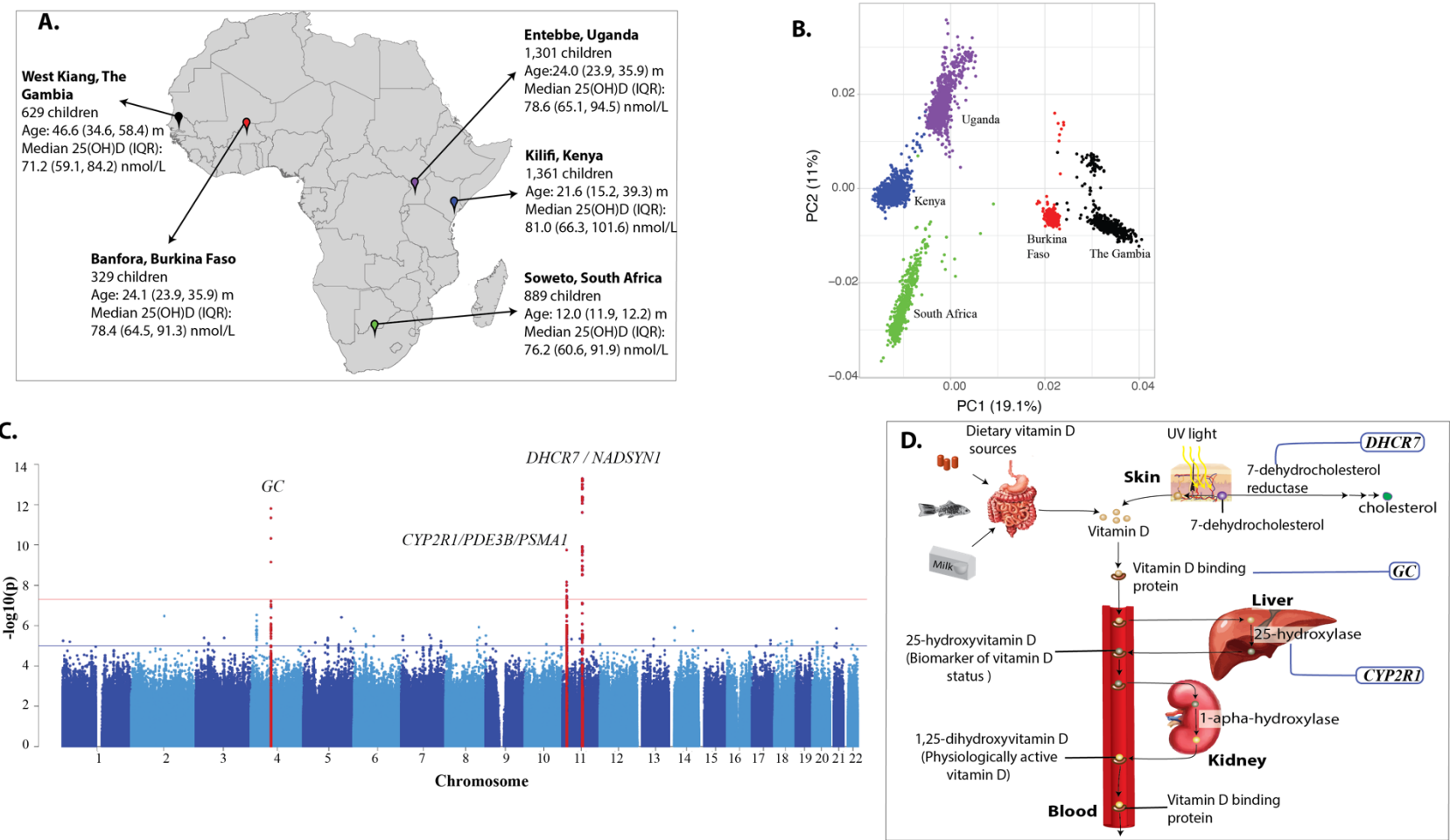
environmental factors, such as dietary differences, sun exposure levels, or lifestyle, could modulate gene–environment interactions in ways that differ across populations.

Our *in vitro* binding assays and molecular docking findings suggest that OSBPL11 interacts with vitamin D compounds, potentially mediating its physiological roles such as intracellular transport of vitamin D metabolites. Unlike other loci in the overall meta-analysis, *OSBPL11* has not previously been reported to be directly linked to vitamin D status or metabolism in the published literature. Studies have shown that lithocholic acid, an oxysterol, can activate vitamin D receptors (VDRs), the same receptors activated by 1,25-dihydroxyvitamin D³⁰, indicating a potential indirect link between vitamin D and oxysterol binding proteins. Furthermore, both vitamin D and oxysterols are implicated in lipid metabolism and demonstrate similar anti-inflammatory and immunomodulatory effects, suggesting the possibility of shared or intersecting biological pathways^{28,31}. OSBPL11, which regulates adipocyte triglyceride storage³² and is strongly associated with obesity and overnutrition, may play a role in vitamin D physiology, as higher rates of obesity have been associated with vitamin D deficiency². Given that OSBPL11 and other members of the OSBP family are implicated in the intracellular transport of lipids¹⁴, this finding suggests that vitamin D compounds might influence lipid distribution and signaling pathways within cells. This study lays the groundwork for further exploration into how alterations in OSBPL11 function may affect vitamin D homeostasis and contribute to clinical outcomes. For instance, dysregulation of vitamin D signaling has been associated with diseases such as osteoporosis and certain cancers^{1,33}, and may explain why obese individuals are on average 35% more likely to be vitamin D deficient than non-obese individuals³⁴. Moreover, the response to vitamin D supplementation in African American participants has been linked to methylation in OSPL6, a protein that, like OSBPL11, belongs to the OSBPL family protein³⁵.

In vivo and *in vitro* studies collectively highlight the possible roles OSBPL11 plays in lipid and glucose metabolism and metabolic health. The high expression of OSBPL11 in adipocytes and its direct involvement in regulating fatty acid metabolism *in vitro* — specifically through the modulation of proteins such as ADIPOQ and FABP4, as well as adipocyte triglyceride storage, suggest that OSBPL11 is crucial for adipocyte function and lipid homeostasis (STRING database). Mouse studies corroborate these findings, showing that OSBPL11 influences glucose homeostasis, triglyceride levels, and total body fat. In humans, elevated expression of OSBPL11 in adipocytes has been associated with conditions related to overnutrition, such as obesity and other diseases of metabolism. Notably, we found that the lead variant at the *OSBPL11* locus, rs2979356, was linked to increased fat mass and triglyceride levels, as well as heightened risks of stroke and hypertension in African ancestry populations although findings varied across populations (Table 2), suggesting that OSBPL11 may contribute to the predisposition to metabolic disorders prevalent in these populations. *OSBPL11* gene polymorphisms have been associated with diastolic blood pressure, hyperglycemia/diabetes, low density lipoproteins-cholesterol (LDL) levels and metabolic syndromes in obese individuals³⁶. Our study extends the understanding of OSBPL11 by revealing its association with 25-hydroxyvitamin D levels and indicating that OSBPL11 may also influence vitamin D metabolism and transport. The shared pathways between lipid and vitamin D metabolism may explain the observed genetic associations and the overlapping physiological roles of vitamin D and oxysterols¹⁴, thereby underscoring the multifaceted role of OSBPL11 in metabolic regulation. Further research is warranted to elucidate the molecular mechanisms underlying OSBPL11's functions and to explore its potential as a therapeutic target for improving metabolic health and reducing disease risk.

The strengths of this study include being the first vitamin D GWAS to be conducted in continental African populations, including community-based children from five countries across Africa. A major highlight of the study is the identification of *OSBPL11*, a novel locus associated with 25(OH)D concentrations, marking the first evidence of its role in vitamin D status. We are also the first study, to the best of our knowledge, to demonstrate that *OSBPL11* binds to vitamin D metabolites in molecular docking or *in vitro* analyses. These findings provide critical insights into the potential role of *OSBPL11* in intracellular vitamin D transport and metabolism. However, our findings should be interpreted in the context of some limitations. First, our GWAS may not have captured low-frequency variants or variants with small effect sizes since vitamin D is a complex trait that is regulated by many genetic and environmental factors. The relatively smaller sample size of our study compared to the much larger European vitamin D GWAS may partially explain our inability to replicate some of their loci. Additionally, the generalisability of our findings to populations with different ethnicities, age groups, dietary habits, sun exposure and other environmental factors remains uncertain. Moreover, many of the individuals in the African diaspora populations included in our replication cohorts were overweight, middle aged and on vitamin D supplements and thus not very representative of African children.

In summary, our study described the genetic determinants of vitamin D concentrations in individuals of African ancestry. We identified a novel association between vitamin D concentrations and the *OSBPL11* gene, marking a significant advance in understanding the mechanisms underlying vitamin D metabolism. This discovery opens new avenues for investigating how *OSBPL11* may influence intracellular vitamin D pathways, particularly its potential role in the transport of vitamin D metabolites within cells. Moreover, the association between the lead variant in *OSBPL11* and cardiometabolic traits, such as triglyceride levels, fat mass, BMI, hypertension, and stroke, suggests a role of *OSBPL11* in lipid metabolism. This finding underscores the possibility that shared pathways between lipid and vitamin D metabolism could contribute to the observed associations. Additionally, our study confirmed previously reported associations between vitamin D concentrations and well-established metabolism genes, namely, *GC*, *CYP2R1*, and *NADSYN1/DHCR7*, further validating their role in vitamin D physiology. Nevertheless, the lack of transferability of some loci—such as the *OSBPL11*—even within some African populations raises important questions about the generalizability of genetic research findings across diverse ancestry groups and emphasizes the need for greater inclusion of underrepresented populations in genomic studies.



428 **Figure 1. Map of Africa showing locations of discovery cohorts included in the current study (A), principal component analysis (PCA) plot (B), Manhattan plot of**
429 **discovery GWAS of 25(OH)D concentrations in African children (C), vitamin D metabolism pathway showing effects of genetic loci (D). Abbreviation PC, principal**
430 **component: IQR, interquartile range. GWAS analyses were adjusted for age, sex, and season of 25(OH)D measurement.**

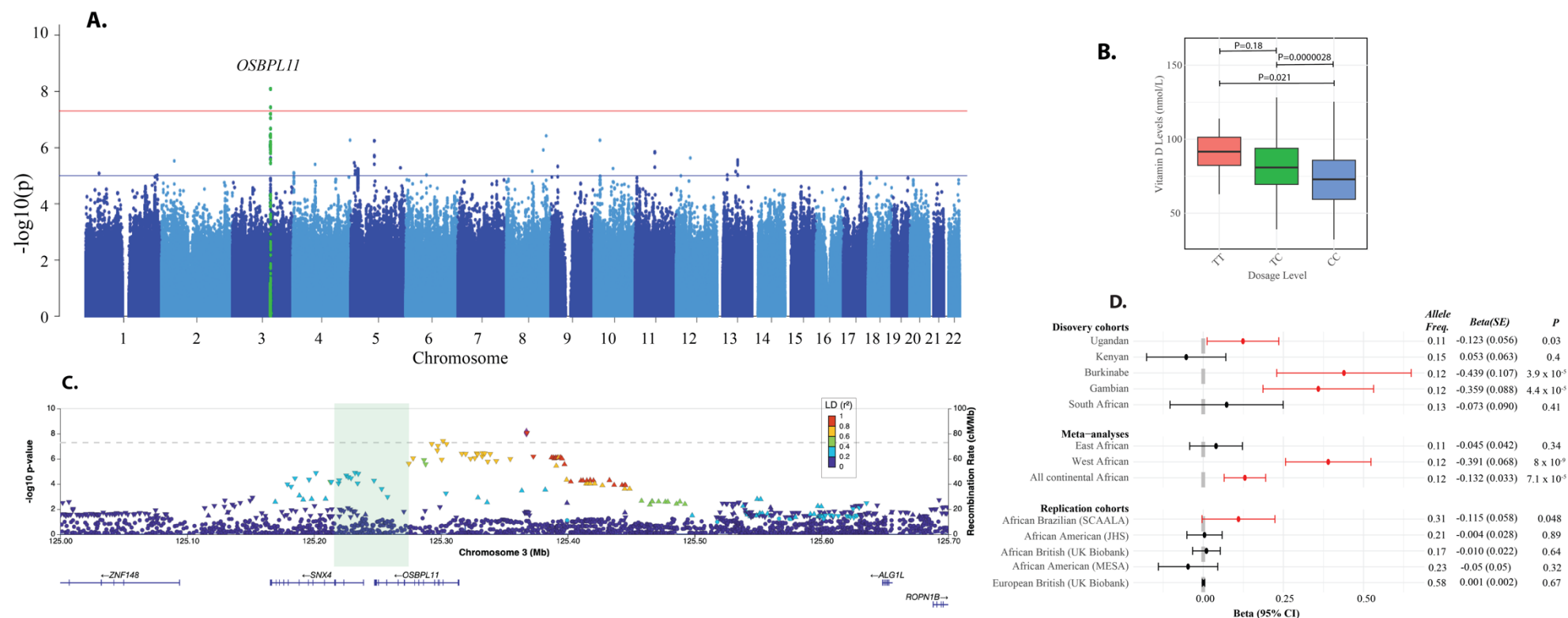


Figure 2. Manhattan plot of meta-analysis GWAS of 25(OH)D concentrations in West African study cohorts (Burkina Faso and The Gambia) (A), boxplot of vitamin D levels by dosages in West African cohorts (B), regional association plot for the *OSBPL11* locus (C), effect of lead variant in the *OSBPL11* locus in discovery and replication cohorts (D), The top variant in the regional plot is shown in a purple diamond shape and the rest are colored according to their LD (R^2) relative to the top variant in individuals of West African ancestry in this study. The direction of the triangle of the top variant in each locus in the regional plot indicates the direction of effect on 25(OH)D concentrations. Plot B boxplots display the median, interquartile range (IQR), and whiskers extending to $1.5 \times \text{IQR}$. Outliers beyond this range were excluded from the plot. Plot D error bars represent 95% confidence intervals (CI), calculated as $\text{Effect} \pm 1.96 \times \text{Standard Error (SE)}$.

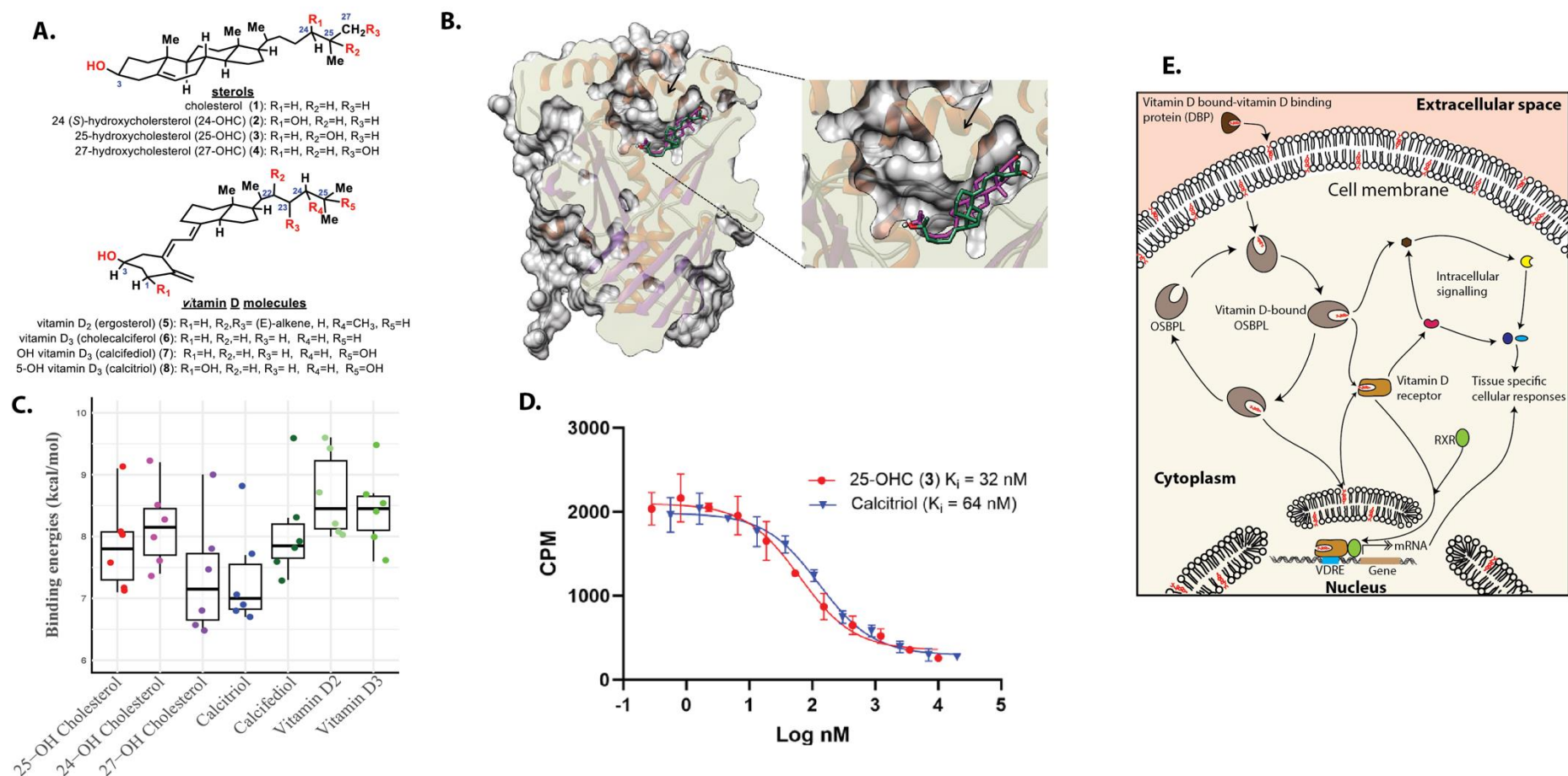


Figure 3. Molecular structures of sterols and vitamin D compounds (A), molecular docking of 25-hydroxycholesterol (magenta) and calcitriol (1,25-OH vitamin D₃) (green) into Alpha Fold modeled OSBPL11 (Q9BXB4 v4) (B), Top six docking scores of common oxysterols and vitamin D metabolites on OSBPL11 Alpha Fold model (C), ³H-25-hydroxycholesterol-competition binding curves of 25-hydroxycholesterol (25-OHC) and calcitriol on OSBPL11 in *in vitro* OSBP/ORP competition binding assay (D), and illustration showing hypothesized mechanism of action of OSBPL11 in vitamin D mediated physiological activities (F). *In vitro* binding was performed on overexpressed OSBPL11-myc-His, in S100 293T lysate, using established 3H-25-OHC *in vitro* OSBP/ORP competition binding assay. Past studies suggest OSBPL11 may be involved in shuttling 1,25-dihydroxyvitamin D between intracellular membranes and ligand receptors (adapted from ¹⁴).

447 **Table 1. Independent genome-wide significant variants for circulating 25(OH)D concentrations in the discovery meta-analysis GWAS**
 448 **for all cohorts (A) and for West Africans (B)**

Gene / Loci	Variant*	Chromosome: position	Effect/ref. allele	Functional effect	Allele Freq.	Effect (Beta)	SE	P value
A. All cohorts*								
<i>GC</i>	rs1352846	4:72617775	A/G	intronic	0.082	-0.287	0.0402	1.56×10^{-12}
<i>CYP2R1/PDE3B/PSMA1</i>	rs12290926	11:14695046	A/G	intronic	0.15	-0.200	0.0313	1.77×10^{-10}
<i>DHCR7/NADSYN1</i>	rs7950991	11:71158612	G/T	intronic	0.17	0.220	0.0292	5.08×10^{-14}
B. West Africa								
<i>OSBPL11</i>	rs2979356	3:125367492	T/C	intergenic	0.12	0.391	0.0678	8.01×10^{-9}

*Overall meta-analysis included cohorts from Kenya, Uganda, Burkina Faso, The Gambia, and South Africa. Additional independent genome-wide significant variants were identified by repeating the associational analyses conditioning for the top variant. No additional signal reached genome-wide significance (*P* value of $< 5 \times 10^{-8}$) for the three loci.

450 **Table 2. Effect of *OSBPL11* lead variant, rs2979356, on select biomarkers of lipid**
 451 **metabolism and cardiovascular health in African ancestry individuals**

	n	Beta	SE	P
Triglycerides				
AADM (overall)	4,292	0.0203	0.0318	0.52
AADM (West Africa)	3,565	-0.00577	0.0365	0.87
UKB (African ancestry)	6,211	-0.0268	0.0245	0.56
AMP-CMDKP	128,439	-0.0155	0.0039	0.0042
AWI-Gen (overall)	10,603	-0.011	0.020	0.58
AWI-Gen (West Africa)	3,763	-0.027	0.034	0.43
GLGC	89,467	-0.0158	0.00598	0.0082
LDL cholesterol				
AADM (overall)	4,257	0.0186	0.0322	0.56
AADM (West Africa)	3,550	0.0223	0.0364	0.54
AWI-Gen (overall)	10,603	-0.044	0.020	0.015
AWI-Gen (West Africa)	3,763	-0.047	0.033	0.16
GLGC	87,759	-0.00619	0.00606	0.31
UKB (African ancestry)	6,006	0.0270	0.0254	0.54
AMP-CMDKP	115,939	0.0055	0.0059	0.35
BMI				
AADM (overall)	5,454	-0.0649	0.155	0.000032
AADM (West Africa)	4,695	-0.0186	0.175	0.000024
AWI-Gen (overall)	10,603	0.012	0.017	0.49
AWI-Gen (West Africa)	3,762	0.024	0.025	0.34
UKB (African ancestry)	6,545	-0.0631	0.0449	0.80
Fat mass				
AADM (overall)	5,295	-0.158	0.302	0.000051
AADM (West Africa)	4,537	-0.0153	0.336	0.00000046
Hypertension				
AWI-Gen (overall)	10,708	0.010	0.009	0.27
UKB (African ancestry)	6,626	-0.0525	0.137	0.15
AMP-CMDKP	31,968	0.0512	0.0247	0.039
Stroke				
UKB (African ancestry)	6,520	-0.00623	0.117	0.019
AMP-CMDKP	4,963	0.482	0.233	0.038

The effect/alternative allele is T/C. *The AWI-Gen Study (Africa Wits-INDEPTH Partnership for Genomic Studies) aimed to investigate the genetic and environmental factors influencing cardiometabolic diseases in sub-Saharan Africa, with participants recruited from Burkina Faso, Ghana, Kenya, Namibia, South Africa, and Tanzania³⁷. AWI-Gen W. Africa included The Gambia and Ghana. The AMP-CMDKP (African American Cardiometabolic Disease Knowledge Portal) aimed to explore genetic contributors to cardiometabolic diseases, recruiting participants from diverse African ancestry populations (<https://hugeamp.org/>). The Global Lipids Genetics Consortium (GLGC) is a large-scale international collaboration focused on identifying genetic variants associated with lipid traits, such as HDL, LDL, triglycerides, and total cholesterol, using data from diverse populations worldwide³⁸. Participants in the AADM study and the UK Biobank study are described in the methods. Other cardiometabolic traits are reported in Supplementary Table 5.

453 **Supplementary Appendix**

454 **Supplementary Table 1. Characteristics of study participants**

	Discovery cohorts					Replication cohorts				
	Kenyans	Ugandans	Burkinabe	Gambians	South Africans	African Brazilian (SCAALA)	African American Adults from JHS	Multi-Ethnic Study of Atherosclerosis (MESA)	African ancestry UK (UKB)	African American (All of US)
No. of participants [§]	962	1288	329	484	607	753	5306	1677	8151	7400
Median 25(OH)D nmol/L (IQR) [*]	78.1 (64.0, 96.5)	78.6 (65.0, 94.6)	78.6 (65.1, 94.5)	72.4 (59.6, 84.9)	72.7 (58.7, 88.8)	27.4 (21.3, 33.1)	32.4 (22.5, 46.2)	17.4 (12.1, 24.2)	31.0 (22.1, 42.9)	26.0 (17.0, 37.0)
Median (IQR) age [†]	21.6 (15.2, 39.3) m	24.0 (23.9, 35.9) m	24.1 (23.9, 35.9) m	46.6 (34.6, 58.4) m	12.0 (11.9, 12.2) m	7.0 (6.0, 9.0) y	55.0 (45.0, 64.0) y	62.0 (53.0, 70.0)	50.0 (45.0, 58.0) y	61.0 (50.0, 69.0) y
Sex: females	473/962 (49.2%)	638/1288 (49.5%)	161/329 (48.9%)	218/484 (45.0%)	291/607 (47.9%)	361/750 (48.1%)	3367/5306 (63.1%)	906 (54.0%)	3393/8151 (41.6%)	6458/8935 (72.28%)
Season [#]										
1 st season	48/962 (5.0%)	302/1284 (23.4%)	72/329 (21.9%)	-	238/607 (39.2%)	179 (24.2%)	1137/5306 (21.4%)	409/1677(24.4%)	1719/8151 (21.1%)	2255/8935 (25.2%)
2 nd season	357/962 (37.1%)	310/1284 (24.1%)	123/329 (37.4%)	-	25/607 (4.1%)	-	1189/5306 (22.4%)	565/1677(33.6%)	2197/8151 (27.0%)	1889/8935 (21.1%)
3 rd season	528/962 (54.9%)	318/1284 (24.6%)	129/329 (39.2%)	536/484 (85.4%)	228/607 (37.6%)	-	1447/5306 (27.3%)	339/1677(20.2%)	1565/8151 (19.2%)	2220/8935 (24.9%)
4 th season	29/962 (3.0%)	354/1284 (27.5%)	5/329 (1.5%)	92/484 (14.7%)	116/607 (19.1%)	562 (75.8%)	1533/5306 (28.9%)	364/1677(21.7%)	2670/8151 (32.7%)	2571/8935 (28.8%)
BMI (IQR)	14.9 (13.7, 16.)	15.7 (14.7, 16.4)	15.6 (14.7, 16.4)	14.6 (13.8, 15.4)	n/a	15.6 (14.3, 16.3)	30.5 (26.9, 35.4)	29.4 (26.1, 33.5)	28.7 (25.7, 32.3)	32.1 (27.1, 38.4)
Malaria [‡]	189/857 (22.1%)	88/1268 (6.9%)	64/303 (21.1%)	65/484 (10.4%)	[¶] No malaria	[¶] No malaria	[¶] No malaria	[¶] No malaria	[¶] No malaria	[¶] No malaria
Inflammation [‡]	256/962 (27.0%)	304/1273 (23.9%)	109/329 (33.9%)	85/484 (14%)	94/607 (15.5%)	230/751(30.5%)	1242/4791(25.9%) ^{§§}	445/1677 (26.6%)	1088/8151 (13%)	n/a
Supplementation ^{††}	n/a ^{††}	n/a ^{††}	n/a ^{††}	n/a ^{††}	n/a ^{††}	n/a ^{††}	1648/5306 (34.4%)	n/a	490/7611 (6.4%)	5640/8935 (63.1%)
Abbreviations; IQR, interquartile range; n/a, not available; 25(OH)D, 25-hydroxyvitamin D. [§] The number of participants in each study cohorts included in the GWAS, based on availability of genotype and phenotype data. [¶] Malaria was defined as the presence of <i>P. falciparum</i> parasites on blood film. [‡] Inflammation was defined as CRP > 5 mg/L or ACT > 0.6 g/L. ACT, but not CRP, was available for The Gambia. [†] Age is presented in months for discovery cohorts and years for replication cohorts. [#] 1 st season corresponds to summer/short rains/dry season; 2 nd season as autumn/dry season; 3 rd season as winter/long rains season; 4 th season as spring or dry season. Seasons in discovery cohorts were based on 3 monthly intervals as follows 1 st season, December to February; 2 nd season, March to May; 3 rd season, June to August; 4 th season, September to November. In South Africa these seasons correspond to summer, autumn, winter, and spring, respectively, in Uganda and Kenya there are two rainy and two dry seasons and in Burkina Faso and The Gambia there is a single rainy and dry season. For MESA study, seasons were defined as follows;1 st season from January to March; 2 nd , April to June; 3 rd , July to September; 4 th , October to December. However, the timing of the rains is often unpredictable and may vary from these times. ^{§§} HsCRP was measured in African American cohorts, which was converted to values equivalent to CRP as previously (Arthroplasty 2014). ^{††} Vitamin D supplementation is very low/non-existent in the African children.										

455

456

457 **Supplementary Table 2. Fine-mapping posterior probabilities of k causal variants at independent signal loci that reached genome-wide**
 458 **significance in the discovery GWAS**

Locus	GWAS index variant	Chromosome: position	K (no. of causal variants) *	Posterior probability
<i>GC</i>	rs1352846	4:72617775	26	1
			27	2.96×10^{-10}
<i>CYP2R1/PDE3B/PSMA1</i>	rs12290926	11:14695046	24	0.54
			25	0.30
			21	0.056
			22	0.040
			23	0.046
			26	0.021
<i>DHCR7/NADSYN1</i>	rs7950991	11:71158612	11	1
			10	2.96×10^{-106}
<i>OSBPL11</i>	rs2979356	3:125367492	8	1

*K (causal variants) 0 to 100 were analyzed. However, only ks with non-zero posterior probabilities are presented in this table. The most probable no of causal variants (k) is highlighted in bold. The k variants listed are predicted as possible credible sets and are not definitively causal.

Supplementary Table 3. Effect of independent African genome-wide significant variants in continental and diaspora African ancestry cohorts and Europeans in the UK Biobank

GENE	GC	CYP2R1/PDE3B/PSMA1	DHCR7/NADSYN1
VARIANT DETAILS			
Lead variant	rs1352846	rs12290926	rs7950991
Chromosome: position	4:72617775	11:14695046	11:71158612
Effect/ref allele	A/G	A/G	G/T
Functional effect	intronic	intronic	intronic
CONTINENTAL AFRICAN ANCESTRY COHORTS			
West Africans meta-analysis			
Allele Freq.	0.039	0.14	0.15
Effect (Beta)	-0.19	-0.25	0.19
SE	0.11	0.063	0.059
P value	0.084	5.45×10^{-05}	0.0012
East Africans meta-analysis			
Allele Freq.	0.10	0.13	0.16
Effect (Beta)	-0.32	-0.17	0.25
SE	0.047	0.041	0.038
P value	3.04×10^{-11}	3.41×10^{-05}	5.06×10^{-11}
South Africans			
Allele Freq.	0.066	0.19	0.200
Effect (Beta)	-0.23	-0.21	0.16
SE	0.11	0.073	0.072
P value	0.040	0.0037	0.029
DIASPORA AFRICAN ANCESTRY COHORTS			
SCAALA (Brazil)			
Allele Freq.	0.163	0.30	0.10
Effect (Beta)	-0.075	-0.12	0.17
SE	0.069	0.056	0.087
P value	0.28	0.039	0.050
JHS (USA)			
Allele Freq.	0.10	0.15	0.17
Effect (Beta)	-0.12	-0.02	0.10
SE	0.038	0.031	0.030
P value	0.0016	0.56	3.97×10^{-4}
MESA (USA)			
Allele Freq.	0.116	0.172	0.155
Effect (Beta)	-0.218	0.0453	0.0188
SE	0.0610	0.0523	0.0530
P value	0.00036	0.39	0.72
All of Us (USA)			
Allele Freq.	0.10	0.16	0.17
Effect (Beta)	-0.124	-0.0593	0.0717
SE	0.026	0.0218	0.0208
P value	2.18×10^{-06}	0.0065	0.00056
UK Biobank (UK)			
Allele Freq.	0.09	0.15	0.16
Effect (Beta)	-0.19	-0.02	0.07
SE	0.03	0.02	0.02
P value	6.66×10^{-12}	0.48	6.40×10^{-4}
EUROPEAN ANCESTRY COHORT			
UK Biobank (UK)			
Allele Freq.	0.291	0.41	n/a*
Effect (Beta)	-0.187	-0.068	n/a*
SE	0.002	0.002	n/a*
P value	2.90×10^{-1689}	3.53×10^{-266}	n/a*

Abbreviation: ref. reference; SE standard error; freq. frequency; n/a not available. UKB European summary data was obtained from Minousaki et al³. The discovery association analyses were adjusted for age, sex and season, and incorporated GRM to account for relatedness. The rs7950991 variant was missing in European GWAS, it is rare in Europeans (MAF<0.001, dbSNP). However, rs12284909, which is in high linkage disequilibrium with rs7950991 in our discovery cohorts ($r^2=0.98$), was genome-wide significant in Europeans (effect/other allele; A/G, A allele frequency=0.006, beta = -0.09, $P=7.6 \times 10^{-12}$).

463 **Supplementary Table 4. Replication of genome-wide significant variants in UK Biobank European-ancestry GWAS on 25(OH)D**
464 **concentrations in a meta-analysis of continental African cohorts**

Gene	Variant	UKB European population GWAS						Continental African populations GWAS			
		Chromosome: position	Effect/ref allele	Allele Freq.	Effect (Beta)	SE	P value	Allele Freq.	Effect (Beta)	SE	P value
<i>CYP2R1</i>	rs201501563	11:14882470	T/C	0.065	-0.035	0.004	1.96×10^{-18}	0.084	-0.215	0.040	6.01×10^{-08}
<i>CYP2R1</i>	rs10832289	11:14669496	T/A	0.41	-0.086	0.002	2.84×10^{-266}	0.088	-0.204	0.039	1.31×10^{-07}
<i>GC</i>	rs11723621	4:72615362	G/A	0.29	-0.156	0.003	2.90×10^{-1689}	0.069	-0.182	0.043	2.61×10^{-05}
<i>SEC23A</i>	rs8018720	14:39556185	C/G	0.82	-0.032	0.003	4.10×10^{-36}	0.86	-0.091	0.031	0.003
<i>RP13-379L11.3</i>	rs2585442	20:52737123	G/C	0.24	0.023	0.002	3.96×10^{-23}	0.069	0.114	0.042	0.007
<i>COG5</i>	rs1858889	7:107117447	C/A	0.50	0.013	0.002	3.03×10^{-11}	0.21	0.068	0.026	0.010
<i>HAL</i>	rs10859995	12:96375682	C/T	0.59	-0.041	0.002	3.03×10^{-91}	0.34	-0.045	0.023	0.050
<i>STAP2</i>	rs57631352	19:4338173	G/A	0.30	-0.013	0.002	1.50×10^{-09}	0.31	0.040	0.023	0.089
<i>MAT1A</i>	rs10887718	10:82042624	T/C	0.53	-0.013	0.002	1.18×10^{-10}	0.43	0.036	0.022	0.100
<i>GC</i>	rs222026	4:72643760	T/A	0.88	-0.051	0.004	1.09×10^{-40}	0.34	-0.036	0.023	0.117
<i>LDLR</i>	rs73015021	19:11192915	G/A	0.12	0.022	0.003	6.29×10^{-14}	0.37	0.034	0.023	0.138
<i>RP11-120M18.2</i>	rs2909218	17:66464546	T/C	0.80	0.017	0.002	2.82×10^{-12}	0.75	-0.037	0.025	0.138
<i>APOC1</i>	rs1065853	19:45413233	T/G	0.081	0.028	0.004	2.24×10^{-14}	0.13	0.047	0.032	0.140
<i>RP13-379L11.3</i>	rs6123359	20:52714706	G/A	0.10	0.024	0.003	7.48×10^{-14}	0.090	0.056	0.039	0.145
<i>TMEM151A</i>	rs523583	11:66070146	C/A	0.47	0.013	0.002	6.60×10^{-12}	0.12	-0.047	0.033	0.159
<i>GCKR</i>	rs11127048	2:27752463	A/G	0.62	0.018	0.002	6.72×10^{-19}	0.76	0.036	0.025	0.159
<i>RP11-21L23.4</i>	rs1149605	11:76485216	C/T	0.17	0.020	0.003	3.36×10^{-15}	0.11	0.048	0.034	0.161
<i>RP4-657M3.2</i>	rs7519574	1:34726552	A/G	0.18	0.017	0.003	4.03×10^{-11}	0.13	-0.045	0.033	0.170
<i>ABO</i>	rs532436	9:136149830	A/G	0.19	-0.015	0.003	1.94×10^{-09}	0.11	-0.045	0.035	0.195
<i>DSG1</i>	rs8091117	18:28919794	A/C	0.064	-0.024	0.004	9.48×10^{-10}	0.43	0.030	0.023	0.202
<i>KLK10</i>	rs10426	19:51517798	A/G	0.21	0.025	0.002	1.59×10^{-26}	0.080	0.050	0.040	0.220
<i>ZPR1</i>	rs964184	11:116648917	C/G	0.87	0.040	0.003	1.30×10^{-43}	0.79	0.031	0.027	0.247
<i>HSD17B11</i>	rs58073039	4:88287363	G/A	0.30	-0.013	0.002	2.84×10^{-10}	0.57	-0.022	0.022	0.319
<i>SCUBE1</i>	rs960596	22:41393520	T/C	0.34	0.012	0.002	2.43×10^{-09}	0.022	-0.074	0.077	0.336
<i>PADI1</i>	rs3750296	1:17559656	C/G	0.34	-0.021	0.002	3.04×10^{-24}	0.44	-0.020	0.022	0.350
<i>EBF2</i>	rs34726834	8:25889606	T/C	0.25	0.014	0.002	3.39×10^{-10}	0.36	-0.021	0.023	0.354
<i>SERPINB11</i>	rs2037511	18:61366207	A/G	0.17	0.016	0.003	8.35×10^{-10}	0.078	-0.037	0.041	0.361
<i>CELSR2</i>	rs7528419	1:109817192	G/A	0.22	0.019	0.002	2.43×10^{-16}	0.30	0.020	0.024	0.397
<i>TM6SF2</i>	rs58542926	19:19379549	T/C	0.077	0.033	0.004	2.63×10^{-19}	0.017	0.072	0.088	0.411
<i>FLJ42102</i>	rs12803256	11:71132868	G/A	0.78	0.087	0.003	1.64×10^{-195}	0.17	0.023	0.030	0.436
<i>TNFAIP8</i>	rs7718395	5:118652574	G/C	0.32	0.013	0.002	1.68×10^{-09}	0.06	-0.036	0.048	0.454
<i>ZPR1</i>	rs2847500	11:120114421	A/G	0.12	-0.021	0.003	1.93×10^{-12}	0.32	-0.017	0.024	0.467
<i>TDRD15</i>	rs12997242	2:21381177	A/G	0.44	-0.012	0.002	2.32×10^{-10}	0.19	0.020	0.028	0.468

<i>BCAR4</i>	rs8063706	16:11909552	T/A	0.27	0.013	0.002	4.27×10^{-09}	0.50	0.016	0.022	0.472
<i>ARNT</i>	rs3768013	1:150815411	A/G	0.37	-0.012	0.002	3.86×10^{-09}	0.53	-0.014	0.021	0.502
<i>MARC_1</i>	rs867772	1:220972343	G/A	0.68	-0.014	0.002	3.31×10^{-11}	0.94	0.029	0.045	0.523
<i>CETP</i>	rs1800775	16:56995236	A/C	0.49	-0.017	0.002	1.57×10^{-17}	0.62	0.012	0.022	0.582
<i>GC</i>	rs705117	4:72608115	T/C	0.85	0.031	0.003	1.12×10^{-27}	0.19	0.015	0.027	0.590
<i>ZNF808</i>	rs8103262	19:53065814	C/T	0.30	0.013	0.002	6.80×10^{-10}	0.42	0.012	0.022	0.602
<i>DOK7</i>	rs78649910	4:3482213	A/T	0.11	-0.018	0.003	3.41×10^{-09}	0.10	-0.017	0.037	0.639
<i>APOC1</i>	rs157595	19:45425460	G/A	0.62	-0.016	0.002	4.25×10^{-15}	0.86	0.015	0.033	0.648
<i>NPAS2</i>	rs6724965	2:101440151	G/A	0.17	-0.017	0.003	1.34×10^{-10}	0.71	0.010	0.023	0.662
<i>HTR5BP</i>	rs7569755	2:118648261	A/G	0.29	0.014	0.002	8.35×10^{-11}	0.41	-0.010	0.023	0.664
<i>MED23</i>	rs3822868	6:131934986	G/A	0.84	0.022	0.003	1.41×10^{-15}	0.023	0.033	0.076	0.667
<i>DNAH11</i>	rs111529171	7:21571932	C/G	0.22	-0.015	0.002	6.26×10^{-11}	0.22	-0.011	0.027	0.682
<i>MRPL3</i>	rs6438900	3:125148287	G/C	0.26	0.014	0.002	1.16×10^{-09}	0.50	0.007	0.022	0.738
<i>LINC00536</i>	rs7828742	8:116960729	G/A	0.60	-0.024	0.002	2.85×10^{-33}	0.71	-0.007	0.024	0.767
<i>PDILT</i>	rs77924615	16:20392332	A/G	0.19	-0.016	0.002	2.28×10^{-10}	0.049	-0.014	0.051	0.782
<i>CPS1</i>	rs1047891	2:211540507	A/C	0.32	-0.014	0.002	1.16×10^{-11}	0.36	-0.006	0.024	0.799
<i>NPHS1</i>	rs3814995	19:36342212	T/C	0.31	-0.015	0.002	1.08×10^{-12}	0.037	-0.015	0.062	0.807
<i>RHOA</i>	rs7650253	3:49431160	A/T	0.70	0.015	0.002	1.76×10^{-10}	0.75	-0.006	0.026	0.827
<i>UGT2B7</i>	rs7699711	4:69947596	T/G	0.45	-0.030	0.002	4.85×10^{-50}	0.75	-0.005	0.026	0.834
<i>CADM2</i>	rs1972994	3:85631142	T/A	0.65	-0.018	0.002	8.04×10^{-18}	0.78	0.006	0.027	0.836
<i>FOXO6</i>	rs56044892	1:41830086	T/C	0.20	0.015	0.002	3.13×10^{-10}	0.074	0.009	0.044	0.839
<i>LIPC</i>	rs1800588	15:58723675	T/C	0.21	-0.030	0.002	3.17×10^{-37}	0.53	-0.004	0.021	0.842
<i>RP13-379L11.3</i>	rs6127099	20:52731402	T/A	0.27	-0.027	0.002	2.22×10^{-32}	0.18	0.006	0.030	0.849
<i>FDPS</i>	rs11264360	1:155284586	A/T	0.24	0.018	0.002	1.12×10^{-15}	0.22	0.003	0.026	0.908
<i>SLCO1B1</i>	rs12317268	12:21352541	G/A	0.15	-0.019	0.003	9.20×10^{-12}	0.16	-0.003	0.030	0.920
<i>TFDP2</i>	rs6773343	3:141825598	T/C	0.72	0.013	0.002	6.28×10^{-09}	0.71	-0.002	0.023	0.928
<i>AC007950.2</i>	rs17765311	15:63789952	C/A	0.34	-0.015	0.002	1.18×10^{-13}	0.022	0.007	0.080	0.933
<i>LIPC</i>	rs261291	15:58680178	C/T	0.36	-0.023	0.002	2.46×10^{-29}	0.43	-0.002	0.022	0.934
<i>DNAH11</i>	rs10818769	9:125719923	G/C	0.86	-0.017	0.003	2.99×10^{-09}	0.053	-0.003	0.048	0.951
<i>RER1</i>	rs6698680	1:2329661	G/A	0.47	-0.012	0.002	7.47×10^{-10}	0.30	-0.001	0.024	0.971
<i>PEAK1</i>	rs62007299	15:77711719	A/G	0.71	-0.014	0.002	3.33×10^{-11}	0.52	-0.001	0.022	0.974
<i>GATA4</i>	rs804280	8:11612698	A/C	0.58	0.016	0.002	9.90×10^{-16}	0.63	-2.5×10^{-4}	0.022	0.991
<i>LINC01004</i>	rs1011468	7:104613791	A/G	0.47	-0.014	0.002	1.39×10^{-12}	0.85	-5.5×10^{-5}	0.029	0.998

European summary data were obtained from Minousaki et al³. Variants that were missing in the current GWAS or had a P value $> 5 \times 10^{-8}$ in the European GWAS were excluded. The variants have been sorted by P value in a meta-analysis of GWAS of continental African populations. Replication was defined as the same direction of effect and P value < 0.05 .

466 **Supplementary Table 5. Effect of *OSBPL11* lead variant, rs2979356, on select biomarkers of lipid metabolism and cardiovascular health**
467 **in African ancestry individuals**

	n	Beta	SE	P
HDL cholesterol				
AADM (overall)	4,295	0.0338	0.0319	0.29
AADM (West Africa)	3,561	-0.0272	0.0368	0.46
AWI-Gen (overall)	10,603	0.004	0.020	0.83
AWI-Gen (West Africa)	3,763	-0.027	0.034	0.43
UKB (African ancestry)	5,754	-0.00359	0.0240	0.055
AMP-CMDKP	128,439	-0.0039	0.0039	0.47
GLGC	90,804	0.00209	0.00595	0.73
Total cholesterol				
AADM (overall)	4,284	0.0154	0.0319	0.63
AADM (West Africa)	3,553	0.0188	0.0366	0.61
AWI-Gen (overall)	10,603	-0.023	0.020	0.20
AWI-Gen (West Africa)	3,763	-0.031	0.033	0.36
UKB (African ancestry)	6,212	0.0174	0.0248	0.32
AMP-CMDKP	128,439	0.0086	0.0039	0.11
GLGC	92,554	-0.00928	0.00589	0.12
HOMA2-%S (AADM)	2,390	-0.0140	0.047	0.77
HOMA2-B (AADM)	2,390	-0.0181	0.036	0.61
TG/HDL ratio (AWI-Gen West Africa)	23,139	0.0277	0.0227	0.23
Fasting Glucose (AWI-Gen West Africa)	3,591	0.024	0.022	0.57
Fasting Insulin (AWI-Gen West Africa)	2,129	0.007	0.021	0.14

The effect/alternative allele is T/C. HOMA2-B estimates beta cell function and insulin production in the body using fasting insulin and glucose levels. HOMA2-%S is a calculation that measures how well peripheral tissues respond to insulin and absorb glucose from the bloodstream.*The AWI-Gen Study (Africa Wits-INDEPTH Partnership for Genomic Studies) aimed to investigate the genetic and environmental factors influencing cardiometabolic diseases in sub-Saharan Africa, with participants recruited from Burkina Faso, Ghana, Kenya, Namibia, South Africa, and Tanzania³⁷. AWI-Gen W. Africa included The Gambia and Ghana. The AMP-CMDKP (African American Cardiometabolic Disease Knowledge Portal) aimed to explore genetic contributors to cardiometabolic diseases, recruiting participants from diverse African ancestry populations (<https://hugeamp.org/>). Participants in the AADM study and the UK Biobank study are described in the methods.

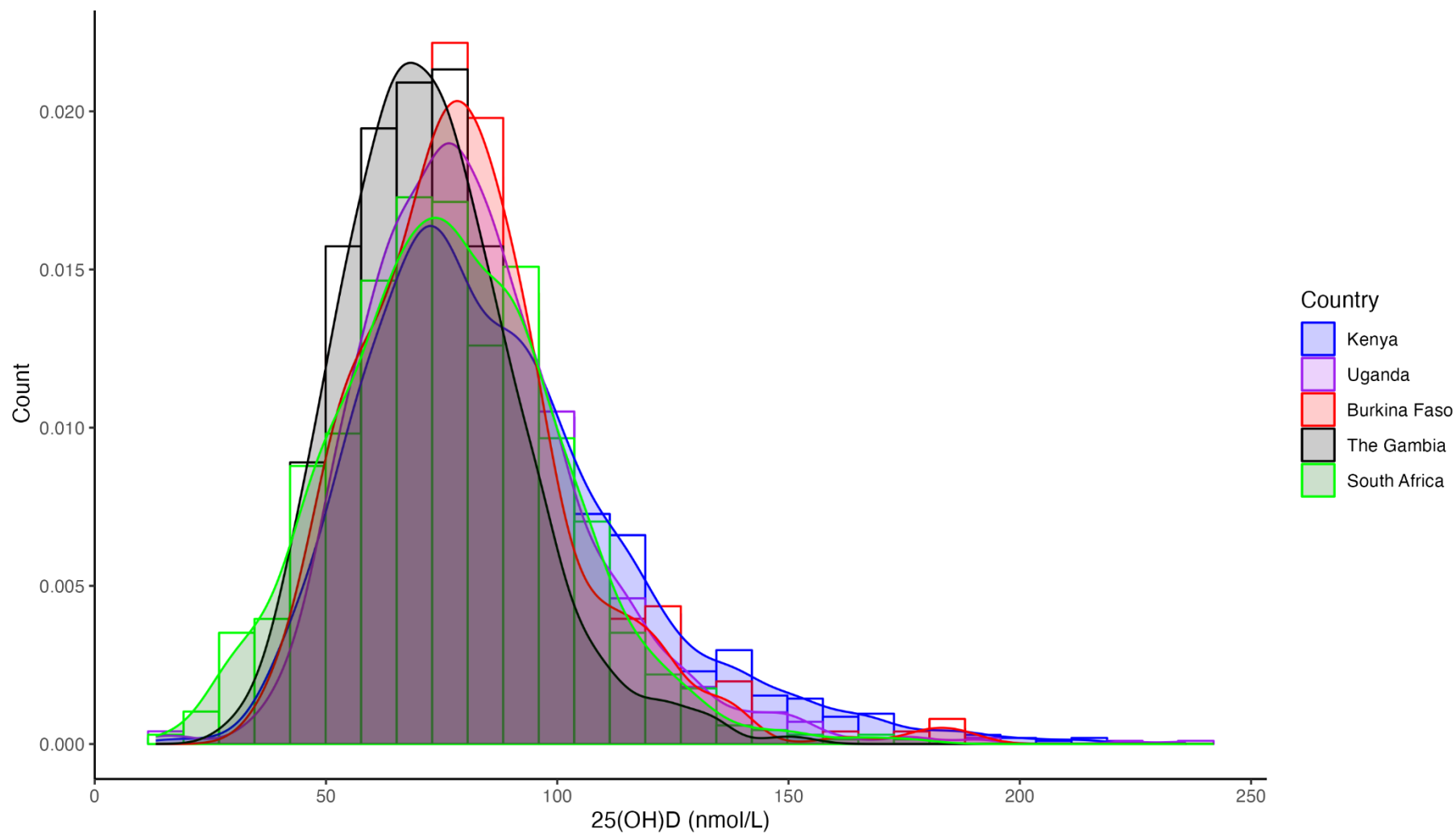
468
469

470 **Supplementary Table 6. Study cohort site, design, site, genotyping and phenotyping**

Study name	Continental African discovery cohorts					African ancestry replication cohorts				
	Kilifi immunology cohort (Kenya)	The Entebbe Mother and Baby Study (Uganda)	The VAC050 ME-TRAP malaria vaccine trial (Burkina Faso)	Malaria study (The Gambia)	Soweto Vaccine Response Study (South Africa)	Jackson Heart Study (USA)	Multi-Ethnic Study of Atherosclerosis (USA)	UK Biobank African ancestry participants (UK)	SCAALA (Brazil)	All of US Research Program (USA)
Study Site	Kilifi, Kenya	Entebbe, Uganda	Banfora, Burkina Faso	West Kiang, The Gambia	Soweto, South Africa	Jackson, Mississippi, USA	USA	UK	Salvador, Brazil	USA
Study design	Community-based cohort	Community-based cohort	Community-based cohort	Community-based cohort	Community-based cohort	Community-based cohort	Community-based cohort	Population cohort	Community-based cohort	Population cohort
Genotyping platform	H3Africa chip v1	Illumina HumanOmni 2.5M-8	Illumina HumanOmni 2.5M-8	H3Africa chip v1	Illumina HumanOmni 2.5M-8	Affymetrix 6.0	Affymetrix 6.0	UK Biobank Axiom™ Array	Illumina HumanOmni 2.5M	Illumina NovaSeq 6000 whole genome sequencing
Number of variants in chip	2.5 million	2.5 million	2.5 million	2.5 million	2.5 million	906,600	906,600	850,000	2.5 million	Whole genome sequence
Vitamin D assay	CMIA (Abbott Architect, USA)	CMIA (Abbott Architect, USA)	CMIA (Abbott Architect, USA)	CMIA (Abbott Architect, USA)	CMIA (Abbott Architect, USA)	LC-MS-MS	LC-MS-MS	DiaSorin Liaison XL assay	Inhibitory enzyme immunosorbent assay (Bolton, UK)	Multiple assays
CMIA, chemiluminescent microparticle immunoassay; SCAALA, Social Change Asthma and Allergy; UKB, UK Biobank; LC-MS-MS, Liquid chromatography with tandem mass spectrometry.										

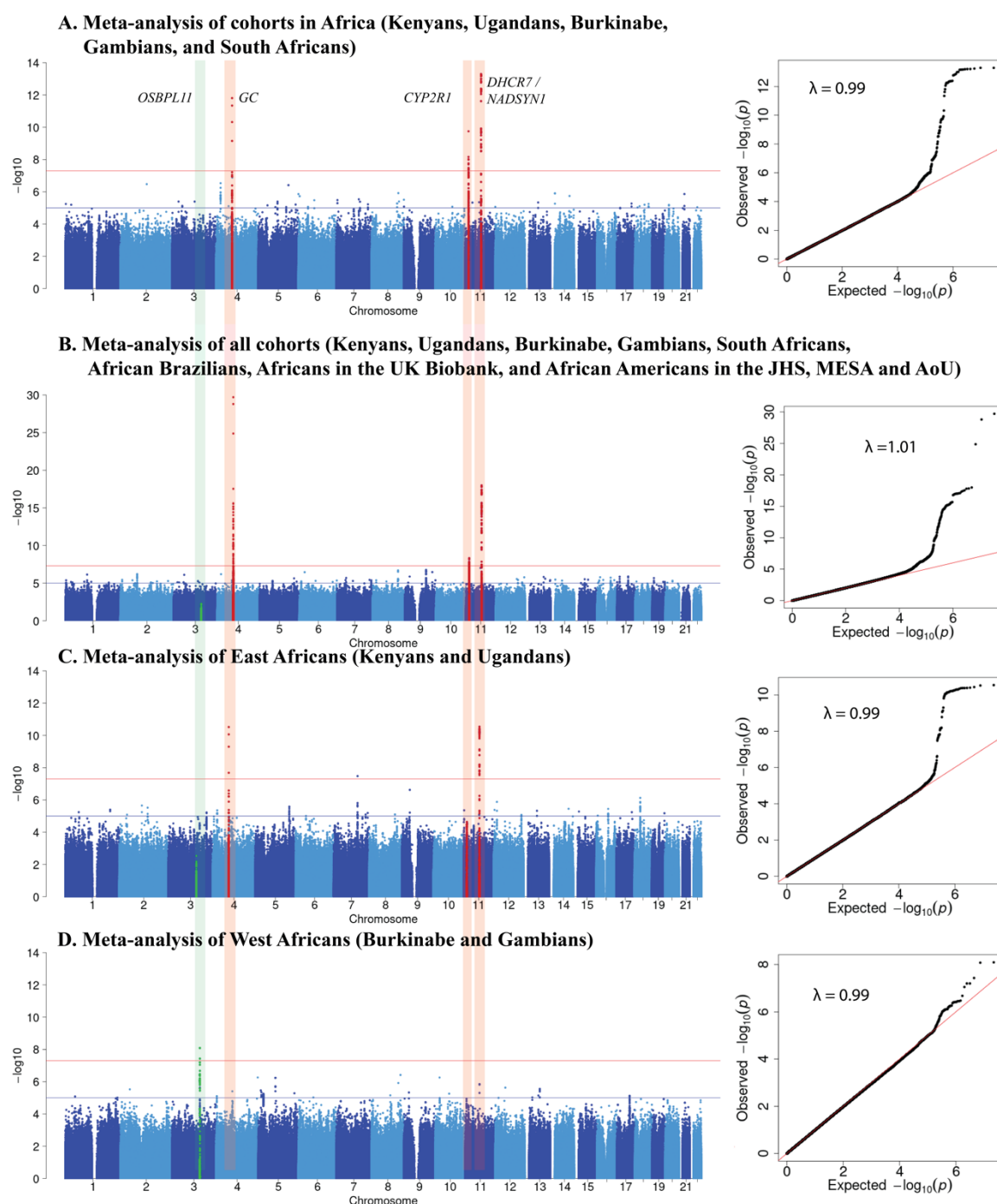
471

472

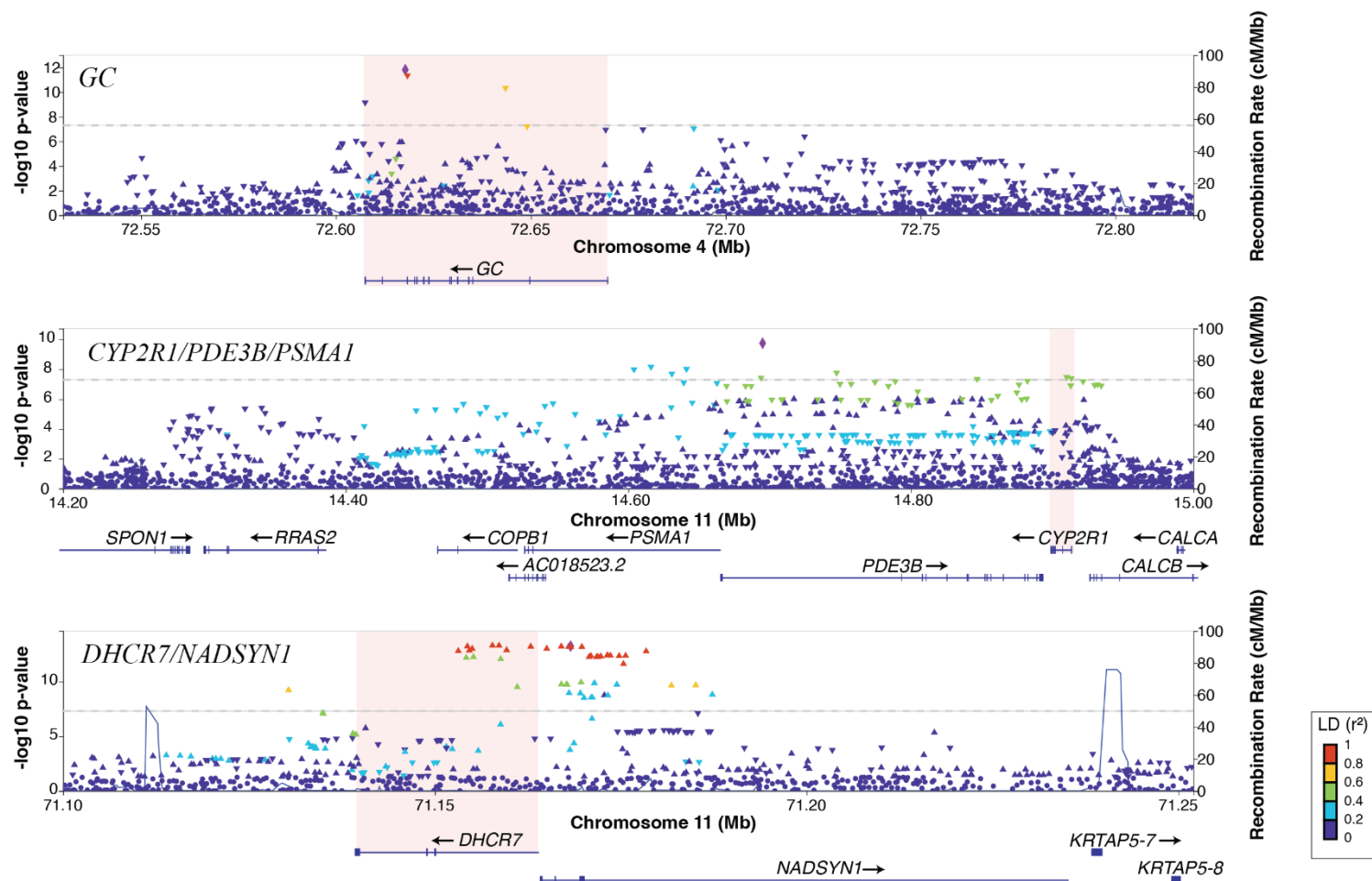


473

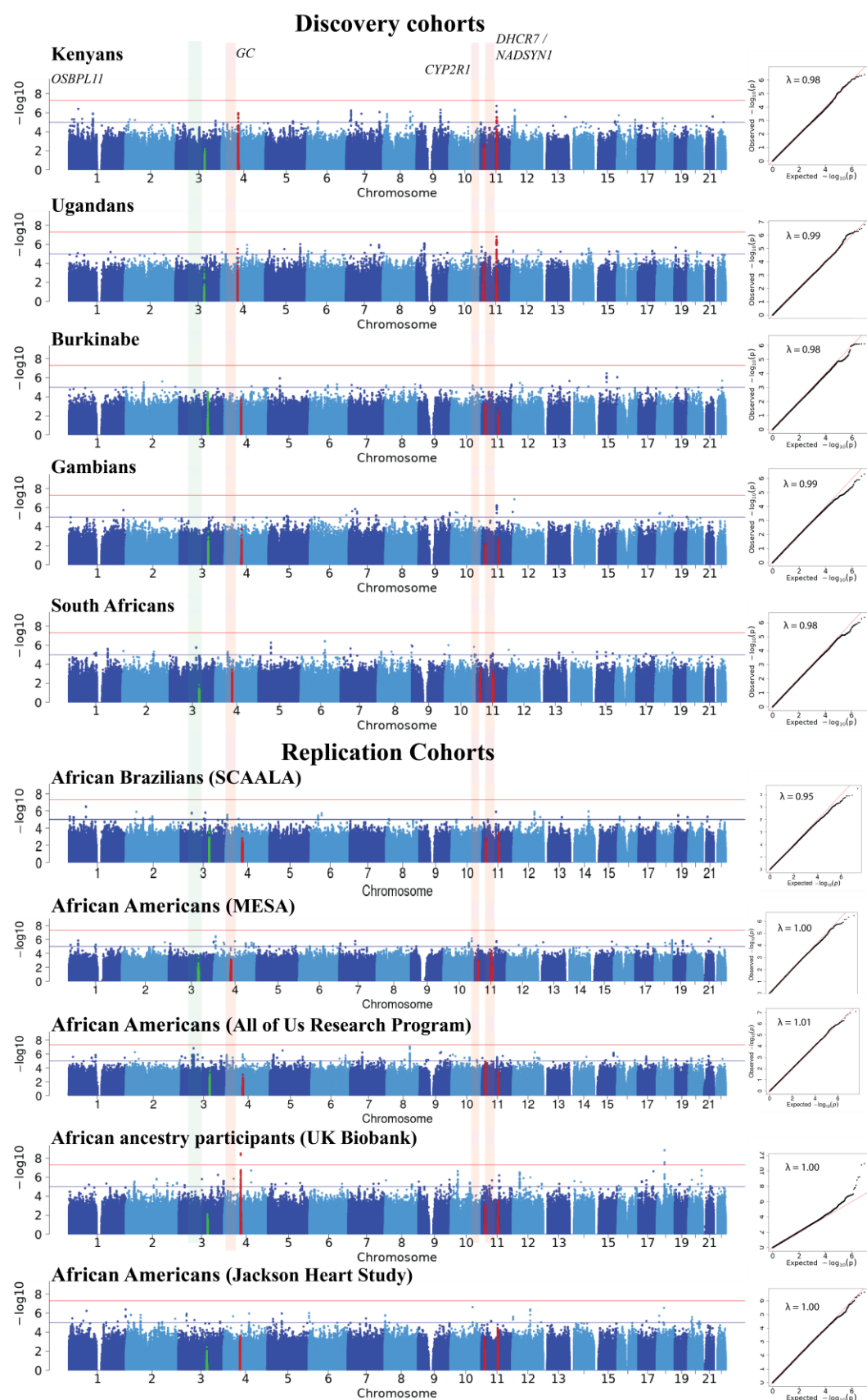
474 **Supplementary Figure 1. Histogram with density plots showing the distribution of 25(OH)D concentrations.** Shapiro-Wilk's test for
 475 normality indicated that 25(OH)D concentrations were skewed ($W=0.95$, P value < 0.001).



Supplementary Figure 2. Manhattan and QQ plots of GWAS of 25(OH)D concentrations meta-analysis of all cohorts in Africa (A), all discovery and replication cohorts (B), East African cohorts (C), and West African cohorts (D). Variants in known loci (*GC*, *CYP2R1/PDE3B* and *DHCR7/NADSYN1*) are indicated in red and novel loci (*OSBPL11*) in green.

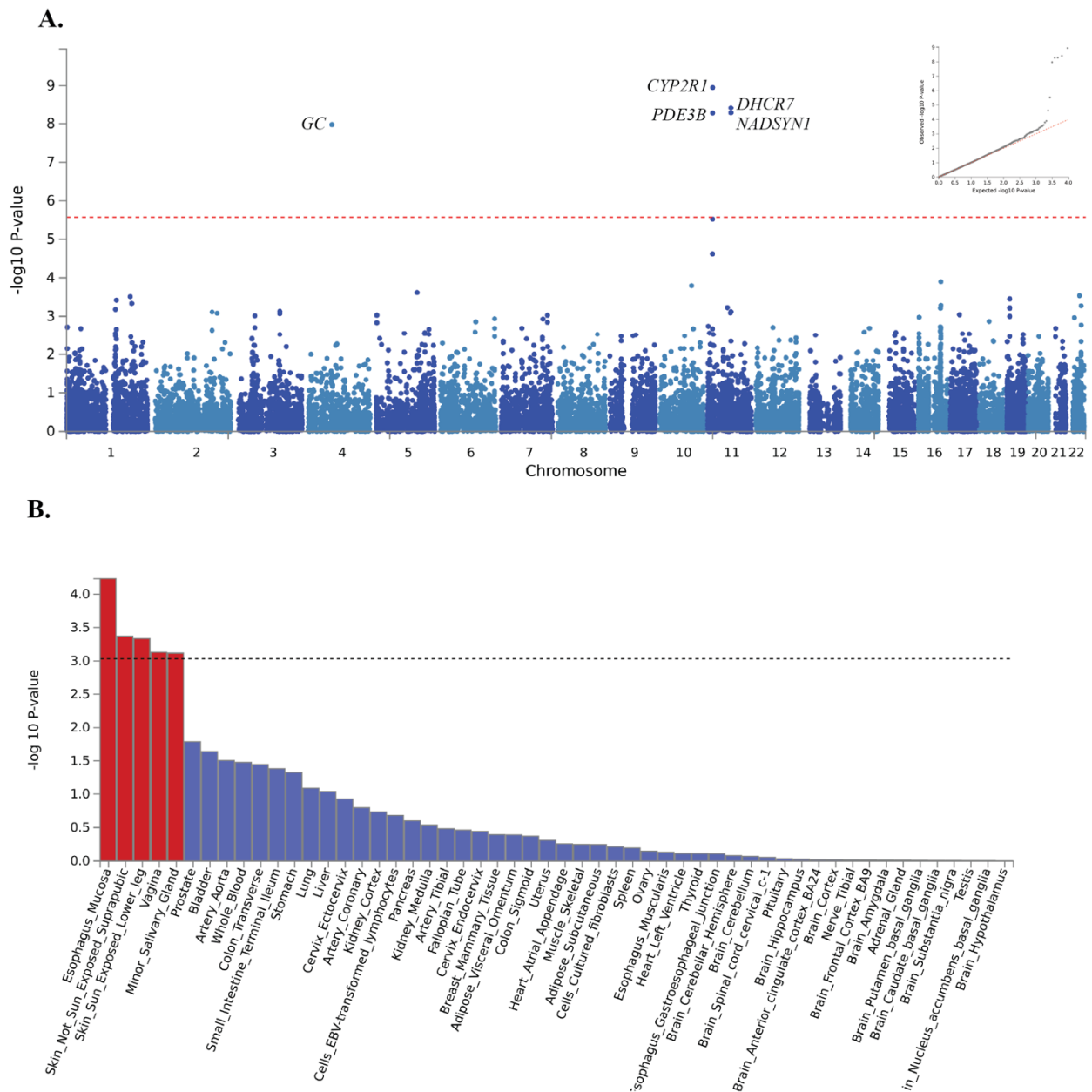


Supplementary Figure 3. Regional association plots for the *GC*, *CYP2R1/PDE3B/PSMA1* and *DHCR7/NADSYN1* loci. The top variant in the regional plots is shown with a purple diamond shape and the rest are coloured according to their LD (R^2) relative to the top variant in individuals of African ancestry in this study; colour varies between absent LD (blue) and total LD (red). The direction of the triangle of the top variant in each locus in the regional plots indicates the direction of effect on 25(OH)D concentrations.



Supplementary Figure 4. Manhattan and QQ plots of GWAS of 25(OH)D concentrations by cohort. Variants in and close to known loci (*GC*, *CYP2R1/PDE3B* and *DHCR7/NADSYN1*) are shown in red and novel loci (*OSBPL11*) are shown in green.

491



492

493

494

495

496

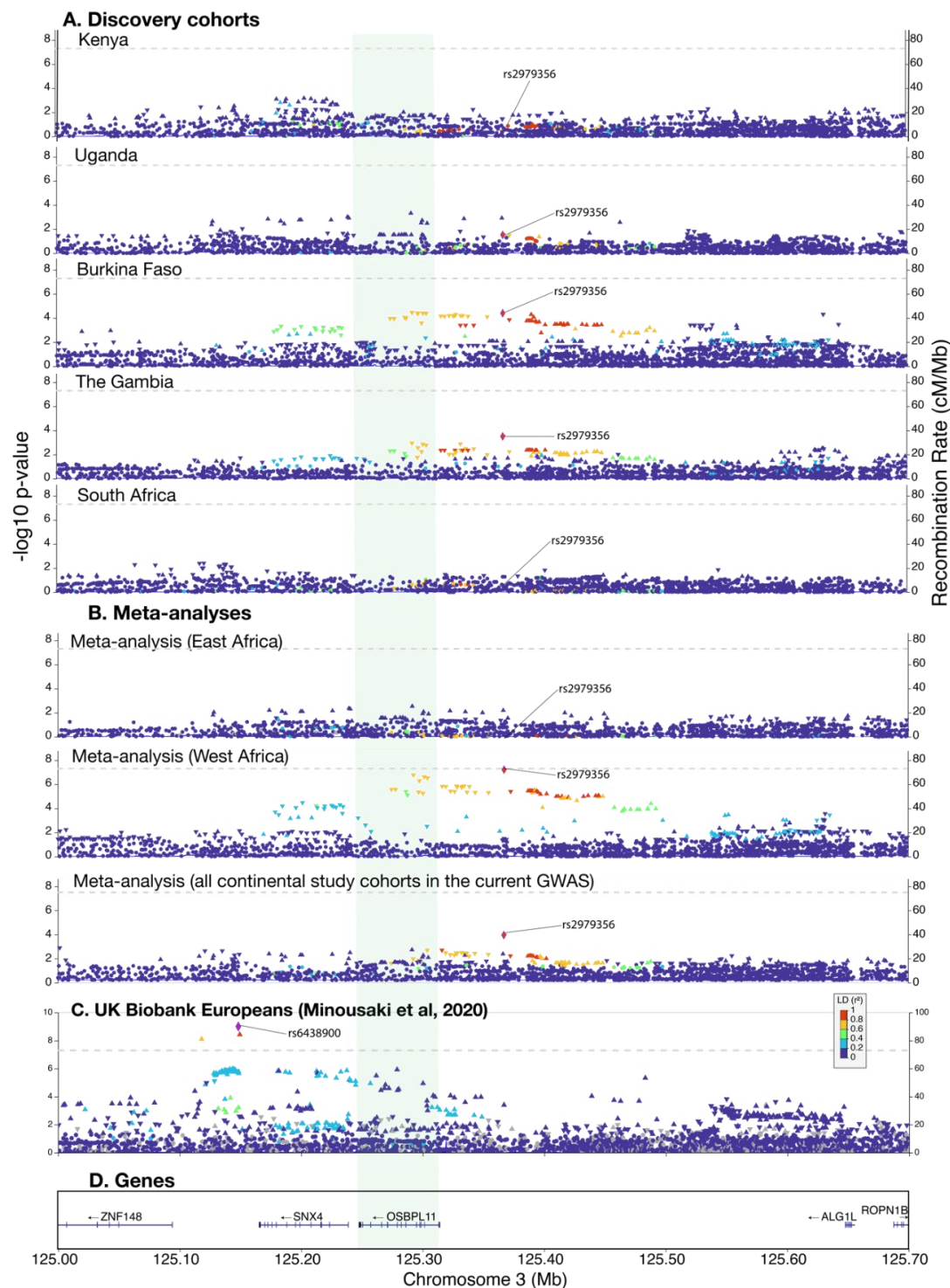
497

498

499

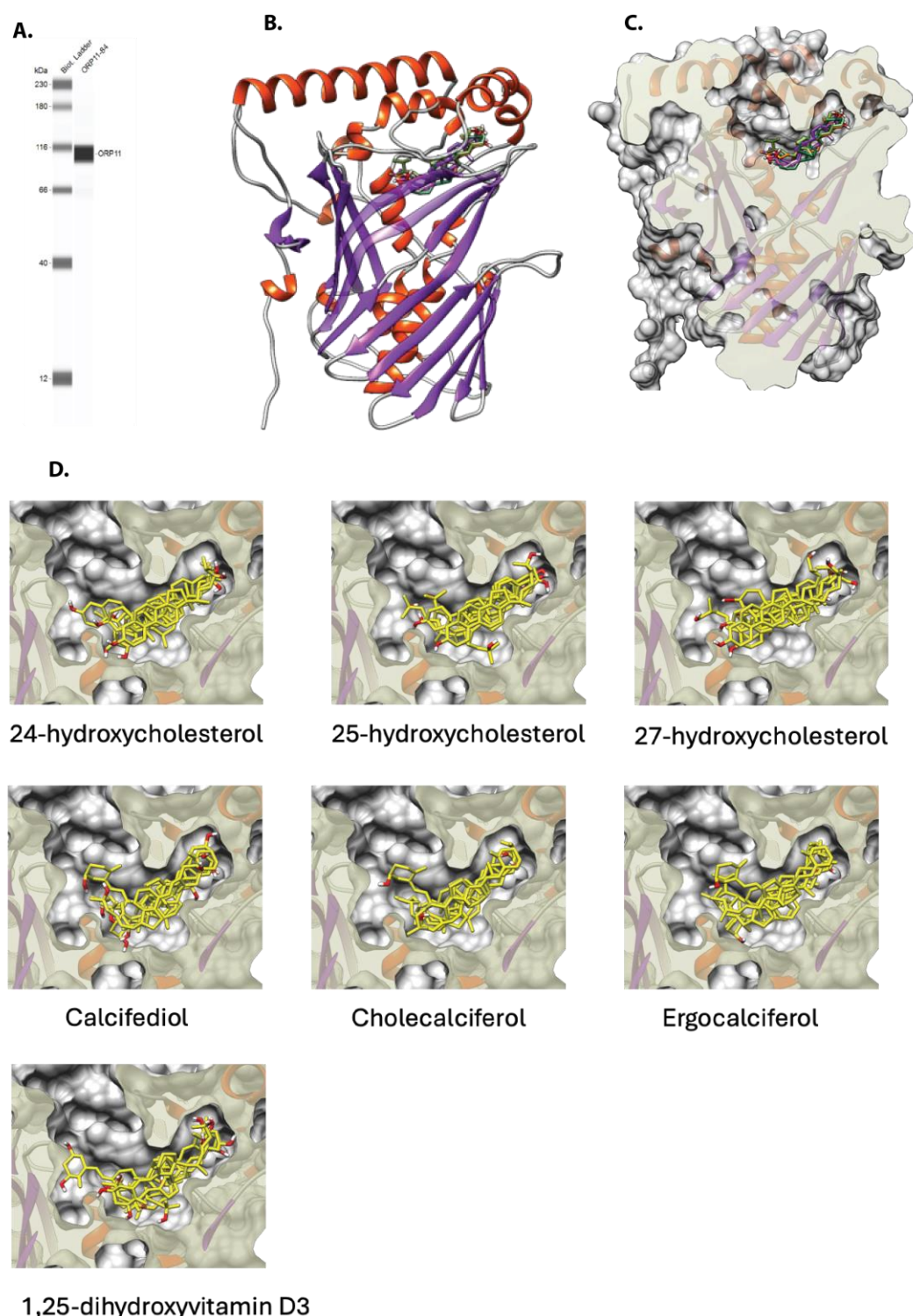
500

Supplementary Figure 5. Manhattan plot of MAGMA gene-based association test results and corresponding QQ plot (A) and MAGMA tissue expression analysis, where tissues that exceed the significance threshold ($P = 0.05/18912 = 2.64 \times 10^{-6}$) are shown in red (B), Dark blue areas highlight genomic risk loci. Green lines represent eQTL associations, while orange lines depict chromatin interactions. Genes mapped by both eQTL and chromatin interactions are highlighted in red. The gene mapping was conducted using FUMA using African ancestry data (<https://fuma.ctglab.nl/>) with the gene-based test performed via MAGMA (v1.08) ³⁹. Input variants were mapped to 18,912 protein-coding genes.

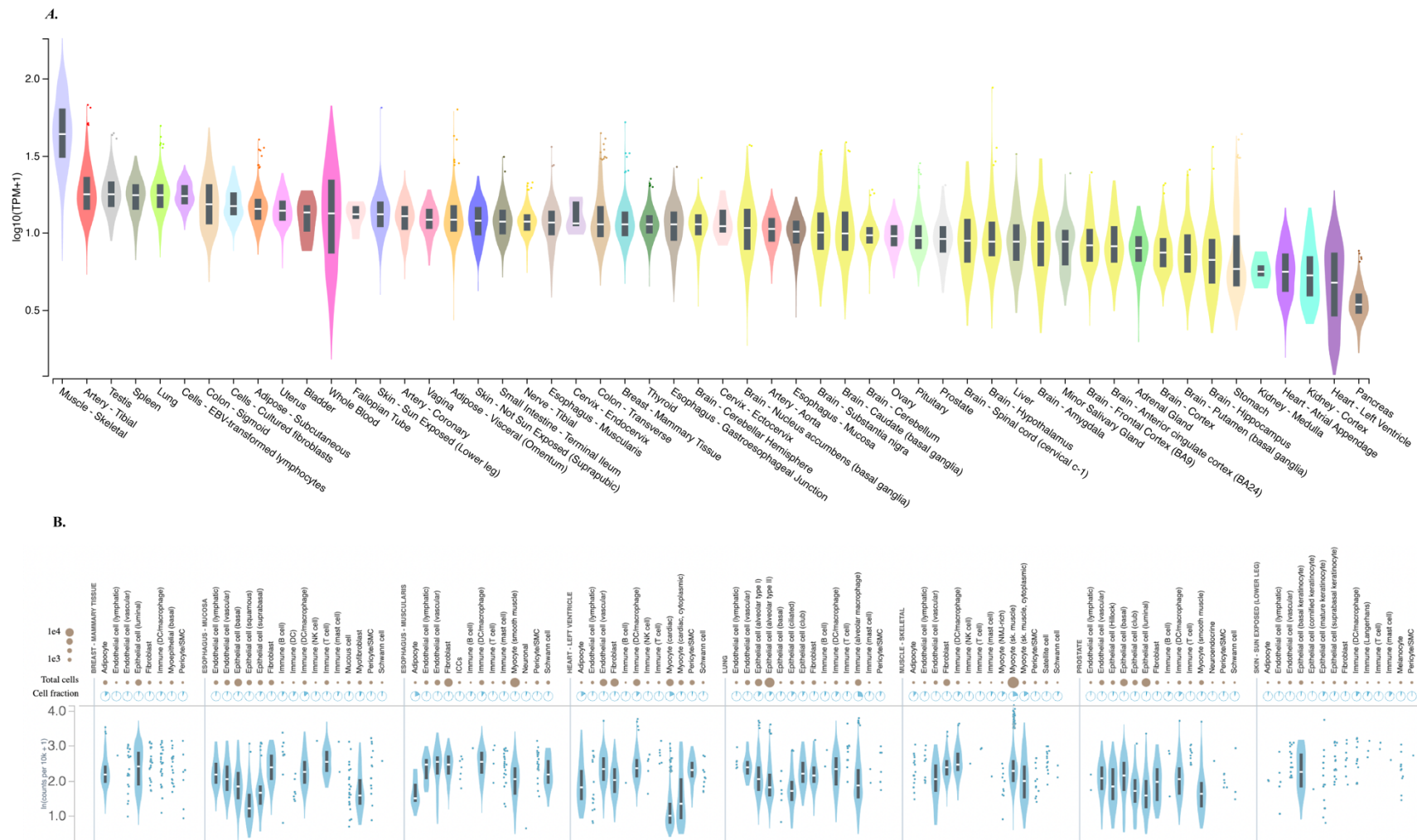


Supplementary Figure 6. Regional association plots for the *OSBPL11* locus in chromosome 3 for individual continental African cohorts (A), meta-analyses (B), UK Biobank Europeans (C), and genes in the regions (D). The direction of the triangle of the top variant indicates the direction of effect on 25(OH)D concentrations. The variants are coloured according to their LD (R^2) relative to the top variant. The reference variant (rs2979356) is illustrated in a purple diamond shape.

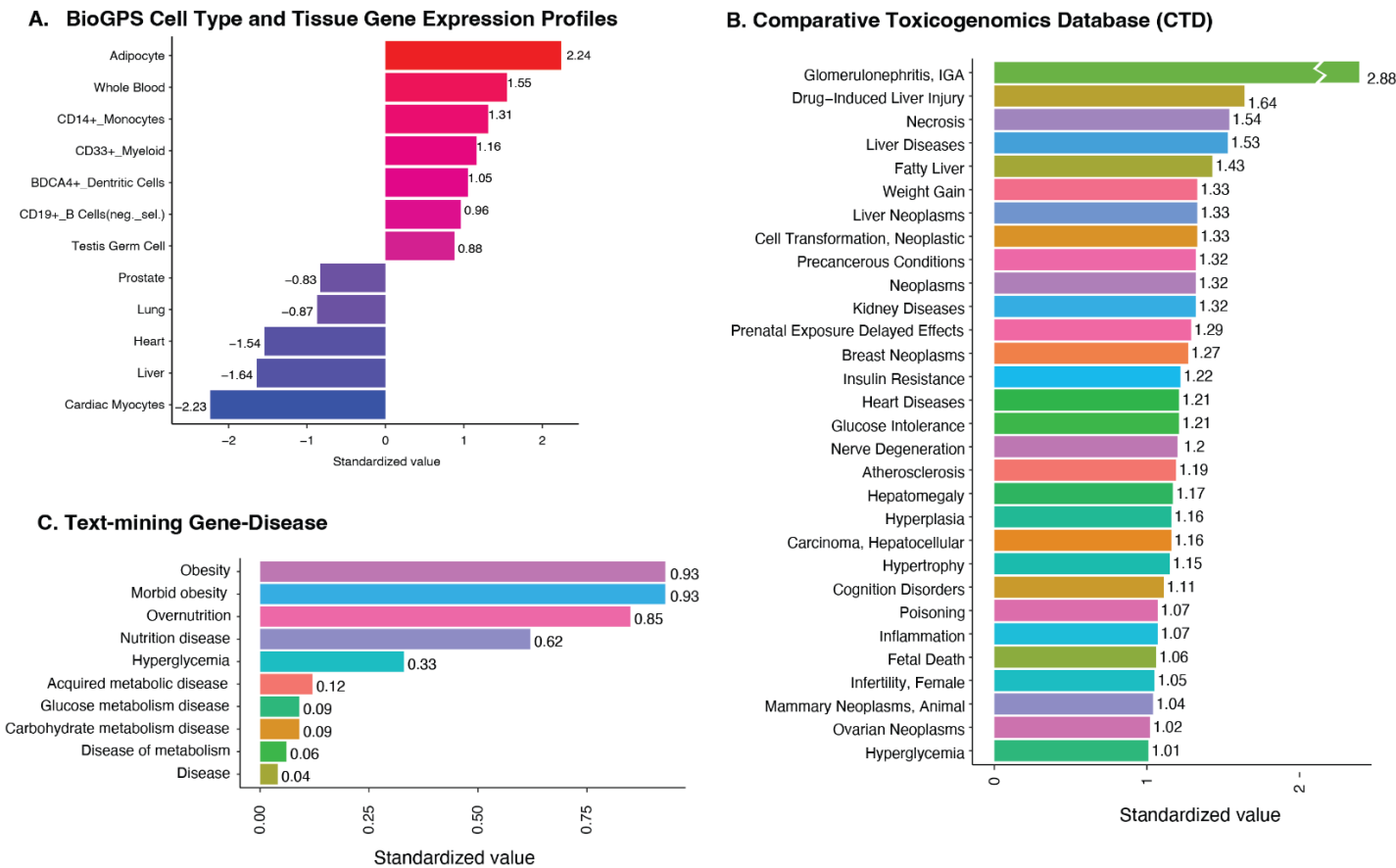
509



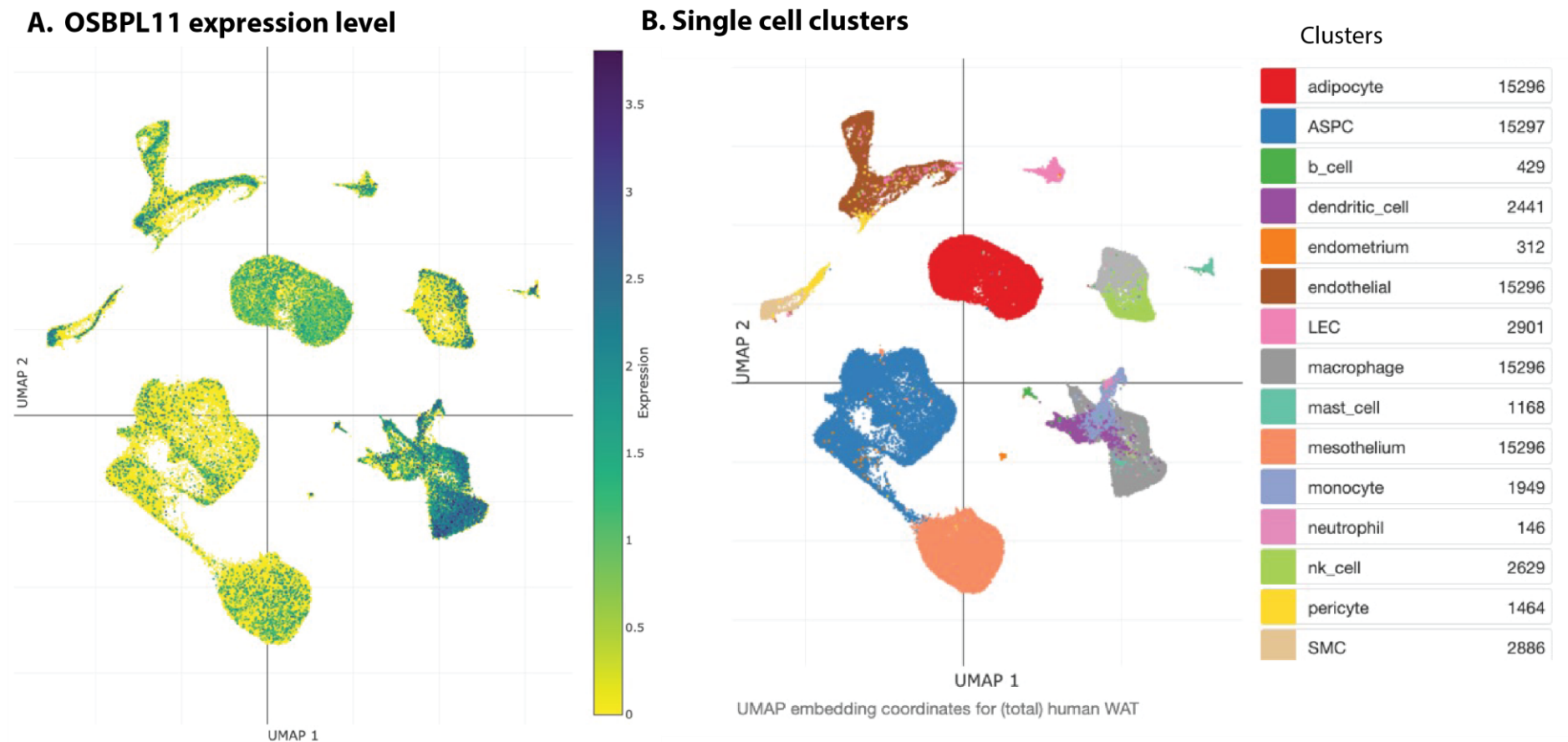
Supplementary Figure 7. OSBPL11 Alpha Fold model. Protein immunoassay of OSBPL11-myc-His overexpressed 293T lysate (A), OSBPL11 AlphaFold model (B), molecular docking of oxysterols and vitamin D compounds into oxysterol binding site in the AlphaFold modeled OSBPL11 (C), visualization of all binding poses for individual sterols and vitamin D analogs (D). Docking was performed using AutoDock Vina (v. 4.2.6, see Web Resources) plugin through UCSF Chimera (v. 1.17.3). Figures were made in UCSF Chimera (see Web Resources).



518
 519 **Supplementary Figure 8. Bulk tissue (A) and single-cell (B) gene expression for *OSBPL11* (ENSG00000144909.7).** Gene expression data
 520 were obtained from GTEX Portal version V8 (<https://gtexportal.org/>).



Supplementary Figure 9. Select *OSBPL11* search results from Harmonizome portal: differential gene-expression profiles in human tissues/cells (A) and mouse cell types in BioGPS datasets (B), [GEO Signatures of Differentially Expressed Genes for Diseases](#) (C), association of *OSBPL11* with diseases in Comparative Toxicogenomics Database (CTD) (D) and Text-mining Gene-Disease dataset (E). Complete Harmonizome (<https://maayanlab.cloud/Harmonizome/>) search results are presented in Supplementary Dataset 6. The standardized values shown represent z-scores calculated in the Harmonizome database, indicating how many standard deviations each gene's measurement is from the mean of the dataset, thus allowing for direct comparison across different datasets.



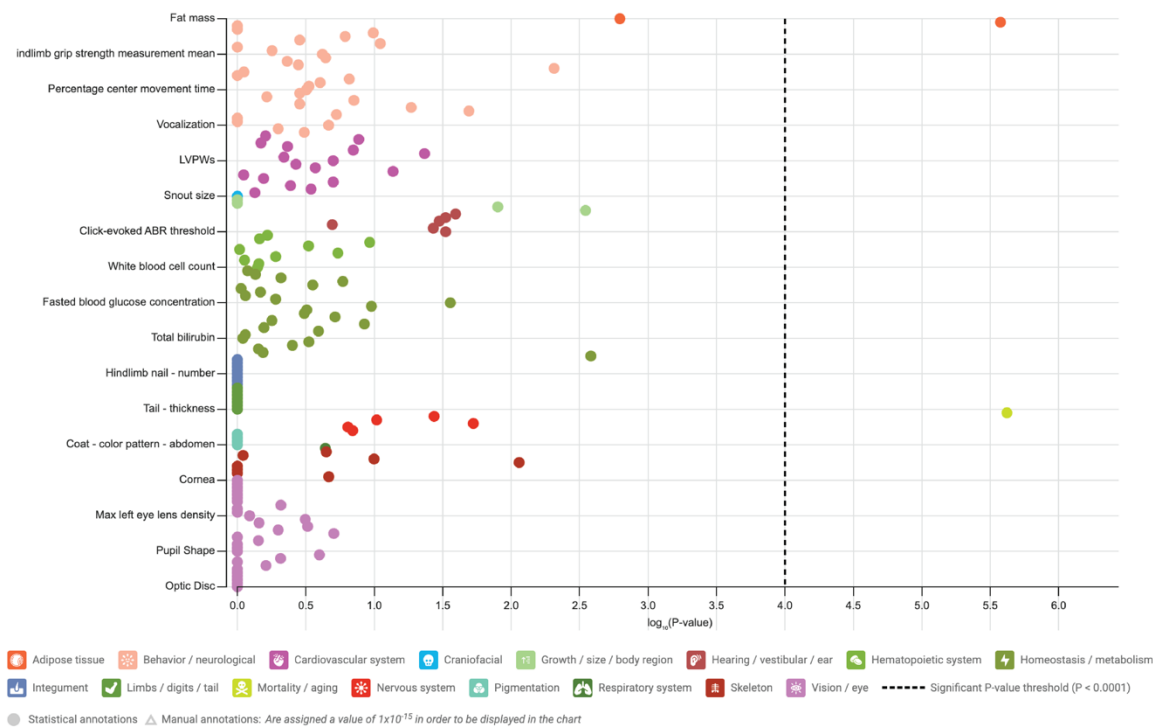
Supplementary Figure 10. Single cell atlas of human white adipose tissue showing expression levels for *OSBPL11*. Data was retrieved from https://singlecell.broadinstitute.org/single_cell/study/SCP1376/a-single-cell-atlas-of-human-and-mouse-white-adipose-tissue.

Abbreviations: OSBPL11, oxysterol binding protein like 11; UMAP, Uniform Manifold Approximation and Projection; B cell, B lymphocyte; T cell, T lymphocyte; LEC, lymphatic endothelial cell; NK cell, natural killer cell; SMC, smooth muscle cell; ASPC, adipose stem and progenitor cell.

A. Significant phenotypes

Phenotype	System	Allele	Zygosity	Life stage	Significant sexes	Significant P-value
decreased circulating triglyceride level	⚡	<i>Osbp11</i> ^{tm1b(EUCOMM)Hmgu}	Heterozygote	Early adult	♂	1.66x10 ⁻⁵
improved glucose tolerance	⚡	<i>Osbp11</i> ^{tm1b(EUCOMM)Hmgu}	Heterozygote	Early adult	♂	6.21x10 ⁻⁵
increased total body fat amount	🌀	<i>Osbp11</i> ^{tm1b(EUCOMM)Hmgu}	Heterozygote	Early adult	♂	2.66x10 ⁻⁶
preweaning lethality, complete penetrance	💀	<i>Osbp11</i> ^{tm1b(EUCOMM)Hmgu}	Homozygote	Early adult	♀	N/A *

B. Physiological systems



Supplementary Figure 11. Phenotypes with significant changes (A) and phenotyping screening results of *Osbp11* mouse knockout mice (B). The International Mouse Phenotyping Consortium (IMPC) applies a panel of phenotyping screens to characterize single-gene knockout mice in comparison to wild types. Data sourced from the IMPC (<https://www.mousephenotype.org/data/genes/MGI:2146553>).

Methods

Ethics statement

Written informed consent was obtained from the parents or guardians of each child prior to their enrolment in this study. Study protocols received approval from the local institutional review boards at each participating site.

Study participants (Discovery GWAS)

This GWAS study included young community-based children living in Kenya (n = 1361), Uganda (n = 1301), Burkina Faso (n = 329), The Gambia (n = 629), and South Africa (n = 889), as previously described⁶. Details of these cohorts are briefly described below.

Kilifi, Kenya (3.5° S, 39.9° E) –This is an ongoing community-based cohort aimed at evaluating immunity to malaria in children from birth to 8 years⁴⁰. The children were followed up weekly for malaria and other infections, and anthropometric measurements and blood samples were collected during annual cross-sectional surveys.

Entebbe, Uganda (0.1° N, 32.5° E) –This cohort was part of the Entebbe Mother and Baby Study (EMaBS) which was a prospective birth cohort study that was originally designed as a randomized controlled trial (ISRCTN32849447) to evaluate the effects of helminths and anthelmintic treatment during pregnancy and early childhood on immunological responses to routine vaccinations. Blood samples and anthropometric measurements were collected at birth and at subsequent annual visits up to age five years⁴¹.

Banfora, Burkina Faso (10.6° N, 4.8° W) –This cohort participated in the VAC050 ME-TRAP Malaria Vaccine trial aimed at testing the effectiveness, safety, and immunogenicity of a prime-boost liver-stage malaria vaccine in children aged between six and 17 months⁴². Blood samples were collected, and anthropometric measurements taken at multiple timepoints after receipt of the experimental vaccine based on another study investigating the genetics of responses against childhood vaccines⁴¹.

West Kiang, The Gambia (13.3° N, 16.0° W) –This was a community-based cohort study aimed at investigating the effect of genetic variants on anaemia during the malaria season in children aged between two and six years from 10 rural Gambian villages⁴³. Blood samples and anthropometric measurements from the start of the malaria season were included in this study.

Soweto, South Africa (26.2° S, 27.9° E)

The Soweto Vaccine Response Study recruited infants from vaccine trials coordinated by the Respiratory and Meningeal Pathogens Unit (<http://www.rmpru.com/>). The study used stored plasma samples collected at 12 months of age from infants that had received all their routine vaccines. Soweto, in South Africa is a non-malaria endemic region and anthropometry was not measured in this cohort⁴¹.

Genotyping and imputation

The Ugandan (n=1390), Burkinabe (n=354) and South African (n=749) discovery cohorts were genotyped on the Illumina HumanOmni 2.5 M-8 (“octo”) BeadChip array version 1.1 as part of GWAS (Illumina Inc., San Diego, USA), while the Kenyan (n=1241) and Gambian cohorts (n=550) were genotyped on the H3A Custom Genotyping Array v1.0 (H3Africa v1.0 contains approximately 2.5 million variants that are optimized for African ancestry

populations). Variant quality control (QC) was performed separately for each cohort using the H3Africa GWAS pipeline version 3 (see Web Resources) to exclude samples with genotype call rate < 97%, heterozygosity >3 SDs around the mean, discrepancy between genotype and allocated sex, and related individuals (identity by descent >0.9). Thereafter, variants passing the following filters were retained: >97% call rate, minor allele frequency >0.01 and Hardy–Weinberg equilibrium (HWE) >0.008. Imputation was performed using the African Genome Resources (AGR) Haplotype Reference Panel using the Sanger Imputation Service (see Web Resources). The AGR haplotype reference panel is based on 4956 samples from African and non-African 1000 Genomes Phase 3 populations, and 2298 additional African genomes from Uganda, Ethiopia, Egypt, Namibia, and South Africa, resulting in an increased number of imputed variants, improved information score and imputation accuracy compared to the 1000 Genomes Phase 3 Version 5 or TOPMED imputation panels. We filtered the resulting imputation dataset for variants with information score (info) ≥ 0.3 and MAF ≥ 0.01 for association analyses.

Laboratory assays

25(OH)D was measured using a chemiluminescent microparticle immunoassay (CMIA, Abbot Architect) in the discovery cohorts (Supplementary Table 6). The CMIA 25(OH)D assay has comparable results to the liquid chromatography tandem mass spectrometry LC-MS/MS method⁴⁴. Overall assay performance ranged from 2.8% to 7.9% for mean 25(OH)D concentrations ranging from 21 to 116 nmol/L. External quality assurance (DEQAS) data showed the 25(OH)D assay to have a mean (SD) bias of -2.7% (7.6) against the all-laboratory trended values, and one of -0.4% (7.7) against the target values. C-reactive protein (CRP) was measured using the MULTIGENT CRP Vario assay (Abbot Architect) and α 1-antichymotrypsin (ACT) using immunoturbidimetry (Cobas Mira Plus Bioanalyser, Roche) in plasma samples stored at -80 °C (Supplementary Table 1). Malaria parasitaemia was detected using Giemsa-stained blood smears.

Genome-wide association study (GWAS) analysis

To identify variants associated with 25(OH)D concentrations, we performed a linear mixed model GWAS within each study cohort as implemented in Genome-wide Complex Trait Analysis software version 1.24.4 (GCTA)⁴⁵. GCTA uses a sparse genomic relationship matrix (GRM) to account for genetic structure within the cohort. The GRM was generated from unimputed genotype data using GCTA. To normalise 25(OH)D concentrations we rank-based inverse-normal transformed (RINT) 25(OH)D concentrations in the model and included age, sex, and season (at the time of vitamin D measurement) as covariates. We adjusted for season since seasonality is associated with vitamin D status in children living in Africa⁶. BMI was not included as it was not associated with vitamin D status in this population⁶. We computed and plotted PCs using PLINK software (version 2.0) to identify and visualize population structure and identify potential outliers and admixed individuals within the discovery cohorts. Vitamin D supplementation is rare in children living in rural African communities, and these data were not available. To identify conditionally independent associations at each genome-wide significant locus, we performed conditional and joint analyses using GCTA-COJO version 1.24.4⁴⁶ using our GWAS summary data. We performed the association analyses using GCTA “cojo-slc” option conditioning on the lead variant and repeated the process if there was a secondary signal that reached genome-wide significance. We considered variants that reached genome-wide significance before and after conditioning to be conditionally independent. We used an inverse-variance-weighted fixed effect approach as implemented in METASOFT⁴⁷ (see Web Resources) to meta-analyse effect estimates across the study cohorts.

Statistical power estimates

The power of the study for discovery was estimated using *Quanto*⁴⁸. For a variant with a minor allele frequency (MAF) of 0.05, the study had 80% power to detect a genetic effect size (GES) of 8.6 (r^2 of 0.011) and 97.7% power to detect a GES of 10.0 (r^2 of 0.015). For a variant with a MAF of 0.10, the study has 80% power to detect a GES of 6.2 and 95% power to detect a GES of 7.0.

Replication in African ancestry cohorts

To evaluate replication of our findings we performed a linear mixed model GWAS as implemented in GCTA in community-based cohorts of participants of African ancestry from the Social Change Asthma and Allergy in Latin America study (SCAALA), the Jackson Heart Study (JHS), the UK Biobank and the All of Us Research Program (Supplementary Table 1). The SCAALA Cohort (Salvador, Bahia, Brazil (12.9° S, 38.5° W)) is a longitudinal study in children initiated in 1997 aimed at assessing the impact of sanitation on health⁴⁹. By 2013, the cohort included 1,206 participants aged 11-19 years, with 750 individuals selected for this analysis. The JHS is an ongoing community-based observational study with 5306 African American adult participants (Jackson, USA, 32.3° N, 90.2° W) aimed at identifying the environmental and genetic risk factors for the development of cardiovascular diseases among an African American population (see Web Resources). Details of the design, recruitment and initial characterization of the JHS study has been described in detail elsewhere⁵⁰. Concentrations of 25(OH)D and CRP were measured in samples collected from clinic visits from 2000-2004. The Multi-Ethnic Study of Atherosclerosis (MESA) is a community-based cohort study initiated in 2000 to investigate subclinical cardiovascular disease. It includes 6,814 participants aged 45-84 years from six US field centers with African American, Hispanic, Chinese, and White ancestry, all free of cardiovascular disease at baseline. African ancestry participants were included in this analysis for replication (dbGaP accession phs000209.v13.p3). The UK Biobank (UKB, n=500,000) is a prospective cohort study following the health and well-being of middle-aged and older adults recruited throughout the UK between 2006 and 2010⁵¹. Concentrations of 25(OH)D and CRP and anthropometry were measured during a baseline evaluation at UKB assessment centres. We performed imputation, quality control, and association analyses using the same parameters and software as for the discovery GWAS. We performed a linear mixed model GWAS (with GRM) as implemented in GCTA for the replication cohorts and adjusted for cohort-specific covariates linked to vitamin D levels: age and sex for SCAALA and age, sex, season, vitamin D supplementation and BMI for UKB and JHS (assessment centre also included for the UKB). Additionally, we conducted a meta-analysis including both the discovery and replication cohorts (Brazilians in SCAALA, African American participants in the JHS and African ancestry participants in the UKB). The All of Us Research Program is a nationwide initiative in the United States aiming to collect health data from one million or more participants to accelerate research and improve health outcomes⁵². This program places a strong emphasis on diversity, with over 80% of participants coming from groups historically underrepresented in biomedical research, including significant representation of individuals of African ancestry. Since the GWAS for the All of Us study had already been performed, we incorporated the summary statistics obtained from their ALL by ALL platform into our analysis⁵². Participants in JHS and MESA were selected based on self-reported race, without exclusion of ancestry outliers. While self-reported population descriptors reflect social identity and should not be interpreted as proxies for genetic ancestry, prior studies have shown that most participants in both cohorts exhibit high proportions of African ancestry⁵³.

Analyses of the All of Us Research Program and UK Biobank have used genetically inferred African ancestry to characterize participants.

Fine-mapping and functional annotations

We used the FUMA online pipeline version 1.4.1⁵⁴ to annotate the GWAS results. We defined independent significant variants as variants with $P < 5 \times 10^{-8}$ and a pairwise $r^2 \leq 0.6$. We further defined lead variants from the independent significant variants as variants with $r^2 < 0.1$. Genomic risk loci were defined by clustering independent significant SNPs that were in LD ($r^2 \geq 0.1$) and <250 kb apart. Each locus is represented by the top lead variant which has the minimum P -value in the locus. Genetic data from African populations in the 1000G phase3 were used as reference data to conduct LD analyses in FUMA. Genomic loci were mapped to genes using positional mapping (loci located in the genes or < 10 kb away) and expression quantitative trait locus (eQTL) mapping (loci were mapped to a gene if the corresponding variants had a significant effect on the expression of the gene using a false discovery rate (FDR) of < 0.05). eQTL mapping was performed using data from 27 tissues (Genotype-Tissue Expression (GTEx) version 8 Whole Blood and Minor Salivary Gland⁵⁵, Database of Immune Cell Expression (DICE)⁵⁶, single-cell RNA eQTL⁵⁷ and Biobank-Based Integrative Omics Study (BIOS) QTL Browser⁵⁸). Additionally, we searched for lead variants (rs2979356) and *OSBPL11* gene expression profiles in 54 human tissues and organs in the GTEx portal v8⁵⁹ (<https://gtexportal.org/>). In addition, we also retrieved the expression profile of *OSBPL11* in the single cell atlas of human and mouse white adipose (see Web Resources).

To identify the most likely causal variant(s) for the four independent loci, we subsequently utilized FINEMAP to estimate the number of most likely causal variants per locus and generated a 95% credible set for each causal variant⁶⁰. We used PolyPhen, a tool designed for predicting the possible impact of amino acid substitutions on the structure and function of human proteins, to systematically examine the credible sets for the three loci. We retrieved data for *OSPL11* from Harmonizome (see Web Resources), which is a collection of information about genes and proteins from 121 datasets provided by 72 online resources. We also analyzed the protein-protein interaction network of *OSBPL11* using data from the STRING database (see Web Resources). We also examined the genomic context and regulatory annotations of rs2979356 using data from the Haploreg database (see Web Resources). We searched for information on rs2979356 and *OSPL11* in the NHGRI-EBI GWAS catalog, PheGenI, dbGaP, ClinVar, PopHumanScan or in EGA, and Relate to identify their associations with diseases, phenotypes, population-specific variations, clinical significance and evidence of natural selection (see Web Resources for links). We additionally investigated the impact of our lead *OSPL11* variant on risk of severe malaria within the MalariaGEN dataset²¹ and on risk of tuberculosis infection in the International Tuberculosis Host Genetics Consortium²² (see Web Resources) to determine whether malaria and/or tuberculosis could be responsible for the observed selection pressure. We searched the International Mouse Phenotyping Consortium (IMPC) server for data on the effect of knocking out *Osbpl11* in mice. IMPC conducted a comprehensive panel of phenotyping screens to characterize the *Osbpl11* single-gene knockout mice in comparison to wild-type controls.

Gene-based analyses

We used the GWAS summary data to calculate gene-based P -values for association between genes and 25(OH)D concentrations using MAGMA (v1.08)³⁹ gene analysis as implemented in FUMA⁵⁴ using standard settings. In total, 18,912 genes were tested, and significance of P -

values assessed based on Bonferroni-corrected significant threshold P -values of $0.05/18,912 = 2.644 \times 10^{-6}$. We also used MAGMA through the FUMA interface to perform gene set analyses to evaluate the association between 25(OH)D concentrations and biological pathways. We tested a total of 10,678 biological pathways (5,917 gene ontology (GO) terms and 4,761 curated gene sets from the Molecular Signatures Database (MsigDB)^{61,62}), and those with P -values below a Bonferroni-corrected significance threshold of 1×10^{-6} were considered significant.

Transferability of established 25(OH)D loci

We looked for evidence of transferability of previously reported loci in African diaspora⁷ and European populations³ to continental Africans. Variants with P -values < 0.05 and the same direction of effect in both European and African populations were considered as transferable. For those variants that were missing or did not show a significant association in our GWAS, we examined all variants with LD $r^2 > 0.3$ and within 250 kb of the reported index variant for association with 25(OH)D concentrations. When multiple SNPs from a region were evaluated, the nominal association P -values were adjusted for the total number of SNPs within the region using the method of effective degrees of freedom^{63,64}. We considered a locus to be replicated if at least one of the tested SNPs had a P value of less than the adjusted association P value.

Molecular docking of *OSBPL11* to common oxysterols and vitamin D metabolites

We performed molecular docking analyses using PyRx version 0.9⁶⁵, a graphical user interface for AutoDock Vina⁶⁶. We retrieved the 3D structure of OSBPL11 (PDB format) from AlphaFold Protein Structure Database (see Web Resources). We prepared and optimized the 3D structure for docking by removing any water molecules and generating missing side chains and loop regions using PyRx. Spatial Data File (SDF) format files of common oxysterols (25-hydroxycholesterol, 7- α -hydroxycholesterol, 27-hydroxycholesterol, and 24(S)-hydroxycholesterol), and vitamin D metabolites (namely vitamin D2, vitamin D3, 1,25-dihydroxyvitamin D and 25-hydroxyvitamin D) were retrieved from the PubChem database (see Web Resources). The ligand structures were prepared using PyRx to generate 3D coordinates and protonate the ligands at pH 7.0. The oxysterol binding site of OSBPL11 was defined using the receptor grid generation tool in PyRx with the binding site centered on the box. The docking parameters were set to default values, and 10 docking runs were performed for each ligand. The docking results were analyzed using the PyMOL Molecular Graphics System (Version 2.4) to visualize the protein-ligand interactions and identify the best binding conformation. We used R version 4.0.0 software to plot the lowest binding energies for each oxysterol and vitamin D metabolite for the binding conformation.

OSBPL11 Alpha Fold modeling and ligand docking

The OSBPL11 Alpha Fold model was downloaded and used without alteration. Ligand docking was carried out using the AutoDock Vina (v. 4.2.6) plugin through UCSF Chimera (v. 1.17.3). Ligands were generated from their CID code and minimized before docking. All components were protonated using Gasteiger. Docking was carried out three times, each with unique seed values; this produced essentially identical poses but with small variations in the resulting score (kcal/mol) values. Figures were made in UCSF Chimera, and the graph was generated using Prism 10 v. 10.2.3.

OSBPL11 *in vitro* ligand binding assay

Materials: [³H]-25-hydroxycholesterol (25[26,27 ³H]-OHC) (catalog # 50-353-012) was purchased from Fischer Scientific. Calcitriol (1,25-dihydroxyvitamin D) was purchased from

MedChemExpress (catalog # HY-113134). Cabot Norit A2 (catalog # DARCO Cas:7440-44-0) charcoal obtained from Sigma Aldrich and Dextran-500 (catalog # D1004) purchased from Spectrum chemicals. Microscint-20 (part.no. 6013621) was purchased from Perkin Elmer. pcDNA3.1 (+)myc/his C was cloned by and purchased from GenScript. Binding assay adapted from established OSBP/ORP *in vitro* binding ³H-25-hydroxycholesterol charcoal/dextran binding assay as described¹⁷⁻¹⁹. OSBPL11cDNA was cloned into pcDNA3.1 (+)myc/his C mammalian expression plasmid, which was transfected into 293T cells with Lipofectamine2000. OSBPL11-myc-His 293T lysate was subjected to a 100,000 g ultracentrifugation, and OSBPL11-myc-His expression verified by protein immunoassay (Supplementary Fig. 11A). For calculation of K_i values in the competition binding curves, a K_D of 25 nM for 25-OHC against OSBP was used, based on the measured K_D of 25-OHC against OSBPL11¹⁸.

Effect of *OSBPL11* lead variant (rs2979356) on cardiometabolic traits in African ancestry populations

In this study, we gathered effect size estimates (beta coefficients and corresponding standard errors) for the lead *OSBPL11* variant, rs2979356, from previously conducted genetic association analyses across multiple African ancestry cohorts. These analyses, carried out by each respective study or consortium (African American Diabetes Mellitus (AADM), UK Biobank, African American Cardiometabolic Disease Knowledge Portal (AMP-CMDKP), Africa Wits-INDEPTH Partnership for Genomic Studies (AWI-Gen), and the Global Lipids Genetics Consortium (GLGC)), focused on cardiometabolic traits such as body composition, lipid profiles, blood pressure, and measures of glucose homeostasis. The AADM study enrolled participants of East and West African ancestry to investigate cardiometabolic diseases. The AWI-Gen study recruited individuals from multiple sub-Saharan African countries (Burkina Faso, Ghana, Kenya, Namibia, South Africa, and Tanzania) to examine genetic and environmental risk factors for cardiometabolic traits. The AMP-CMDKP collates data from diverse African ancestry populations to identify genetic contributions to cardiometabolic diseases. GLGC is a large-scale collaboration focusing on lipid traits in ethnically diverse populations. Finally, the UK Biobank (UKB) African ancestry subset includes participants of African descent within the broader UKB cohort. After confirming consistency in variant reference alleles, each study's results were extracted in their final, post-quality control form. No additional analyses were performed by the authors; rather, the compiled data were integrated into a single summary table (Table 2), highlighting the associations of rs2979356 with various phenotypes and emphasizing its potential relevance to both lipid metabolism and cardiovascular health in African ancestry populations.

Web resources (links to web resources that were used in the study)

For H3Africa GWAS pipeline version 3 see <https://github.com/h3abionet/h3agwas>. For FUMA see <https://fuma.ctglab.nl/>. All of Us Research Program see <https://databrowser.researchallofus.org/>. For Sanger Imputation Service see <https://imputation.sanger.ac.uk/>. Description of Jackson Heart Study see www.jacksonheartstudy.org. For METASOFT see <http://genetics.cs.ucla.edu/meta/index.html>. For Open Targets Genetics database <https://genetics.opentargets.org/>. For GTEx portal v8 see <https://gtexportal.org/home/>. For GTEx Portal see <https://gtexportal.org/>. For single cell atlas of human and mouse white adipose tissue https://singlecell.broadinstitute.org/single_cell/study/SCP1376/a-single-cell-atlas-of-human-and-mouse-white-adipose-tissue. For AlphaFold Protein Structure Database see <https://alphafold.ebi.ac.uk/>. For AutoDock Vina <https://vina.scripps.edu/>. For UCSF Chimera <https://www.cgl.ucsf.edu/chimera/>. For Ensembl Variant Effect Predictor see

<https://www.ensembl.org/vep>. For Harmonizome see <https://maayanlab.cloud/Harmonizome/>. For NHGRI-EBI GWAS catalog see <https://www.ebi.ac.uk/gwas>. For STRING see <https://string-db.org/>. For ClinVar see <https://www.ncbi.nlm.nih.gov/clinvar/>. For PheGenI see <https://www.ncbi.nlm.nih.gov/gap/phegeni>. For dbGaP see <https://www.ncbi.nlm.nih.gov/gap/>. For data from Accelerating Medicines Partnership in Common Metabolic Diseases Knowledge Portal (AMP-CMDKP) see <https://hugeamp.org/>. For EGA see <https://ega-archive.org/>. For PheWAS see <https://gwas.mrcieu.ac.uk/phewas>. For HaploReg v4.2 see <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>. For LDpair see <https://ldlink.nih.gov/?tab=ldpair>. For African Functional Genomics Resource (AFGR) see <https://github.com/smontgomlab/AFGR>. For JHS eQTL see <http://jhsqtl.genetics.unc.edu/eQTL.php>. For PubChem database see <https://pubchem.ncbi.nlm.nih.gov/>. For PhenoScanner database see <http://www.phenoscaner.medschl.cam.ac.uk/>. For FINEMAP see <http://christianbenner.com/>. For PolyPhen see <http://genetics.bwh.harvard.edu/pph2/>. For SIFT see <https://sift.bii.a-star.edu.sg/>. For dbSNP see <https://www.ncbi.nlm.nih.gov/snp/>. For PopHumanScan see <https://pophumanscan.uab.cat/>. For MalariaGEN see <https://apps.malariagen.net/>. For International Mouse Phenotyping Consortium data see <https://www.mousephenotype.org/data/genes/MGI:2146553>.

Funding

This work was funded by Wellcome (grant numbers SHA 226014, TNW 202800, AJM 106289, AME 064693, 079110, and 095778), as well as core awards to the KEMRI-Wellcome Trust Research Programme (203077). RMM, MG, ARB, CR, and AA's work was made possible by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH), which is itself supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology, and the Office of the Director at the National Institutes of Health. RMM's work was supported by the Developing Excellence in Leadership, Training and Science (DELTAS) Africa Initiative (DEL-15-003). The authors are solely responsible for the views expressed in this publication, which do not necessarily reflect the views of AAS, NEPAD Agency, Wellcome, or the UK government. The authors have opted to apply a CC-BY public copyright license to any accepted manuscript version arising from this submission for the purpose of Open Access. The funders did not play a role in the study design, data collection, data analysis, data interpretation, or writing of the report. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Data availability

The GWAS summary statistics are available at GWAS catalog (Accession number XXX). The data and analyses scripts underlying this article are available in Harvard Dataverse at XXXXX and applications for data access can be made through the Kilifi Data Governance Committee cgmrc@kemri-wellcome.org.

Acknowledgement

We express our gratitude to our colleagues at the KEMRI-Wellcome Trust Research Programme, specifically Jedidah Mwacharo and Barnes Kitsao, for their assistance in retrieving archived samples. Additionally, we extend our thanks to the teams at the UVRI/MRC Entebbe Mother and Baby Study, the Malaria Vectored Vaccines Consortium (MVVC) in Burkina Faso, and the Respiratory and Meningeal Pathogens Unit (RMPRU) in

Johannesburg, South Africa. We would also like to acknowledge the Oxford Genomics Centre at the Wellcome Centre for Human Genetics, which was funded by a Wellcome Trust grant, reference 203141/Z/16/Z, for their role in generating and processing the genotype data for samples from Kenya and The Gambia. MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-RR-025005, and UL1-TR-000040. Funding support for the Vitamin D dataset was provided by grant HL096875. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. We thank the International Tuberculosis Host Genetics Consortium (ITHGC) for providing access to GWAS summary statistics. The Jackson Heart Study is supported by Contracts HHSN268201800010I, HHSN268201800011I, HHSN268201800012I, HHSN268201800013I, HHSN268201800014I, HHSN268201800015I from the National Heart Lung and Blood Institute (NHLBI) with additional support from the National Institute of Minority Health and Health Disparities (NIMHD). The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institute of Minority Health and Health Disparities (NIMHD); the National Institutes of Health; or the U.S. Department of Health and Human Services. This study was published with the permission of the Office of the Director of KEMRI.

Description of Additional Supplementary Files

Title: Supplementary Dataset 1. FUMA gene-based analysis results

Description: Gene-based analysis results. Analyses were conducted with MAGMA v1.08, using default parameters (SNP-wide mean model) and with the African 1000 Genome Phase3 selected as reference panel. Columns are: GENE, gene Ensembl ID; CHR, chromosome; START/STOP, annotation boundaries of the gene on that chromosome; NSNPs, number of SNPs annotated to the gene that were found in the data and were not excluded based on internal SNP QC; NPARAM, number of relevant parameters used in SNP-wise (mean) model; N, sample size; ZSTAT, the Z-value for the gene, based on its (permutation) P-value (this is what is used as the measure of gene association in the gene-level analyses); P, the gene P-value, using asymptotic sampling distribution (if available); SYMBOL, Gene name.

Title: Supplementary Dataset 2. MAGMA gene-set analysis output

Description: Gene sets were updated for MsigDB v7.0. Total 15496 gene sets (Curated gene sets: 5500, GO terms: 9996) were tested. Curated gene sets consist of 9 data resources including KEGG, Reactome and BioCarta. GO terms consists of three categories, biological processes (bp), cellular components (cc) and molecular functions (mf). All parameters were set as default (competitive test). Columns are: VARIABLE, name of the gene set, gene covariate or interaction. Names in this column are capped at 30 (by default) characters to keep the output file more readable. FULL_NAME, the full variable name; only included if the variable names exceed the maximum length for the VARIABLE column (and the abbreviate option is not set to file mode); TYPE, denotes the type of variable, either SET or COVAR for normal gene sets and gene covariates provided in the input files; or INTER-SS or INTER-SC, for internally created interaction terms (set by set and set by covariate, respectively); NGENES, the number of genes in the data that are in the set (for gene sets and

set by covariate interactions), that are in the interaction set (for set by set interactions), or for which non-missing values were available (for gene covariates); BETA, the regression coefficient of the variable; BETA_STD, the semi-standardized regression coefficient, corresponding to the predicted change in Z-value given a change of one standard deviation in the predictor gene set / gene covariate (i.e.. BETA divided by the variable's standard deviation). SE, the standard error of the regression coefficient. P, p-value for the parameter / variable.

Title: Supplementary Dataset 3. FUMA results for annotation of genes to tissue enriched expression.

Description: Tissue expression analysis. MAGMA gene-property test was performed for average gene-expression per category (e.g. tissue type or developmental stage) conditioning on average expression across all categories (one-side). Analysis was conducted with MAGMA v1.06. Columns are: VARIABLE, name of the gene set, gene covariate or interaction; TYPE, the type of variable, either SET or COVAR for normal gene sets and gene covariates provided in the input files; NGENES, the number of genes in the data that are in the set (for gene sets and set by covariate interactions), that are in the interaction set (for set by set interactions), or for which non-missing values were available (for gene covariates); BETA, the regression coefficient of the variable; BETA_STD, semi-standardized regression coefficient, corresponding to the predicted change in Z-value given a change of one standard deviation in the predictor gene set / gene covariate (i.e.. BETA divided by the variable's standard deviation); SE, standard error of the regression coefficient; P, P-value for the parameter/variable; FULL_NAME: the full variable name; only included if the variable names exceed the maximum length for the VARIABLE column. General parameters and settings for the analysis are included at the top. These are: TOTAL_GENES, the total number of genes included in the analysis; TEST_DIRECTION, the testing direction used for the different types of parameters; CONDITIONED_INTERNAL/RESIDUALIZED/HIDDEN/VARIABLES, the internal covariates and external variables which the analysis was conditioned on.

Title: Supplementary Dataset 4. GWAS summary statistics and model-averaged posterior summaries for variants predicted to be causal

Description: File, the GWAS summary statistics and model-averaged posterior summaries for each variant one per line. Columns are: RSID, the variant identifiers (rsID number); CHROMOSOME, the chromosome names; POSITION, the base pair positions; ALLELE1, the "first" allele of the variants; ALLELE2, the "second" allele of the variants; MAF, the minor allele frequencies; BETA, the estimated effect sizes as given by GWAS software; SE, the standard errors of effect sizes as given by GWAS software; Z, the z-scores; PROB, the marginal Posterior Inclusion Probabilities (PIP). The PIP for the l^{th} variant is the posterior probability that this variant is causal; LOG10BF, the log10 Bayes factors. The Bayes factor quantifies the evidence that the l^{th} variant is causal with log10 Bayes factors greater than 2 reporting considerable evidence; MEAN, the marginalized shrinkage estimates of the posterior effect size mean for the same allele as in beta; SD, the marginalized shrinkage estimates of the posterior effect size standard deviation; MEAN_INCL, the conditional estimates of the posterior effect size mean for the same allele as in column beta; SD_INCL, the conditional estimates of the posterior effect size standard deviation.

Title: Supplementary Dataset 5: eQTL Summary Statistics for Multiple Tissues

Description: This dataset contains the eQTL (expression quantitative trait loci) summary statistics for variant rs2979356 across various tissues. Each row represents a unique tissue

and provides information about the associated gene, the variant, the effect size, and the significance of the association. Columns: Gencode Id: The unique identifier for each gene from the Gencode database; Gene Symbol, the common symbol or name of the gene; Variant Id, The identifier for the variant, indicating its genomic location and specific alteration; variant Id, The dbSNP identifier for the single nucleotide polymorphism (SNP); P-Value, The statistical significance of the eQTL effect in each tissue. Lower values indicate higher significance; NES (Normalized Effect Size), The magnitude of the eQTL effect. Positive values indicate an increase in gene expression, and negative values indicate a decrease; Tissue, the specific tissue where the eQTL effect was observed.

Title: Supplementary Dataset 6: Gene Associations for OSBPL11 from the Harmonizome Database

Description: This file contains gene association data for OSBPL11 across various biological datasets, retrieved from the Harmonizome database. Each row represents a specific association between OSBPL11 and a biological trait, condition, or molecular feature. The columns are as follows: ASSOCIATION, the biological trait, condition, or molecular feature associated with OSBPL11 (e.g., tissue types, gene expression profiles, transcription factor binding sites); DATASET, the source dataset from which the association was derived (e.g., Allen Brain Atlas, ENCODE, GTEx, TCGA); THRESHOLD VALUE, the cutoff for determining the significance of the association, with values of 1 or -1 indicating positive or negative associations; STANDARDIZED VALUE, The magnitude of the association between OSBPL11 and the trait, with positive values indicating upregulation and negative values indicating downregulation in the relevant context. Null values indicate cases where no significant data were available.

Title: Supplementary Dataset 7: PheWAS Summary Statistics for rs2979356 Associated with Various Health Traits.

Description: This file contains PheWAS (Phenome-Wide Association Study) summary statistics for the variant rs2979356, which has been linked to multiple health traits. Each row represents a specific trait association. The columns are as follows: ID: the unique identifier for each trait or study; TRAIT, the health trait associated with rs2979356; POSITION, the base pair position of the variant on chromosome 3; P: the p-value, indicating the statistical significance of the association; SE: the standard error of the effect size estimate; N, the sample size used in each analysis (if available); BETA, the estimated effect size of the variant on the trait; CHR, the chromosome number where rs2979356 is located; EA, the effect allele of rs2979356; NEA, the non-effect allele of rs2979356; EAF, the effect allele frequency in the population.

References

- 1 Mogire, R. M. *et al.* Prevalence of vitamin D deficiency in Africa: a systematic review and meta-analysis. *Lancet Glob Health* **8**, e134-e142 (2020).
[https://doi.org/10.1016/S2214-109X\(19\)30457-7](https://doi.org/10.1016/S2214-109X(19)30457-7)
- 2 Holick, M. F. Vitamin D deficiency. *N. Engl. J. Med.* **357**, 266-281 (2007).
<https://doi.org/10.1056/NEJMr070553>
- 3 Manousaki, D. *et al.* Genome-wide Association Study for Vitamin D Levels Reveals 69 Independent Loci. *Am. J. Hum. Genet.* **106**, 327-337 (2020).
<https://doi.org/10.1016/j.ajhg.2020.01.017>
- 4 Revez, J. A. *et al.* Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nat Commun* **11**, 1647 (2020).
<https://doi.org/10.1038/s41467-020-15421-7>
- 5 Braithwaite, V. S., Jones, K. S., Schoenmakers, I., Silver, M., Prentice, A. & Hennig, B. J. Vitamin D binding protein genotype is associated with plasma 25OHD concentration in West African children. *Bone* **74**, 166-170 (2015).
<https://doi.org/10.1016/j.bone.2014.12.068>
- 6 Mogire, R. M. *et al.* Prevalence and predictors of vitamin D deficiency in young African children. *BMC Med.* **19**, 1-14 (2021). <https://doi.org/10.1186/s12916-021-01985-8>
- 7 Parlato, L. A. *et al.* Genome-wide association study (GWAS) of circulating vitamin D outcomes among individuals of African ancestry. *Am. J. Clin. Nutr.* **117**, 308-316 (2023). <https://doi.org/10.1016/j.ajcnut.2022.12.001>
- 8 O'Brien, K. M., Sandler, D. P., Shi, M., Harmon, Q. E., Taylor, J. A. & Weinberg, C. R. Genome-Wide Association Study of Serum 25-Hydroxyvitamin D in US Women. *Front Genet* **9**, 67 (2018). <https://doi.org/10.3389/fgene.2018.00067>
- 9 Hong, J. *et al.* Transethnic Evaluation Identifies Low-Frequency Loci Associated With 25-Hydroxyvitamin D Concentrations. *J. Clin. Endocrinol. Metab.* **103**, 1380-1392 (2018). <https://doi.org/10.1210/jc.2017-01802>
- 10 Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D. & Mountain, J. L. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am. J. Hum. Genet.* **96**, 37-53 (2015).
<https://doi.org/10.1016/j.ajhg.2014.11.010>
- 11 Choudhury, A. *et al.* High-depth African genomes inform human migration and health. *Nature* **586**, 741-748 (2020). <https://doi.org/10.1038/s41586-020-2859-7>
- 12 Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26-31 (2019). <https://doi.org/10.1016/j.cell.2019.02.048>
- 13 Bick, A. G. *et al.* Genomic data in the All of Us Research Program. *Nature* **627**, 340-346 (2024). <https://doi.org/10.1038/s41586-023-06957-x>
- 14 Raychaudhuri, S. & Prinz, W. A. The diverse functions of oxysterol-binding proteins. *Annu. Rev. Cell Dev. Biol.* **26**, 157-177 (2010).
<https://doi.org/10.1146/annurev.cellbio.042308.113334>
- 15 Wang, H. *et al.* ORP2 delivers cholesterol to the plasma membrane in exchange for phosphatidylinositol 4, 5-bisphosphate (PI (4, 5) P2). *Mol. Cell* **73**, 458-473 (2019).
- 16 Kobayashi, J. *et al.* Ligand recognition by the lipid transfer domain of human OSBP is important for enterovirus replication. *ACS infectious diseases* **8**, 1161-1170 (2022).
- 17 Roberts, B. L. *et al.* Differing activities of oxysterol-binding protein (OSBP) targeting anti-viral compounds. *Antiviral Res.* **170**, 104548 (2019).

1084 18 Severance, Z. C. *et al.* Structure–Activity Relationships of Ligand Binding to
1085 Oxysterol-Binding Protein (OSBP) and OSBP-Related Protein 4. *J. Med. Chem.* **66**,
1086 3866–3875 (2023).
1087 19 Burgett, A. W. G. *et al.* Natural products reveal cancer cell dependence on oxysterol-
1088 binding proteins. *Nat. Chem. Biol.* **7**, 639–647 (2011).
1089 20 Diercks, A. H. *et al.* Oxysterol binding protein regulates the resolution of TLR-
1090 induced cytokine production in macrophages. *Proceedings of the National Academy*
1091 *of Sciences* **121**, e2406492121 (2024).
1092 21 Malaria, G. E. N. P. f. C. P. Genomic epidemiology of artemisinin resistant malaria.
1093 *elife* **5**, e08714 (2016).
1094 22 Schurz, H. *et al.* Multi-ancestry meta-analysis of host genetic susceptibility to
1095 tuberculosis identifies shared genetic architecture. *Elife* **13**, e84394 (2024).
1096 23 Bikle, D. D. & Schwartz, J. Vitamin D Binding Protein, Total and Free Vitamin D
1097 Levels in Different Physiological and Pathophysiological Conditions. *Front.*
1098 *Endocrinol. (Lausanne)* **10**, 317 (2019). <https://doi.org/10.3389/fendo.2019.00317>
1099 24 Palmer, N. D. *et al.* Genome-wide association study of vitamin D concentrations and
1100 bone mineral density in the African American-Diabetes Heart Study. *PLoS One* **16**,
1101 e0251423 (2021). <https://doi.org/10.1371/journal.pone.0251423>
1102 25 Engelman, C. D. *et al.* Genome-wide association study of vitamin D concentrations in
1103 Hispanic Americans: the IRAS family study. *J. Steroid Biochem. Mol. Biol.* **122**, 186-
1104 192 (2010). <https://doi.org/10.1016/j.jsbmb.2010.06.013>
1105 26 Kim, Y. A. *et al.* Unveiling Genetic Variants Underlying Vitamin D Deficiency in
1106 Multiple Korean Cohorts by a Genome-Wide Association Study. *Endocrinol Metab*
1107 *(Seoul)* **36**, 1189–1200 (2021). <https://doi.org/10.3803/EnM.2021.1241>
1108 27 Traglia, M. *et al.* Genetic contributions to maternal and neonatal vitamin D levels.
1109 *Genetics* **214**, 1091–1102 (2020).
1110 28 Bikle, Daniel D. Vitamin D Metabolism, Mechanism of Action, and Clinical
1111 Applications. *Chem. Biol.* **21**, 319–329 (2014).
1112 <https://doi.org/10.1016/j.chembiol.2013.12.016>
1113 29 Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161-
1114 164 (2016). <https://doi.org/10.1038/538161a>
1115 30 Adachi, R. *et al.* Selective activation of vitamin D receptor by lithocholic acid acetate,
1116 a bile acid derivative. *J. Lipid Res.* **46**, 46–57 (2005).
1117 [https://doi.org:https://doi.org/10.1194/jlr.M400294-JLR200](https://doi.org/https://doi.org/10.1194/jlr.M400294-JLR200)
1118 31 Samadi, A. *et al.* A Comprehensive Review on Oxysterols and Related Diseases.
1119 *Curr. Med. Chem.* **28**, 110–136 (2021).
1120 <https://doi.org/10.2174/0929867327666200316142659>
1121 32 Zhou, Y. *et al.* OSBP-related proteins (ORPs) in human adipose depots and cultured
1122 adipocytes: evidence for impacts on the adipocyte phenotype. *PLoS One* **7**, e45352
1123 (2012). <https://doi.org/10.1371/journal.pone.0045352>
1124 33 El-Sharkawy, A. & Malki, A. Vitamin D signaling in inflammation and cancer:
1125 Molecular mechanisms and therapeutic implications. *Molecules* **25**, 3219 (2020).
1126 34 Pereira-Santos, M., Costa, P. R. & Assis, A. M.
1127 35 Chen, L., Dong, Y., Chen, J., Huang, Y. & Zhu, H. Epigenetics Predicts Serum 25-
1128 Hydroxyvitamin D Response to Vitamin D(3) Supplementation in African Americans.
1129 *Mol. Nutr. Food Res.* **64**, e1900738 (2020). <https://doi.org/10.1002/mnfr.201900738>
1130 36 Bouchard, L. *et al.* Association of OSBP11 gene polymorphisms with
1131 cardiovascular disease risk factors in obesity. *Obesity* **17**, 1466–1472 (2009).
1132 37 Ramsay, M. *et al.* H3Africa AWI-Gen Collaborative Centre: a resource to study the
1133 interplay between genomic and environmental risk factors for cardiometabolic

diseases in four sub-Saharan African countries. *Global health, epidemiology and genomics* **1**, e20 (2016).

38 Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675-679 (2021).

39 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
<https://doi.org/10.1371/journal.pcbi.1004219>

40 Bejon, P. *et al.* Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS Med.* **7**, e1000304 (2010).
<https://doi.org/10.1371/journal.pmed.1000304>

41 Mentzer, A. J. *et al.* High-resolution African HLA resource uncovers HLA-DRB1 expression effects underlying vaccine response. *Nat. Med.* **30**, 1384-1394 (2024).
<https://doi.org/10.1038/s41591-024-02944-5>

42 Tiono, A. B. *et al.* First field efficacy trial of the ChAd63 MVA ME-TRAP vectored malaria vaccine candidate in 5-17 months old infants and children. *PLoS One* **13**, e0208328 (2018). <https://doi.org/10.1371/journal.pone.0208328>

43 Atkinson, S. H. *et al.* Seasonal childhood anaemia in West Africa is associated with the haptoglobin 2-2 genotype. *PLoS Med.* **3**, e172 (2006).
<https://doi.org/10.1371/journal.pmed.0030172>

44 Hutchinson, K., Healy, M., Crowley, V., Louw, M. & Rochev, Y. Verification of Abbott 25-OH-vitamin D assay on the architect system. *Practical Laboratory Medicine* **7**, 27-35 (2017). <https://doi.org/10.1016/j.plabm.2017.01.001>

45 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76-82 (2011).

46 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 (2012). <https://doi.org/10.1038/ng.2213>

47 Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *The American Journal of Human Genetics* **88**, 586-598 (2011).

48 Gauderman, W. J. Sample size requirements for association studies of gene-gene interaction. *Am. J. Epidemiol.* **155**, 478-484 (2002).

49 Galvão, A. A. *et al.* Genetic polymorphisms in vitamin D pathway influence 25(OH)D levels and are associated with atopy and asthma. *Allergy Asthma Clin. Immunol.* **16**, 62 (2020). <https://doi.org/10.1186/s13223-020-00460-y>

50 Wilson, J. G. *et al.* Study design for genetic analysis in the Jackson Heart Study. *Ethn. Dis.* **15**, 30-37 (2005).

51 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).

52 All of Us Research Program, I. The “All of Us” research program. *N. Engl. J. Med.* **381**, 668-676 (2019).

53 Smith, J. G. *et al.* Genome-wide association studies of the PR interval in African Americans. *PLoS genetics* **7**, e1001304 (2011).

54 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8** (2017).
<https://doi.org/10.1038/s41467-017-01261-5>

55 Consortium, G. T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).

1184 56 Schmiedel, B. J. *et al.* Impact of genetic polymorphisms on human immune cell gene
1185 expression. *Cell* **175**, 1701-1715 (2018).

1186 57 Van Der Wijst, M. G. P. *et al.* Single-cell RNA sequencing identifies celltype-specific
1187 cis-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493-497 (2018).

1188 58 Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative
1189 trait loci in whole blood. *Nat. Genet.* **49**, 139-145 (2017).

1190 59 Consortium, G. T. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis:
1191 multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).

1192 60 Benner, C., Spencer, C. C. A., Havulinna, A. S., Salomaa, V., Ripatti, S. & Pirinen,
1193 M. FINEMAP: efficient variable selection using summary data from genome-wide
1194 association studies. *Bioinformatics* **32**, 1493-1501 (2016).

1195 <https://doi.org/10.1093/bioinformatics/btw018>

1196 61 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
1197 interpreting genome-wide expression profiles. *Proceedings of the National Academy*
1198 *of Sciences* **102**, 15545-15550 (2005).

1199 62 Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. &
1200 Mesirov, J. P. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-
1201 1740 (2011).

1202 63 Ramos, E. *et al.* Replication of genome-wide association studies (GWAS) loci for
1203 fasting plasma glucose in African-Americans. *Diabetologia* **54**, 783-788 (2011).

1204 64 Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M. & Bladé, I. The
1205 effective number of spatial degrees of freedom of a time-varying field. *Journal of*
1206 *climate* **12**, 1990-2009 (1999).

1207 65 Dallakyan, S. & Olson, A. J. Small-molecule library screening by docking with PyRx.
1208 *Chemical biology: methods and protocols*, 243-250 (2015).

1209 66 Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of
1210 docking with a new scoring function, efficient optimization, and multithreading. *J.*
1211 *Comput. Chem.* **31**, 455-461 (2010).

1212

1213