The Role of Common and Rare Genetic Variation in Vitamin D Status and Clinical Implications of Vitamin D Genomics

Despoina Manousaki

Department of Human Genetics

Faculty of Medicine

McGill University, Montréal

December 2019

A thesis submitted to McGill University in partial fulfillment

of the requirements of the degree of Doctor of Philosophy

Abstract/ Résumé

Abstract

Vitamin D insufficiency affects as many as 40% of otherwise healthy adults and has been associated to multiple adverse health outcomes, such as cancer, cardiovascular disease, autoimmune disorders and bone disease. Personal, social, and cultural factors are important determinants of vitamin D availability via their effects on sun exposure and diet. However, only about a quarter of the variability in the biomarker of vitamin D in humans, 25 hydroxyvitamin D (25OHD), is attributable to these known environmental factors, while half of its variability has been attributed to genetics. However, until 2017, the known common genetic determinants of 25OHD only explained 2.4% of its variability. We performed genome-wide association studies (GWAS), aiming to identify novel common (minor allele frequency (MAF) >5%), or low-frequency/rare genetic determinants (MAF <5%) of large effect on 25OHD. To do this, we first undertook a GWAS meta-analysis on 42,326 individuals with available whole genome sequencing data or imputed genotypes from 19 European cohorts. We identified a low-frequency variant in the 25 hydroxylase gene (*CYP2R1*), affecting 1 out of 20 Europeans, conferring the largest effect on 25OHD levels described to date. We showed that carrying one copy of the effect allele of this variant doubles the risk of vitamin D insufficiency and increases by 40% the risk of multiple sclerosis, a disease with a well-established association with low 25OHD levels. Next, with the release of 25OHD data from 401,460 White British participants in UK Biobank, we performed the largest to date GWAS on 25OHD levels, and combined the results of this GWAS to our previous GWAS in a fixed-effects meta-analysis. We therefore identified 69 25OHD-related loci, among which 63 were novel. These 69 loci harbor 138 conditionally independent variants, among which 53 are low-frequency or rare. Hence, the genomic heritability of 25OHD levels was estimated to

16.1%. By leveraging results of these GWAS, we interrogated the causality of 25OHD levels in coronary artery disease, asthma, atopic dermatitis and elevated IgE levels, using a Mendelian randomization study design, and generated results with direct translational impact. We therefore present the findings in this thesis as novel contributions to the genetic determinants of vitamin D status in Europeans, and to the translational research seeking to determine the causal role of vitamin D on four human diseases/traits.

Résumé

L'insuffisance en vitamine D affecte jusqu'à 40% des adultes en bonne santé et a été associée à de nombreux effets néfastes sur la santé, tels que le cancer, la maladie cardiovasculaire, les troubles auto-immuns et la maladie osseuse. Les facteurs personnels, sociaux et culturels sont des déterminants importants de la disponibilité de la vitamine D par leurs effets sur l'exposition au soleil et le régime alimentaire. Cependant, seulement un quart de la variabilité du biomarqueur de la vitamine D chez l'homme, la 25 hydroxyvitamine D (25OHD), est attribuable à ces facteurs environnementaux connus, tandis que la moitié de sa variabilité a été attribuée à la génétique. Néanmoins, les déterminants génétiques communs connus de 25OHD jusqu'en 2017 n'expliquaient que 2,4% de sa variabilité. Nous avons effectué des études d'association pangénomique (GWAS) pour identifier des nouveaux déterminants génétiques fréquents ou des variantes de basse fréquence ou rares ayant un large impact sur les taux de 25OHD. Pour le faire, nous avons d'abord combiné des données sur 42,326 individus séquencés ou imputés au génome entier provenant de 19 cohortes Européennes. Nous avons identifié une variante de basse fréquence dans le gène de la 25 hydroxylase (*CYP2R1*), affectant un Européen sur 20, conférant le plus grand effet sur les taux de 25OHD décrit à ce jour. Nous avons montré que le fait de porter une copie de l'allèle à effet de cette variante double le risque d'insuffisance en vitamine D et augmente de 40% le risque de sclérose en plaques, une maladie avec une association bien établie avec des taux bas de 25OHD. Ensuite, avec la relâche des données sur les taux de 25OHD des 401,460 participants blancs britanniques dans UK Biobank, nous avons effectué la plus large GWAS à ce jour sur les taux de 25OHD, et nous avons combiné ces résultats à ceux de notre GWAS précédente, dans une méta-analyse d'effets fixes. Nous avons identifié 69 loci associés aux taux de 25OHD, dont 63 sont nouveaux. Ces 69 loci incluent 138 variants independents, dont 53 sont de basse fréquence ou

rares. Dorénavant, l'héritabilité génomique de 25OHD a été estimée à 16.1%. En exploitant les résultats de nos études GWAS, nous avons interrogé le lien de causalité des taux de vitamine D dans la maladie coronarienne, l'asthme, la dermatite atopique et les taux enlevés de IgE, à l'aide de la randomisation Mendélienne et nous avons généré des résultats ayant un impact translationnel direct. Nous présentons donc les résultats de cette thèse comme de nouvelles contributions aux déterminants génétiques du statut en vitamine D chez les Européens, ainsi qu'aux travaux de recherche translationelle visant à déterminer le rôle causal de la vitamine D sur 4 maladies/traits humains.

Table of Contents

List of Abbreviations

25OHD: 25 Hydroxyvitamin D

1,25[OH]$_2$D: 1,25 Dihydroxyvitamin D

ALSPAC: Avon Longitudinal Study of Parents and Children

BMDCS: Bone Mineral Density in Childhood Study

BMI: Body Mass Index

BPROOF: B-Vitamins for the PRvention Of Osteoporotic Fractures

BWA: Burrows-Wheeler Aligner

CAD: Coronary Artery Disease

CaMOS: Canadian Multicentre Osteoporosis Study

CARDIoGRAM: Coronary ARtery DIsease Genome-wide Replication and Meta-analysis

CHS: Cardiovascular Health Study

CI: Confidence Interval

CIHR: Canadian Institutes of Health Research

CYP24A1: Cytochrome P450 Family 24 Subfamily A Member 1

CYP27B1: Cytochrome P450 Family 27 Subfamily B Member 1

CYP2R1: Cytochrome P450 Family 2 Subfamily R Member 1

DBP: Diastolic Heart Pressure

DHCR7: 7-Dehydrocholesterol Reductase

DIAGRAM: DIabetes Genetics Replication And Meta-analysis

DNA: Deoxyribonucleic Acid

DXA: Dual-Energy X-ray absorptiometry

EA: Effect Allele

EAF: Effect Allele Frequency

EAGLE: EArly Genetics and Lifecourse Epidemiology

FEV1: Forced Expiratory Volume in 1second

FHS: Framingham Heart Study

GC: GC vitamin D binding protein

GCTA: Genome-wide Complex Trait Analysis

GCTA-COJO: Conditional and Joint Genome-wide Complex Trait Analysis

GEMIN2: Gem Nuclear Organelle Associated Protein 2

GIANT: Genetic Investigation of Anthropomorphic Traits

GOOD: Gothenburg Osteoporosis and Obesity Determinants

GWAMA: Genome-Wide Association Meta-Analysis

GWAS: Genome-Wide Association Study

HAL: Histidine Ammonia-Lyase

HRC: Haplotype Reference Consortium

HWE: Hardy-Weinberg Equilibrium

ICBP: International Consortium for Blood Pressure

IgE: Immunoglobulin E

JDRF: Juvenile Diabetes Research Foundation

LD: Linkage Disequilibrium

LDSC: LD score regression

LMM: Linear mixed model

MAF: Minor allele frequency

MR: Mendelian randomization

MrOS: Osteoporotic Fractures in Men USA

MROS GBG: Osteoporotic Fractures in Men Gothenburg

MROS MALMO: Osteoporotic Fractures in Men Malmo

MS: Multiple Sclerosis

NEO: Netherlands Epidemiology of Obesity

OR: Odds Ratio

PTH: Parathyroid Hormone

QQ-plot: Quantile-quantile plot

PDE3B: Phosphodiesterase 3B

PGC: Psychiatric Genetics Consortium

PIVUS: Prospective Investigation of the Vasculature in Upssala Seniors

RCT: Randomized Controlled Trial

REML: Residual maximum likelihood

RS: Rotterdam Study

SE: Standard Error

SEC23A: Sec23 homolog A, coat protein complex II component

SNP: Single Nucleotide Polymorphism

SNV: Single Nucleotide Variant

TUK: TwinsUK

ULSAM: Upssala Longitudinal Study of Adult Men

URL: Universal Resource Locator

VDR: Vitamin D Receptor

VDBP: Vitamin D Binding Protein

WGS: Whole Genome Sequencing

List of Figures

List of Tables

Table 3: Sensitivity Analyses Testing MR Assumptions.

Acknowledgments

This doctoral thesis would not have been possible without the endless support of some exceptional people, and I will now thank them in gratuitous detail.

I would like to begin by thanking my supervisor, Dr. Brent Richards. Thank you, Brent, for being my mentor, and for providing me with a space in which to grow into the researcher I am today. I came to McGill as a Master's student with very narrow exposure to research, very few publications, and very limited background in Human Genetics and Epidemiology. I had never thought of pursuing a career of clinician scientist, but your mentorship and your personal example convinced me. I did not know if I would pursue to a PhD, but you counseled me to fast-track. Today, I am so glad that you did, as I am leaving McGill with what feels like a plethora of riches. I completed successfully my graduate courses, published over a dozen manuscripts, won more awards than I ever thought before, and most importantly, started developing skills to become an independent PI. Your supervision has thus positioned me for success in academia. This wouldn't be possible without you, so thank you.

Next, I would like to thank Dr. Vincenzo Forgetta, the bioinformatics wizard in the lab. Vince has been always my bigger helper whenever I was in an impasse in my analyses, with whom I have spent countless hours discussing research and a lot that had nothing to do with research. Vince, without your help and support, I wouldn't be sitting here writing these acknowledgments. You have been an invaluable mentor during my graduate studies, so thank you.

I would like to thank my supervisory committee: Drs. Goltzman, Shurr and Thanassoulis, but also Dr. Greenwood and Dr. Polychronakos, whose door was always open for discussion. They all have been important sources of knowledge for my research.

During my five years at McGill, I have come to know several members of our lab. Aside from Vince, there have been many that I would like to thank: Lauren, Agustin, Julia, Stephanie, Sirui, Adil, Tricia, Laetitia, Julyan, Tomoko, Haoyu. Thank you all for becoming a part of my life and for helping me along the way.

To the people closest to me, thank you. To my husband George, and my children, Joseph and Emmanuel, for their endless support, for inspiring me and for influencing me in different but incredible ways. To my parents, for always challenging me to pursue my studies. To my brother George, a neurologist and assistant professor, for discussing on my research.

Finally, I would like to thank my sources of funding, without which none of this work would have been possible. The Réseau de médecine génétique appliquée, the Canadian Pediatric Endocrinology Group, and Juvenile Diabetes Research Funding, have all provided me with stipend funding. The American Society for Human Genetics, European Society of Pediatric Endocrinology, and various faculties, divisions, or departments of McGill have all provided me with awards. Each award was accepted with humility and gratitude, and I thank them.

Format of the Thesis

The work described here was performed under the supervision of Dr. Brent Richards. It is a manuscript-based thesis format as described in the Thesis Preparation Guidelines by the Department of Graduate and Postdoctoral Studies. This thesis contains six chapters. Chapter 1 is an introduction to this thesis. Chapters 2, 3 and 4 have been published in the American Journal of Human Genetic, Circulation Cardiovascular Genetics and PloS Medicine, respectively. Chapter 5 is a manuscript which is under review at the American Journal of Human Genetics. Chapter 6 is a discussion of Chapters 2 to 5. Chapter 7 is a conclusion with future aims for Chapters 2, to 5. A summary of other publications can be found in the Appendix.

Contribution of Authors

Chapter 2 is a manuscript authored by *Despoina Manousaki, Tom Dudding , Simon Haworth, Yi-Hsiang Hsu, Ching-Ti Liu, Carolina Medina-Gómez Trudy Voortman, Nathalie van der Velde, Håkan Melhus, Cassiane Robinson-Cohen, Diana L. Cousminer Maria Nethander,, Liesbeth Vandenput, Raymond Noordam, Vincenzo Forgetta, Celia MT Greenwood, Mary L. Biggs, Bruce M. Psaty, Jerome I. Rotter, Babette S. Zemel, Jonathan A. Mitchell, Bruce Taylor, Mattias Lorentzon, Magnus Karlsson, Vincent V.W. Jaddoe, Henning Tiemeier, Natalia Campos-Obando, Oscar H.Franco, Andre G. Utterlinden, Linda Broer, Natasja M. van Schoor, Annelies C. Ham, M. Arfan Ikram, David Karasik, Renée de Mutsert, Frits R. Rosendaal, Martin den Heijer ,Thomas J. Wang, Lars Lind, Eric S.Orwoll, Dennis O. Mook-Kanamori, Karl Michaëlsson, Bryan Kestenbaum, Claes Ohlsson, Dan Mellström, Lisette CPGM de Groot, Struan F.A. Grant,Douglas P. Kiel, M. Carola Zillikens, Fernando Rivadeneira, Stephen Sawcer , Nicholas J Timpson and J. Brent Richards.* It was published in the American Journal of Human Genetics on August 3rd, 2017. BMP, JIR, BSZ, JAM, BT, ML, MK, TJW, LL, ESO, RN, DOMK, KM, BK, CO, DMe, NvdV, LCPGMdG, MAI, HT, VVWJ, OF, AU, SFAG, DPK, MCZ, FR, SS, NJT, JBR were principal investigators of the studies which participated in the GWAS meta-analysis. TD, SH, Y-HH, C-TL, CM-G, TV, HM, CR-C, DLC, MN, LV, VF, CMTG, MLB, NC-O, LB, NMvS, AH, RdM, FRR, MdH and I undertook the GWAS analyses, while VF and I were the lead analysts, who undertook the meta analysis. All authors revised and reviewed the paper. Specifically, as first author of this paper, I collected data from the individual cohorts, performed the GWAS meta-analysis, which identified the *CYP2R1* low-frequency variant, and undertook the follow-up analyses. I wrote the abstract, introduction, methods, results and discussion, and all relevant tables and figures, both main and supplementary.

Chapter 3 is a manuscript authored by *Despoina Manousaki, Lauren E. Mokry, Stephanie Ross, David Goltzman and J. Brent Richards*. It was published in Circulation: Cardiovascular Genetics on August 9th 2016. JBR conceived the experiment, with input from DG. JBR. LEM, SR & I conducted the analyses. JBR and I wrote the first draft of the manuscript. All authors revised and reviewed the paper. Specifically, I performed the Mendelian randomization analysis, described the results. I wrote the manuscript, specifically the abstract, introduction, the above listed methods and results, the discussion, and all relevant tables and figures, both main and supplementary.

Chapter 4 contains a manuscript authored by *Despoina Manousaki, Lavinia Paternoster, Marie Standl, Miriam F. Moffatt, Martin Farrall, Emmanuelle Bouzigon, David P Strachan, Florence Demenais, Mark Lathrop, William O.C.M. Cookson and J. Brent Richards.* Mendelian randomization shows no role for low vitamin D levels in asthma, atopic dermatitis or IgE levels. It was published in PLoS Med. on May 9 2017. J JBR and WOCMC conceived the experiment. JBR & DM conducted the analyses. JBR and I wrote the first draft of the manuscript. MFM, BC, FD, EB, LP, MS, WOCMC, MF, DS and ML reviewed the manuscript. Similar to Chapter 3, I performed the Mendelian randomization analysis, described the results. I wrote the manuscript, specifically the abstract, introduction, the methods and results, the discussion, all relevant tables and figures, both main and supplementary.

Finally, chapter 5 contains an unpublished manuscript titled: "Genome-wide association study for vitamin D levels reveals 63 novel loci", authored by *Despoina Manousaki, Ruth Mitchell, Tom Dudding, Simon Haworth, Adil Harroud, Vince Forgetta, Rupal L. Shah, Jian'an Luan, Claudia Langenberg, Nicholas J. Timpson and J. Brent Richards.* JBR and NJT conceived the experiment, and I, RM, TD, SH undertook the analyses. All authors revised the manuscript. RM and I were the main analysts, performed the GWAS in UK Biobank, the meta-analysis with the

previous GWAS and *in silico* follow-up, as well as the first draft of the manuscript. CL, RS and JL undertook the 1,25 dihydroxyvitamin D GWAS and the look-up for the cojo-independent 25OHD SNPs.

Original Contribution to Knowledge

This doctoral thesis identified novel genetic determinants of 25OHD levels through two large GWAS meta-analyses and used information from 25OHD GWAS to explore if low 25OHD levels are causally associated with clinical outcomes in Europeans.

Chapter 2 is titled "Low Frequency Coding Variation in CYP2R1 has Large Effects on Vitamin D Level and Risk of Multiple Sclerosis". It describes how we analyzed data from 42,326 participants with either genome-wide genotypes or whole-genome sequencing data and available 25OHD to identify a low-frequency variant with large effect on a known locus (*CYP2R1*) and 2 novel loci involving common variants.

Chapter 3 is titled "Mendelian Randomization Studies do not Support a Role for Vitamin D in Coronary Artery Disease", and Chapter 4 is titled "Mendelian randomization shows no role for low vitamin D levels in asthma, atopic dermatitis or IgE levels". These 2 Chapters include 2 distinct Mendelian randomization papers, testing the causal role of vitamin D on coronary artery disease, and asthma, atopic dermatitis and IgE levels respectively.

Chapter 5 is titled "Genome-wide association study for vitamin D levels reveals 63 novel loci". It describes how we performed the largest to date GWAS meta-analysis on 25OHD levels, by undertaking a GWAS on 401,460 White British individuals from UK Biobank, and combining these results to our previous GWAS described in Chapter 2. In the same paper, we undertook an *in silico* functional follow-up analysis using summary-level results of this GWAS, to identify enrichments in gene expression, gene pathways, and genetic correlations between 25OHD levels and other GWAS traits and diseases.

Chapter 6 is titled: "Conclusions and future steps in vitamin D genomics". It includes the description of a project for a pharmacogenetics trial, exploring the response to vitamin D supplementation in carriers of the low-frequency variant in *CYP2R1* identified by the GWAS

described in Chapter 2. It also describes ideas for other future projects, which will expand our knowledge on the genetic architecture of vitamin D in humans.

Chapter 1: General introduction

1.1 Vitamin D levels and their genetics

Vitamin D insufficiency affects as many as 40% of otherwise healthy adults in developed countries[1]. The musculoskeletal consequences of inadequate vitamin D concentrations are well established, and include childhood rickets, osteomalacia, and fractures[2]. A growing number of other extra-skeletal disorders have also been linked to vitamin D insufficiency, and include autoimmune disorders, increased risk of falls, and several cancers[2]. Results of a 2007 meta-analysis suggested that vitamin D supplementation substantially reduced mortality in adults[3]. Personal, social, and cultural factors are important determinants of vitamin D availability via their effects on sun exposure and diet. Sufficient exposure to ultraviolet light or adequate intake from diet or supplements is needed to maintain vitamin D status. Concentrations of the widely accepted biomarker for vitamin D status, 25-hydroxyvitamin D (25OHD), are highest in the summer and lowest in the winter in northern latitudes. However, only about a quarter of the inter-individual variability in 25OHD concentration is attributable to season of measurement, geographical latitude, body mass index, age, sex and reported vitamin D intake[4; 5]. Moreover, results of previous twin and family studies suggest that genetic factors contribute substantially to this variability, with estimates of heritability as high as 53%[4; 6].

25OHD is a steroid pro-hormone and a fat-soluble metabolite of cholecalciferol, which is predominately synthesized in the skin from 7-dehydrocholesterol after exposure to ultra-violet light or obtained from dietary sources including fortified foods, supplements and oily fish. Cholecalciferol is hydroxylated first to 25OHD in the liver, then to the hormonal form 1,25-dihydroxyvitamin D 1,25$[OH]_2$D in the kidney. CYP2R1 is the most important liver 25-hydroxylase; CYP27B1 is the key renal 1-hydroxylase. Both 25OHD and 1,25$[OH]_2$D are

catabolized by CYP24A1. 1,25$[OH]_2$D is the ligand for the vitamin D receptor (VDR), a transcription factor, binding to sites in the DNA called vitamin D response elements (VDREs). There are thousands of these binding sites regulating hundreds of genes in a cell-specific fashion, among which genes influencing the calcium and phosphorus metabolism, cell proliferation, differentiation, apoptosis and immune modulation[2]. Understanding the genetic etiology of low vitamin D levels could have important public health implications by prioritizing individuals who would benefit from supplementation.

Although several rare Mendelian disorders cause functional vitamin D insufficiency and, along with candidate gene studies, have pointed to specific vitamin D pathway genes, our knowledge on the common genetic determinants of 25OHD levels expanded in the genome-wide association study (GWAS) era. In recent years, multiple GWAS of circulating levels of serum 25OHD have been conducted on participants of Europeans ancestry [7-9].The first large GWAS on 25OHD levels was performed in 2010 by the SUNLIGHT consortium,[7] which was co-led by Dr. Brent Richards. This multicenter GWAS meta-analysis on 33,996 individuals of European ancestry demonstrated four single nucleotide polymorphisms (SNPs) affecting 25OHD levels. This study was designed to identify only common variants (ie, those variants having a minor allele frequency (MAF) of > 5%) and explained only 2.4% of the variability of 25OHD levels. In 2018, the largest to date GWAS by Jiang et al [9], comprised of 79,366 individuals, identified 2 additional common genetic variants, and estimated the heritability of 25OHD from the entire genome to 7.5%, which is still far from the ~50% estimate reported in twin studies. In summary, the known to date vitamin D variants map in or near genes having an established role in vitamin D synthesis (*DHCR7/NADSYN1* (rs12785878) and *CYP2R1* (rs10741657)), transportation (*GC* (rs2282679)) and degradation (*CYP24A1* (rs17216707)), as well as outside of the vitamin D metabolism pathway: *SEC23A* (Sec23 homolog A, coat protein

complex II component, rs8018720) involved in endoplasmic reticulum (ER)-Golgi protein trafficking, and *AMDHD1* (amidohydrolase domain containing 1, rs10745742) an enzyme involved in the histidine, lysine, phenylalanine, tyrosine, proline and tryptophan catabolic pathway. This suggests that genes outside the vitamin D metabolism pathway may contribute to the genetic regulation of serum 25OHD homeostasis, and these genes can be identified by increasing with larger GWAS discovery sample sizes. Moreover, recently, low frequency and rare genetic variants (MAF of ≤5% and ≤1%, respectively) of large effect have been identified for biomedically relevant traits providing an opportunity to better understand the biologic mechanisms influencing disease susceptibility in the general population[10; 11]. This suggests that not yet identified low-frequency or rare genetic variants of large effects in known or novel loci can exert large effects on 25OHD levels. These variants can now be identified with larger GWAS sample sizes, the use of whole genome sequencing, and better imputed genotypes, for instance by using the combined 1000 Genomes/UK10K panel[12], the Haplotype Reference Consortium Panel (HRC)[13], or the more recent combined UK10K/HRC panel).

1.2 Role of vitamin D levels in human disease and causal inference

Improved understanding of the genetic determinants of 25OHD has helped re-assess the role of Vitamin D in the etiology of complex diseases. Vitamin D deficiency has been associated with musculoskeletal disorders (childhood rickets, osteomalacia and fractures)[2], autoimmune disease (in particular multiple sclerosis[14; 15] and type 1 diabetes[16; 17]), and cancer[18]. There is also evidence that vitamin D deficiency influences infectious disease[19], cardiovascular disease[20] , neurodevelopmental[21; 22] and neurodegenerative conditions[23]. However, observed associations could be due to unmeasured confounders or reverse causality. Randomized control trials have investigated the benefits of vitamin D supplementation but with inconsistent conclusions[24; 25]. Additionally, large trials are expensive and have been rarely performed, as vitamin D supplementation is off-patent and thus offers little financial incentive. Epidemiological methods for causal inference such as Mendelian randomization (MR) may provide an alternative method to assess the likely benefits of vitamin D supplementation for these health outcomes. MR[26; 27] is an established technique that uses human genetics to ascertain whether a given biomarker, such as vitamin D, is implicated in disease etiology. This method relies on a simple tenet: if a biomarker is etiologically involved in disease process, then the genetic factors which influence the biomarker will influence disease risk. MR is not susceptible to reverse causation or confounding as disease state does not change DNA sequence, and genotypes are randomly assorted following meiosis (due to Mendel's 2nd law)[28], respectively. Thus, MR is comparable to a randomized controlled trial, where the random assortment of genetic variants replicates the random allocation of study participants to different therapeutic arms. It becomes obvious that a deeper understanding of the genetic determinants contributing to variation in circulating vitamin D levels will enable an improved instrumentation of vitamin D in MR studies, will allow better genomic prediction of vitamin D levels and will provide insights into biological mechanisms. Specific applications include MR

experiments investigating causal relationships between vitamin D and an outcome, and in particular has confirmed a protective effect of vitamin D against multiple sclerosis[14; 15]. The evidence from these studies partially led to changes in the 2017 guidelines of the Multiple sclerosis society of Canada to include vitamin D supplementation as a preventive strategy for high-risk individuals(https://mssociety.ca/library/document/Vka6RXcnOizNm9sIwuWvroxejlhLqTJ8/original.pdf). Since the two most recent 25OHD-associated genetic variants have been identified, more than fifteen MR studies have been published utilizing these additional genetic variants to give greater precision of estimation[29-44].

Despite this progress, as mentioned above, the SNPs identified to date continue to explain little of the 25OHD heritability (7.5%), and although the samples size of the most recent GWAS by Jiang et al [9] was at least double that of the previous GWASes, its yield in novel vitamin D loci was very limited. This suggests that vitamin D might be a metabolite with a simple oligo-genic architecture, that it is influenced by few genetic variants (which is a challenge to downstream MR analyses), or might have a more polygenic structure, but many variants with small effects remain to be identified in future much larger GWAS.

1.3 Rationale and objectives

In the light of all the above, this doctoral Thesis had three objectives:

First, to detect low-frequency/rare variants (MAF <5% and <1% respectively) associated with 25OHD levels. We *hypothesized* that these variants could be identified in novel or previously known genetic loci, and could have large effects on 25OHD levels. To address this objective, we undertook a large-scale meta-analysis of 19 genome-wide association studies (GWAS) with either whole-genome sequencing data or deeply imputed genotypes, totaling 42,326 Europeans. Thus, we aimed to characterize genetic variants strongly associated with vitamin

D in novel genes, and use "fine-mapping" to identify novel genetic variants of large effect in known vitamin D related genes. At the moment this study was published, it was an unprecedented resource in depth of imputation, and number of genetic variants analyzed. Therefore, Chapter 2 represents major advancements in the field of vitamin D genomics.

In Chapters 3 and 4 of the present thesis, we addressed the *second objective* of this Thesis, which was to elucidate the role of low vitamin D levels on common complex human diseases. We *hypothesized* that 25OHD levels could be associated with conditions, such as coronary artery disease, asthma, atopic dermatitis and IgE levels. To test this hypothesis, we applied a MR design, to generate results free from confounding and reverse causation, which are typical bias in observational epidemiology. Thus, we created results with a direct clinical impact, since this evidence from MR informs public health strategies on the use of vitamin D supplementation for prevention of these diseases.

Taking advantage of the release of data on 25OHD levels from UK Biobank in April 2019, *our third objective* was to expand the phenotypic variance of 25OHD explained by common and rare genetic variation, more deeply investigate the genetic architecture (oligogenic or polygenic) of vitamin D, and enhance our genetic instruments for vitamin D for future MR studies. Our *hypothesis* was that 25OHD levels might actually be more polygenic than previously thought, and that loci outside the ones directly related to vitamin D metabolism might be involved in their genetic control. In Chapter 5, we first undertook a GWAS of serum 25OHD levels in 401,460 White British individuals from UK Biobank. We then meta-analyzed the results of this GWAS to those of our previous GWAS (described in Chapter 1), achieving a sample size of 443,734 European individuals. This allowed us to replicate all 6 known

25OHD-associated loci and identify 63 novel loci. Chapter 5 represents thus an unprecedented resource in terms of sample size and number of SNPs analyzed, to study genetics of vitamin D, and to expand our knowledge on the genocopies of vitamin D status.

Finally, in Chapter 7, we describe a future project, which consists in a direct follow-up of the findings of Chapter 2. Specifically, our *future objective* is to undertake a pharmacogenetics study, where we will seek to validate the findings of Chapter 2, by testing the response of carriers of the novel low-frequency variant on *CYP2R1* to oral vitamin D supplementation. The *hypothesis* behind this study is that we expect that the change in 25OHD levels of carriers will be smaller than this of the non-carriers, after a period of vitamin D oral supplementation, since this variant affects the 25 hydroxylation of vitamin D, which is the step which converts dietary vitamin D to 25OHD. If successful, this project may also have direct clinical impact, since early identification and treatment of individuals carrying this variant (~5% of the general European population) with the active form of vitamin D (1,25 dihyxroxyvitamin D or calcitriol), will allow for prevention of vitamin D insufficiency and of all the negative health outcomes causally related with low 25OHD in this patient group.

In summary, the *scientific hypothesis* of this Thesis was twofold: First, that through large GWAS we could gain a better knowledge of the genetic architecture of vitamin D levels. Second, that by using the genetic determinants (common, low-frequency or rare) of vitamin D status we could test causal associations between vitamin D levels and human diseases.

Chapter 2: Low Frequency Synonymous Coding Variation in *CYP2R1* has Large Effects on Vitamin D Level and Risk of Multiple Sclerosis.

This chapter contains a manuscript published under the same title in the American Journal of Human Genetics in 2017.

Date of publication: Aug 3 2017; Epub date: Jul 27 2017

Volume 101(issue2), pages 227-238

doi: 10.1016/j.ajhg.2017.06.014.

PMID: 28757204

Preface: Bridge Between Chapter 1 and Chapter 2

Until 2017, our knowledge on common genetic variation affecting vitamin D levels was mainly the result of a large GWAS meta-analysis on 25OHD levels, published in 2010 in Lancet by the SUNLIGHT consortium, co-led by Dr. Brent Richards at McGill University. Given the previous success of our research group in collaborating with the SUNLIGHT consortium, we aimed to expand our knowledge on low-frequency and rare genetic variation of vitamin D, taking advantage of the recent availability of whole genome sequencing data from some of the SUNLIGHT consortium studies, and of recent advances in genomic imputation. We thus performed a GWAS on 25OHD levels, with the largest to date number of SNPs, using imputed and whole genome sequencing data of up to 42,326 participants, and focusing on discovery of low-frequency and rare genetic variants with large effects on 25OHD levels, in novel or previously known vitamin D loci.

Title page

Title: Low Frequency Synonymous Coding Variation in CYP2R1 has Large Effects on Vitamin D Level and Risk of Multiple Sclerosis.

Short title: Low-frequency Variant Confers Large Effect on Vitamin D levels

Authors: Despoina Manousaki[1,2,47], Tom Dudding[3,47], Simon Haworth[3,47], Yi-Hsiang Hsu [4,5,6,47], Ching-Ti Liu[7,47], Carolina Medina-Gómez [8,9,10,47] Trudy Voortman[9,10,47], Nathalie van der Velde[8,11,47], Håkan Melhus[12,47], Cassianne Robinson-Cohen[13,47], Diana L. Cousminer [14,15,47] Maria Nethander[16,17,47], Liesbeth Vandenput[16,47] Raymond Noordam[18,47], Vincenzo Forgetta[1,2], Celia MT Greenwood[1,2,19,20] Mary L. Biggs[21], Bruce M. Psaty[22,23], Jerome I. Rotter[24], Babette S. Zemel[25,26], Jonathan A. Mitchell[25,26], Bruce Taylor[27], Mattias Lorentzon[16,28,29], Magnus Karlsson[30] , Vincent V.W. Jaddoe [9,10] , Henning Tiemeier[9,10,31], Natalia Campos-Obando[8], Oscar H.Franco[10], Andre G. Utterlinden[8,9,10], Linda Broer[8], Natasja M. van Schoor[32], Annelies C. Ham[8], M. Arfan Ikram[10,33], David Karasik[4], Renée de Mutsert[34], Frits R. Rosendaal[34], Martin den Heijer[35] ,Thomas J. Wang[36], Lars Lind[12,48], Eric S.Orwoll [37,38,48], Dennis O. Mook-Kanamori[34,39,48], Karl Michaëlsson[40,48], Bryan Kestenbaum [13,48], Claes Ohlsson[16,48], Dan Mellström[16,28,48] , Lisette CPGM de Groot[41,48], Struan F.A. Grant[14, 25, 42,48] ,Douglas P. Kiel[4,5,6,43,48], M. Carola Zillikens[8,48], Fernando Rivadeneira[8,9,10,48], Stephen Sawcer[44,48] , Nicholas J Timpson[3,48] and J. Brent Richards[1,2,45,46,48]

[1]Department of Human Genetics, McGill University, Montreal, QC H3A 1B1, Canada

[2]Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, QC H3T 1E2, Canada

[3]Medical Research Council Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, BS8 2BN, UK

[4]Institute for Aging Research, Hebrew SeniorLife, Boston, MA 02131, USA

[5]Harvard Medical School, Boston, MA 02115, USA

[6]Broad Institute of MIT and Harvard, Boston, MA 02142, USA

[7]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA

[8]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands

[9]The Generation R Study Group, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands

[10]Department of Epidemiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands

[11] Department of Internal Medicine, Section of Geriatrics, Academic Medical Center, Amsterdam, 1105 AZ, The Netherlands

[12]Department of medical sciences, Uppsala university, Uppsala, 751 85, Sweden

[13] Kidney Research Institute, Division of Nephrology, University of Washington, Seattle, WA 98195, USA

[14]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[15]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[16]Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, 40530, Sweden

[17]Bioinformatics Core Facility, Sahlgrenska Academy, University of Gothenburg, Gothenburg, 41390, Sweden

[18]Department of Internal Medicine, section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, 2333 ZA, the Netherlands

[19]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC H3A 1A2, Canada

[20]Department of Oncology, McGill University, Montreal, QC H4A 3T2, Canada

[21] Cardiovascular Health Research Unit, Departments of Medicine and Biostatistics, University of Washington, Seattle, WA 98101, USA

[22]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA 98101, USA

[23]Kaiser Permanente Washington Health Research Unit, Seattle, WA 98101, USA.

[24]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics at Harbor-UCLA Medical Center, Torrance, CA 90502, USA

[25]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[26]Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[27] Menzies Institute for Medical Research University of Tasmania, Locked Bag 23, Hobart, Tasmania 7000, Australia

[28]Geriatric Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 43180, Mölndal, Sweden

[29]Geriatric Medicine, Sahlgrenska University Hospital, 43180, Mölndal, Sweden.

[30]Clinical and Molecular Osteoporosis Research Unit, Department of Clinical Sciences, Lund University, and Department of Orthopaedics, Skåne University Hospital, 22241, Malmö, Sweden

[31]Department of Child and Adolescent Psychiatry/Psychology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands

[32]Department of Epidemiology and Biostatistics and the EMGO Institute of Health and Care Research, VU University Medical Center, Amsterdam, 1081 HV, The Netherlands

[33]Department of Radiology and Nuclear Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands

[34]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, 2333 ZA, The Netherlands

[35]Department of Endocrinology, VU University Medical Center, Amsterdam, 1081 HV, The Netherlands

[36] Division of Cardiovascular Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA

[37]Bone and Mineral Unit, Oregon Health & Science University, Portland, OR 97239, USA

[38]Department of Medicine, Oregon Health & Science University, Portland, OR 97239, USA

[39]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, 2333 ZA, The Netherlands

[40]Department of surgical sciences, Uppsala university, 75105, Uppsala, Sweden

[41]Division of Human Nutrition, Wageningen University, Wageningen, 6708 WE, The Netherlands

[42]Division of Endocrinology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[43]Beth Israel Deaconess Medical Center, Boston, MA 02215, USA

[44]University of Cambridge, Department of Clinical Neurosciences, Box 165, Cambridge Biomedical Campus, Hills Road, Cambridge, CB2 0QQ, UK

[45]Department of Twin Research and Genetic Epidemiology, King's College London, London, WC2R 2LS, United Kingdom

[46]Department of Medicine, McGill University, Montreal, QC H3G 1Y6, Canada

[47]These authors contributed equally to this work

[48] These authors contributed equally to this work

Corresponding author:

Brent Richards, MD, MSc

Pavillon H-413, Jewish General Hospital

3755 Cote Ste Catherine

Montreal, QC, Canada, H3T 1E2

T: +1 514 340 8222 ext. 4362 /F: +1 514 340 7529

brent.richards@mcgill.ca

Number of tables: 5, number of figures: 7. This article has Supplemental Data

2.1 Abstract

Vitamin D insufficiency is common, correctable and influenced by genetic factors, and it has been associated with risk of several diseases. We sought to identify low-frequency genetic variants that strongly increased the risk of vitamin D insufficiency and tested their effect on risk of multiple sclerosis, a disease influenced by low vitamin D concentrations. We used whole-genome sequencing data from 2,619 individuals through the UK10K program and deep imputation data from 39,655 genome-wide genotyped individuals. Meta-analysis of the summary statistics from 19 cohorts identified a low-frequency synonymous coding p.Asp120Asp variant (rs117913124[A], minor allele frequency=2.5%) in *CYP2R1* which conferred a large effect on 25-hydroxyvitamin D (25OHD) levels (-0.43 standard deviations of standardized natural log-transformed 25OHD, per A allele, P-value = 1.5×10^{-88}). The effect on 25OHD was four-times larger and independent of the effect of a previously described common variant near *CYP2R1*. By analyzing 8,711 individuals we showed that heterozygote carriers of this low-frequency variant have an increased risk of vitamin D insufficiency (OR=2.2, 95% CI 1.78-2.78, P=1.26×10^{-12}). Individuals carrying one copy of this variant had also an increased odds of multiple sclerosis (OR=1.4, 95%CI 1.19-1.64, P=2.63×10^{-5}) in a sample of 5,927 cases and 5,599 controls. In conclusion, we describe a low-frequency coding variant in *CYP2R1*, which exerts the largest effect upon 25OHD levels identified to date in the general European population and implicates vitamin D in the etiology of multiple sclerosis.

2.2 Introduction

Vitamin D insufficiency affects approximately 40% of the general population in developed countries [1]. This may have important public health consequences, since vitamin D insufficiency has been associated with musculoskeletal consequences and several common diseases, such as multiple sclerosis (MIM:126200), types 1 and 2 diabetes (MIM:222100 and MIM:125853) and several cancers [2]. Further, repletion of vitamin D status can be achieved safely and inexpensively. Thus, understanding the determinants of vitamin D insufficiency, and their effects, can provide a better understanding of the role of vitamin D in disease susceptibility with potentially important public health benefits.

Approximately half of the variability in the concentration of the widely accepted biomarker for vitamin D status, 25-hydroxyvitamin D (25OHD), has been attributed to genetic factors in twin and family studies [3; 4]. Four common genetic variants (minor allele frequency [MAF] >5%) in loci near four genes known to be involved in cholesterol synthesis (*DHCR7* [MIM;602858]), hydroxylation (*CYP2R1* [MIM:608713]), vitamin D transport (*GC* [MIM:139200]) and catabolism (*CYP24A1* [MIM:126065]) are strongly associated with 25OHD levels, yet explain little of its heritability [5]. Low-frequency and rare genetic variants (defined as variants with a MAF of ≤5% and ≤1% respectively) have recently been found to have large effects on clinically relevant traits [6-8] providing an opportunity to better understand the biologic mechanisms influencing disease susceptibility in the general population.

Therefore, the principal objective of the present study was to detect low-frequency and rare variants with large effects on 25OHD levels, through a large-scale meta-analysis and describe their biological and clinical relevance. Similar to an earlier genome-wide association study

(GWAS) studying common genetic variation (MAF ≥5%) by the SUNLIGHT Consortium [5], we sought to increase understanding of the genetic etiology of vitamin D variation within the general population, however, our current study focused on genetic variation with a MAF <5%. This has only recently been made possible through whole-genome sequencing and the use of improved genotype imputation for low frequency and rare variants, with the recent availability of large whole genome sequencing reference panels [9]. The second objective of this study was to better understand if low-frequency genetic variants with large effects on 25OHD could predict a higher risk of vitamin D insufficiency in their carriers, and whether vitamin D intake through diet may interact with such genetic factors to prevent, or magnify, vitamin D insufficiency. Finally, we sought to understand whether these genetic determinants of 25OHD levels are implicated in multiple sclerosis, a disease influenced by low 25OHD levels[10].

To do so, we first undertook an association study of whole-genome sequence data and deeply imputed genome-wide genotypes to identify novel genetic determinants of vitamin D in 42,274 individuals. We next tested if these genetic variants conferred a higher risk of vitamin D insufficiency in 8,711 subjects and whether this insufficiency showed effect modification by dietary intake. Last we assessed their effect on multiple sclerosis in a separate sample of 5,927 cases and 5,599 controls.

2.3 Material and Methods

2.3.1 Cohorts

All human studies were approved by each respective institutional or national ethics review committees, and all participants provided written informed consent. To investigate the role of rare and low-frequency genetic variation on 25OHD levels in individuals of European descent, we used whole genome sequencing (WGS) data at mean read depth of 6.7 x in 2,619 subjects from two cohorts in the UK10K project [11] with available 25OHD phenotypes **(Table 1).** We also used imputation reference panels to impute variants that were missing, or poorly captured, from previous GWAS in 39,655 subjects **(Table 1 and Figure 1).** The participating individuals were drawn from independent cohorts of individuals of European descent**.** Detailed description of each of the participating studies is provided in **Table S1.**

2.3.2 25OHD Measurements

The methods applied to measure 25OHD levels differed among the participating cohorts (**Tables S1 and S6)**. The four methods used were tandem mass spectrometry (in BMDCS, MrOS and BPROOF), combined high-performance liquid chromatography with mass spectrometry (in ALSPAC, BPROOF, CHS, ULSAM, NEO, Generation R), chemiluminescence immunoassay (DiaSorin, Inc, Stillwater, MN) (in TUK, PIVUS, FHS, MrOS Malmo, MrOS GBG and GOOD) and an electrochemiluminescence immunoassay (COBAS, Roche Diagnostics GmbH) (in RSI, RSII and RSIII). Detection limits for the different methods are provided in the **Table S6**.

2.3.3 Whole-Genome Sequencing, Genotyping and Imputation

ALSPAC WGS and TUK WGS cohorts had been sequenced at an average read depth of 6.7x through the UK10K consortium (www.UK10K.org) using the Illumina HiSeq platform, and aligned to the GRCh37 human reference using Burrows-Wheeler Aligner 31[12]. Single-nucleotide variant (SNV) calls were completed using samtools/bcftools[13], and VQSR[14] and GATK were used to recall these variants. The whole genome sequencing for the ALSPAC and TwinsUK cohorts has been described in detail in a previous publication from our group[7]. **Table S8** summarizes the data generation method for sequencing-based cohorts.

Participating studies separately genotyped samples and imputed them to WGS-based reference panels. The most recent imputation panels, such as the UK10K and 1000Genomes Project (v3) combined panel, which in total contained 7,562 haplotypes from the UK10K Project and 2,184 haplotypes from the 1000 Genomes Project [9], and the Haplotype Reference Consortium (HRC) panel, with 64,976 haplotypes[15], enabled more accurate imputation of low frequency variants, when compared to the UK10K or the 1000Genomes reference panel alone[9]. Specifically, 11 out of the 17 participating cohorts were imputed to the UK10K and 1000 Genomes reference panel (total number of imputed individuals included in the meta-analysis N=25,589). Three of the participating cohorts were imputed using the HRC panel (total number of imputed individuals N=5,717). Finally, 2 cohorts were imputed to the 1000Genomes panel (N=7,536), and 1 cohort was imputed to the UK10K panel (N=863). **(Table S1**). Details on genotyping methods and imputation for the 17 participating cohorts are presented in **Table S6.** Info scores for the imputed SNVs per participating cohort are presented in **Table S7.** To assess the quality of imputation, we tested the non-reference discordance rate for the low frequency genome-wide significant SNVs and found this to be 0% (**Table S9**).

2.3.4 Association Testing for 25OHD levels and Meta-analysis

A GWAS was conducted separately by each cohort using an additive genetic model for 25OHD levels. Because 25OHD concentrations were measured using different methods, log-transformed 25OHD levels were standardized to z-scores, after being adjusted for age, sex, BMI, and season of measurement. Specifically, the phenotype for each GWAS study was prepared according to the following steps:1) 25OHD levels were log-transformed to ensure normality 2) Linear regression models were used to generate cohort-specific residuals of log transformed 25OHD levels adjusted for covariates (age, sex, BMI and season). Season was treated as a non-ordinal categorical variable (summer: July to September, fall: October to December, winter: January to March, and spring: April to June). 3) The mean of log transformed 25OHD levels was added to the residuals to create the adjusted 25OHD phenotype. 4) The above phenotype was then normalized within each cohort (mean of zero with SD of one) to make the phenotype consistent across cohorts, since 25OHD levels have been measured in different cohorts in our consortium using different methods. 5) Finally, outliers beyond 5 standard deviations were removed from step (4).

For comparison purposes, we computed the average 25OHD levels, adjusted for age, sex, BMI and season of measurement, in one cohort of our meta-analysis (TUK WGS) in carriers and non-carriers of the lead SNV(s).

The software used by each cohort to perform a GWAS is listed in **Table S1**. Single variant tests were undertaken for variants with MAF>0.1%, using an additive effect of the minor allele at each variant in each cohort. The type of software employed for single variant testing for each

cohort is shown in **Table S1**. Studies with related individuals used software that accounted for relatedness. Cohort-specific genomic inflation factors (lambdas) are also shown in **Table S1** (the mean lambda was 1.015).

We then meta-analyzed association results from all discovery cohorts (N total = 42,274). This stage included validation of results file format, filtering files by the above QC criteria, comparison of trait distributions among different studies, identification of potential biases (large betas and/or standard errors, inconsistent effect allele frequencies, extreme lambdas). Meta-analysis quality control of the GWAS data included the following SNV-level exclusion criteria: i) Info score <0.4, ii) HWE P-value $<10^{-6}$ iii) Missingness >0.05, and iv) MAF <0.5%. Alignment of the SNVs across studies was done using the chromosome and position information for each variant according to genome build hg19. SNVs in the X chromosome were not included in the meta-analysis. Fixed–effects meta-analysis was performed using the software package GWAMA[16] adjusting for genomic control. We tested bi-allelic SNVs with MAF $\geq$ 0.5% for association, declaring genome-wide statistical significance at $P \leq 1.2 \times 10^{-8}$ for variants present in more than one study. This stringent p-value threshold was set to adjust for all independent SNVs above the MAF threshold of 0.5%.[17]

Conditional analysis was undertaken for the four previously described lead vitamin D SNVs from the SUNLIGHT consortium using the GCTA package [18]. This method uses an approximate conditional analysis approach from summary-level statistics from the meta-analysis and linkage disequilibrium corrections between SNVs estimated from a reference sample. We used UK10K individuals as the reference sample to calculate the linkage disequilibrium information of SNVs. The associated regions flanking within 400kb of the top

SNVs from SUNLIGHT were extracted and the conditional analyses were conducted within these regions. Conditional analyses of individual variants presented in **Table 2** and **Table S5** were conducted using GCTA v 0.93.9 using default parameters. Haplotype block analyses were used for the candidate variants of interest by deriving phased haplotypes from 1013 individuals from the TUK WGS cohort using a custom R package.

2.3.5 Effects on Vitamin D insufficiency

To investigate the effect of genome-wide significant SNVs on vitamin D insufficiency (defined as 25OHD levels below 50 nmol/L), we used data from 4 cohorts: TUK Imputed, TUK WGS, BPROOF and MrOS (n_{total}=8,711). Logistic regression of this binary phenotype was performed against the SNVs, adjusting for the following covariates: age, sex, BMI, and season of measurement. Meta-analysis of cohort-level summary statistics was performed in R [19] using the epitools [20] and metafor packages[21].

2.3.6 Interaction analysis with Vitamin D intake

We undertook an interaction analysis of our candidate SNV(s) with vitamin D dietary intake (continuous and tertiles) in 9,224 individuals from five of the cohorts participating in our discovery phase (Framingham, PIVUS, ULSAM, BPROOF and RSIII). A detailed description of the method to capture vitamin D intake in each one of the participating cohorts appears in **Table S6**. Linear regression was conducted in each of these studies under an additive genetic model. The following variables and co-variables were included in the model: log-transformed serum 25OHD as the dependent variable; SNV genotype (coded as 0, 1 or 2) as an independent variable; SNV (genotype)* dietary vitamin D intake (continuous or tertiles respectively) as an interaction term; age, sex, BMI, season of 25OHD measurement, dietary vitamin D intake

(continuous or tertiles), supplemented vitamin D (yes/no), and total energy intake as covariates. The results from the 5 studies were meta-analyzed using a fixed-effects model using the metafor tool of the R statistical package.

2.3.7 Effects on Multiple Sclerosis

We tested the effect of the genome-wide significant SNVs on the risk of multiple sclerosis in 5,927 cases and 5,599 controls, assuming an additive genetic model. Controls were obtained from the UK Biobank [22] by random selection of participants without multiple sclerosis. The cases were obtained from UK Biobank[22], previously published MS GWAS[23; 24] and newly genotyped UK patients. Prior to genotype imputation of the genotyped cases, numerous quality control criteria were applied to ensure unbiased genotype calls between cohorts. These included retaining only SNVs with MAF > 1% and excluding SNVs or samples with high missingness [25]. Further, samples were assessed for population stratification using EIGENSTRAT [26; 27] and outliers were removed. Genotype data was then imputed using the Sanger Imputation Service[15] with the combined UK10K and 1000 Genomes Phase 3 reference panels[9; 28], the same reference panel used for the UK Biobank controls. Genotype data was phased using EAGLE2[29] and imputed using PBWT[30]. Association testing was undertaken using SNPTEST[31] on the combined case/control dataset, testing the additive effect of each allele on multiple sclerosis status, and including the top 10 principal components from EIGENSTRAT [26; 27] to adjust for population stratification and batch effects.

2.4 Results

2.4.1 GWAS

After strict quality control, the genomic inflation factor for the meta-analysis of 19 GWAS studies was 0.99, suggesting lack of bias due to population stratification (**Figure 2**). Through meta-analysis of 11,026,511 sequenced and imputed variants from our discovery cohorts **(Table 1)**, we identified a signal at the chromosome 11p.15.2 locus, harboring variants associated with 25OHD levels (lead low-frequency SNV p.Asp120Asp [rs117913124(A)], MAF = 2.5%, allelic effect size = -0.43 standard deviations of the standardized log-transformed 25OHD levels [SD], P = 1.5×10^{-88}, **Figure 3 and Table 2)**. The direction of effect was consistent across all discovery cohorts **(Table 3 and Figure 3A)** and the mean imputation information score for the imputed studies was 0.97. This low-frequency synonymous coding variant is in exon 4 of the *CYP2R1* and is ~14 kb from the previously identified common *CYP2R1* variant, rs10741657 (r^2 between these two SNVs= 0.03) **(Figure 4).** To our knowledge**,** the rs117913124 SNV has not previously been associated with any vitamin D-related traits in humans.

A comparison of the average 25OHD levels, adjusted for age, sex, BMI and season of measurement, in non-carriers and heterozygote carriers of the A allele of rs117913124 in the TUK WGS appears in **Figure S1.** The average 25OHD levels, adjusted for age, sex, BMI and season of measurement were computed in 542 individuals from the Twins UK WGS cohort, among which 510 were no carriers and 32 were heterozygote carriers of the A allele of rs117913124 (no homozygote carriers present in this cohort). After removing outliers (adjusted 25OHD levels below and above 3 SD from the mean), we included in our analysis 449 non-carriers and 30 heterozygote carriers (for a total of 479 individuals). A linear regression model with the adjusted 25OHD levels as the dependent variable and the dose of the "A" allele of rs117913124 (numeric factor, 1 or 0) as the independent variable demonstrated a 8.3 nmol/L

decrease in the adjusted 25OHD levels per "A" allele. The mean adjusted 25OHD levels were 64.3 nmol/L in non-carriers vs 56.0 nmol/L in heterozygote carriers.

Two-way conditional analysis between the *CYP2R1* common (rs10741657) and low-frequency (rs117913124) variants revealed that the two association signals are largely independent. Specifically, after conditioning on rs10741657, rs117913124 remained strongly associated with 25OHD level (P_{cond} = $2.4x10^{-78}$); after conditioning on rs11791324, the effect of rs10741657 on 25OHD level remained significant (P_{cond}= $4.0 x10^{-33}$ versus $P_{pre-cond}$= $8.8 X 10^{-45}$) **(Table 2 and Table S5).** Further, no other low frequency variant in the region remained significant when conditioning on rs117913124 (**Table 2**). To further disentangle the role of rs117913124 from rs10741657 on 25OHD levels, we undertook a haplotype analysis based on WGS data from 3,781 individuals from the TUK WGS and ALSPAC WGS cohorts. We found that the 25OHD decreasing allele A of rs117913124 was always transmitted in the same haplotype block with the 25OHD decreasing allele G of the common *CYP2R1* variant rs10741657. By using 25OHD data from the TUK WGS cohort, we compared the 25OHD levels among carriers of the various haplotype blocks. We observed evidence of decrease in the 25OHD levels in carriers of the A allele of the rs117913124 compared to non-carriers independent of the presence of the effect allele G of the common *CYP2R1* variant (**Table 4)** No other low-frequency or rare variants were identified in the three previously described vitamin D-related loci at *DHCR7*, *GC* and *CYP24A1*. The mean effect size of the four previously reported common genome-wide significant SNVs (MAF ≥ 5%) from the SUNLIGHT consortium was -0.13 SD and the largest effect size was -0.25 SD (for the *GC* variant) in our meta-analysis **(Table S3 and Figure 3B).** The effect size of rs10741657(G), the known common *CYP2R1* variant, was -0.09 SD. Hence, the observed effect size of rs117913124 is 3-fold larger than the above mean, 4-fold larger that of the common *CYP2R1*

variant and almost twice that of the largest previously reported effect of the *GC* variant. Last, the percentage of the variance of the 25OHD phenotype explained by the low-frequency *CYP2R1* variant was more than double than the percentage of the variance explained by the *CYP2R1* common variant (0.9% vs 0.4%).

We also identified 18 genome-wide significant low-frequency and rare SNVs on the same chromosome 11 region as rs117914124 located in the neighboring *PDE3B* (MIM:602047) **(Table 2, Table S4 and Figure 4B).** Signals from these SNVs in *PDE3B* were independent of the common variant at *CYP2R1* (**Table 2).** We then created haplotype blocks with rs117913124 and SNVs at *PDE3B* based on haplotype information from the 3,781 individuals from the TUK WGS and ALSPAC WGS cohorts (**Table S2**). We found that the 25OHD decreasing allele (A) of the rs117913124 was always inherited with the 25OHD decreasing allele (A) of its perfect proxy rs116970203 (r^2=1). Therefore, rs116970203 is not likely to have a distinct effect from rs117913124 on 25OHD levels. On the other hand, the 25OHD decreasing alleles of the remaining four low-frequency variants (all having a MAF of approximately 1.4%) were not always inherited in the same haplotype block as the rs117913124 and rs116970203 and were in moderate linkage disequilibrium with the rs117913124 (all r^2< 0.6, **Figure 4B and Figure 4C**). Each of the four alleles is in almost perfect linkage disequilibrium with the remaining three (all r^2 >0.96). This implied that these four SNVs might influence 25OHD levels independently of the rs117913124. Nevertheless, as mentioned above, when conditioning on the lead low-frequency *CYP2R1* SNV rs117913124, the P-values of the 4 *PDE3B* SNVs became non-significant and their betas decreased substantially (**Table 2**), demonstrating that they likely do not represent an independent signal at the chromosome 11 locus.

2.4.2 rs117913124 and risk of vitamin D insufficiency

To further investigate the clinical significance of the low-frequency *CYP2R1* variant rs117913124, we tested its effect on a binary outcome for vitamin D insufficiency (defined as 25OHD levels < 50 nmol/L) in 8,711 individuals from 4 studies (TUK WGS, TUK IMP, BPROOF and MROS). rs117913124 was strongly associated with an increased risk of vitamin D insufficiency (OR = 2.20, 95% CI 1.8-2.8, P =1.2 x 10^{-12}) (**Figure 5),** after control for relevant covariates as described in the Methods section**.**

2.4.3 Common 25OHD-associated SNVs

We report two additional loci associated with 25OHD levels (**Table 5**). Variants leading these associations were common and exerted a rather small effect on 25OHD: first, a variant in chromosome 12 (rs3819817[C], intronic to *HAL* [MIM:609457]), with a MAF of 45%, a beta of 0.04 and a P-value of 3.2 x 10^{-10}. Second, a variant in chromosome 14 (rs2277458[G], intronic to *GEMIN2* [MIM:602595]), with a MAF of 21%, a beta of -0.05 and a P-value of 6.0 x 10^{-9}. Both variants were present in all 19 studies, and the direction of the effect was the same among the 19 studies (**Figure 6**). Neither the *HAL* nor the *GEMIN2* loci are previously known to be associated with 25OHD levels. Of note, neither variant was present in the HapMap imputation reference used in the SUNLIGHT study.

2.4.4 Interaction analysis

CYP2R1 encodes the enzyme responsible for 25-hydroxylation of vitamin D in the liver [32], a necessary step in the conversion of dietary vitamin D and vitamin D oral supplements to the active metabolite, 1,25 dihydroxy-vitamin D. Therefore, we hypothesized that individuals heterozygous or homozygous for rs117913124 in *CYP2R1* would not show a response in their 25OHD levels to vitamin D intake compared to non-carriers. In other words, we expected

carriers of the effect allele of rs117913124 to have steadily lower 25OHD levels, independently of their vitamin D intake. To investigate this hypothesis, we tested the presence of interaction of rs117913124 with vitamin D dietary intake (continuous values and tertiles) on 25OHD levels in 9,224 individuals from 5 studies **(Figure S2)**. We found no interaction between rs117913124 and dietary vitamin D intake (beta = -0.0002; P-value for interaction = 0.41 for continuous vitamin D intake and beta = 0.012; P-value = 0.60 for tertiles of vitamin D intake). Since the two common 25OHD-associated SNVs are located in genes (*HAL* and *GEMIN2*) with no known role in the processing of dietary vitamin D, we found no biological rationale for undertaking a gene-diet interaction analysis for these variants.

2.4.5 25OHD-associated variants and risk of multiple sclerosis

We tested whether the *CYP2R1* low-frequency variant rs117913124 and the common variants rsrs3819817 and rs2277458 in *HAL* and *GEMIN2*, respectively, influenced the risk of multiple sclerosis. In a sample of 5,927 multiple sclerosis cases and 5,599 controls, we found that the 25OHD decreasing allele at rs117913124[A], was associated with an increased odds of multiple sclerosis: OR = 1.40 (95%CI: 1.19-1.64); P-value = 2.6×10^{-5}. By way of comparison, the OR of multiple sclerosis for the common *CYP2R1* variant was 1.03 (95%CI: 0.97-1.08); P-value 0.03 in the same multiple sclerosis study, and has previously been reported to be 1.05 (95%CI: 1.02-1.09); P-value 0.004 in a separate study [33]. Thus, the effect per allele of rs117913124 on multiple sclerosis was 12.4-fold larger than that attributed to the already known common variant at *CYP2R1*. With regards to the two common SNVs, the 25OHD decreasing allele [T] at the *HAL* variant rs3819817 was not clearly associated with risk of multiple sclerosis, however there was a trend in the expected direction: OR = 1.05 (95%CI: 1.00-1.11); P-value = 0.07. We found no association between the 25OHD decreasing allele

[G] at the *GEMIN2* variant rs2277458 and risk of multiple sclerosis: OR = 1.03 (95%CI: 0.96-1.11); P-value = 0.34.

2.5 Discussion

Through the largest meta-analysis of genome-wide association studies for 25OHD levels in European populations to date, we have identified a low-frequency, synonymous coding genetic variant of large effect that strongly associates with 25OHD levels. This variant has an effect size four-fold larger than that described for the common variant in the same gene (*CYP2R1*) and is associated with two-fold increase in risk of vitamin D insufficiency and a 40% increase in the odds of developing multiple sclerosis. The biologic plausibility of these findings is supported by the fact that the low-frequency variant is located in *CYP2R1*, the major hepatic 25-hydroxylase for vitamin D [32]. These findings are of clinical relevance since 5% of the general European population carry this variant in either the homozygous or heterozygous state, and it is associated with a clinically relevant increase in the risk of multiple sclerosis.

Our study was enabled by large imputation reference panels (UK10K/1000 Genomes and HRC), which offer at least 10-fold more European samples than the 1000 Genomes reference panel alone. We did not identify genome-wide significant variants of large effect on 25OHD in novel genes in Europeans, although we found variants with smaller effects in two loci not previously known to be associated to 25OHD. Yet we did identify low-frequency variants in a known vitamin D related-gene with much larger effects than the previously described common variants.

CYP2R1 encodes the enzyme responsible for 25-hydroxylation of vitamin D, and is one of the two main enzymes responsible for vitamin D hepatic metabolism [32] (**Figure 7**). Rare mutations in *CYP2R1* have already been described to cause rickets (MIM: 27744) [32; 34]. Due to the important role of *CYP2R1* in the conversion of dietary vitamin D and vitamin D oral supplements to the active form of vitamin D, we hypothesized that carriers of the low-frequency *CYP2R1* variant might respond poorly to vitamin D replacement therapy. We tested this hypothesis by undertaking an interaction analysis between the *CYP2R1* low frequency variant and dietary vitamin D intake, which showed no clear interaction. However, we note that gene by environment interaction studies are generally underpowered, measurement error in dietary data is common, and this interaction was further limited by time differences between dietary intake assessment and measurement of 25OHD levels. Therefore, whether this genetic variant influences 25OHD response to vitamin D administration requires further study.

Although the aim of the present study was to describe variants of low MAF and large effect on 25OHD, we report two common genetic variants of small effect size on chromosome 12 (*HAL* gene) and chromosome 14 (*GEMIN2* gene) that reached genome-wide level significance in our meta-analysis. Although there is no existing evidence of implication of *GEMIN2* in vitamin D related physiologic pathways, *HAL* is expressed in the skin and is involved in formation of urocanic acid, a "natural sunscreen" [35; 36]. Thus, this could constitute a plausible pathophysiologic mechanism implicating *HAL* in vitamin D synthesis in the skin. Additional functional follow-up of the signals in chromosomes 12 and 14 is needed to characterize the genes and/or mechanisms underlying these associations.

Our findings may have clinical relevance for several reasons: First, individuals carrying at least one copy of the low-frequency *CYP2R1* variant have lowered levels of 25OHD by a clinically relevant degree. Specifically, the risk of vitamin D insufficiency is doubled in these individuals. Second, their risk of multiple sclerosis is also increased in accordance with previous evidence supporting a causal role for vitamin D in the risk of multiple sclerosis[10]. Third, these findings affect ~5% of individuals of European descent. And last, rs117913124 could be used as an additional genetic predictor of low 25OHD levels, along with the previously identified common vitamin D-related variants, in Mendelian randomization studies investigating the causal role of low vitamin D levels in human disease.

Our study also has its limitations. First, although the scope of our study was detection of low-frequency and rare variants, we opted to include in our meta-analysis two whole genome sequencing studies with a relatively low read depth of 6.7x, as well as three studies imputed to older imputation panels (1000Genomes and UK10K). These studies have a limited capacity to capture very rare variants, which might explain why we failed to identify such associations. The gene-diet interaction analysis, as mentioned above, may have lacked statistical power, in addition to the limitations arising from the time-difference between dietary vitamin D intake assessments and 25OHD measurements. Since our analysis is restricted to populations of European ancestry, we cannot make any assumptions concerning the effect of rs117913124 in non-European populations. Nonetheless, based on the 1000Genomes reference, this variant is rare in Africans (MAF = 0.3%) and has not been described in East Asians (MAF = 0%). Therefore, large sample sizes of these populations will be required to describe with any certainty the effect of this variant on 25OHD level in these populations. Finally, in the absence of functional experiments showing the exact function of the rs117913124 on *CYP2R1* and given that this synonymous polymorphism does not affect protein sequence, we cannot

unequivocally confirm that this low-frequency variant is causal, however, given that this is a coding variant in a well-documented 25OHD-associated gene, it seems most likely that it exerts its effect on *CYP2R1*.

In conclusion, our findings demonstrate the utility of whole-genome sequencing-based discovery and deep imputation to enable the characterization of genetic associations, offering an improved understanding of the pathophysiology of vitamin D, an enriched set of genetic predictors of 25OHD levels for future study, and enabling the identification of groups at increased risk for vitamin D insufficiency and multiple sclerosis.

2.6 Supplemental Data Description

Supplemental Data of this article include 2 figures, 9 Tables, Funding information, Author Information and Acknowledgements.

2.7 Acknowledgements

The authors have no conflicts of interest. Detailed acknowledgments are included in the Supplemental Data.

2.8 Web Resources

URL for Online Mendelian Inheritance in Man: http://www.omim.org

URL for the UK10K program: http://www.uk10k.org

URL for VQSLOD:

http://www.broadinstitute.org/gsa/wiki/index.php/Variant_quality_score_recalibration

URL for GWAMA: http://www.geenivaramu.ee/en/tools/gwama

URL for GCTA: http://cnsgenomics.com/software/gcta/

2.9 References

1. Forrest, K.Y., and Stuhldreher, W.L. (2011). Prevalence and correlates of vitamin D deficiency in US adults. Nutr Res 31, 48-54.

2. Rosen, C.J., Adams, J.S., Bikle, D.D., Black, D.M., Demay, M.B., Manson, J.E., Murad, M.H., and Kovacs, C.S. (2012). The nonskeletal effects of vitamin D: an Endocrine Society scientific statement. Endocr Rev 33, 456-492.

3. Shea, M.K., Benjamin, E.J., Dupuis, J., Massaro, J.M., Jacques, P.F., D'Agostino, R.B., Sr., Ordovas, J.M., O'Donnell, C.J., Dawson-Hughes, B., Vasan, R.S., et al. (2009). Genetic and non-genetic correlates of vitamins K and D. Eur J Clin Nutr 63, 458-464.

4. Livshits, G., Karasik, D., and Seibel, M.J. (1999). Statistical genetic analysis of plasma levels of vitamin D: familial study. Ann Hum Genet 63, 429-439.

5. Wang, T.J., Zhang, F., Richards, J.B., Kestenbaum, B., van Meurs, J.B., Berry, D., Kiel, D.P., Streeten, E.A., Ohlsson, C., Koller, D.L., et al. (2010). Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet 376, 180-188.

6. Sidore, C., Busonero, F., Maschio, A., Porcu, E., Naitza, S., Zoledziewska, M., Mulas, A., Pistis, G., Steri, M., Danjou, F., et al. (2015). Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat Genet 47, 1272-1281.

7. Zheng, H.F., Forgetta, V., Hsu, Y.H., Estrada, K., Rosello-Diez, A., Leo, P.J., Dahia, C.L., Park-Min, K.H., Tobias, J.H., Kooperberg, C., et al. (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature 526, 112-117.

8. Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., and Hobbs, H.H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305, 869-872.

9. Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J.L., Danecek, P., Malerba, G., Trabetti, E., Zheng, H.F., et al. (2015). Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat Commun 6, 8111.

10. Mokry, L.E., Ross, S., Ahmad, O.S., Forgetta, V., Smith, G.D., Goltzman, D., Leong, A., Greenwood, C.M., Thanassoulis, G., and Richards, J.B. (2015). Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLoS Med 12, e1001866.

11. Consortium, U.K., Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R., Xu, C., Futema, M., et al. (2015). The UK10K project identifies rare variants in health and disease. Nature 526, 82-90.

12. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987-2993.

13. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics 27, 2156-2158.

14. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43, 491-498.

15. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48, 1279-1283.

16. Magi, R., and Morris, A.P. (2010). GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 11, 288.

17. Xu, C., Tachmazidou, I., Walter, K., Ciampi, A., Zeggini, E., Greenwood, C.M., and Consortium, U.K. (2014). Estimating genome-wide significance for whole-genome sequencing studies. Genet Epidemiol 38, 281-290.

18. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88, 76-82.

19. Team, R.C. (2013). R: A language and environment for statistical

computing. In. (R Foundation for Statistical Computing, Vienna, Austria.

20. Aragon T.J., Wollschlaeger D., Omidpanah A. (2017). epitools: Epidemiology Tools. https://cran.r-project.org/package=epitools

21. Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. . Journal of Statistical Software 36, 1-48.

22. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). 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.

23. International Multiple Sclerosis Genetics Consortium, Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I., Gabriel, S.B., Mirel, D.B.,

et al. (2007). Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med 357, 851-862.

24. Australia, and New Zealand Multiple Sclerosis Genetics Consortium (2009). Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet 41, 824-828.

25. Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., and Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. Nat Protoc 5, 1564-1573.

26. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904-909.

27. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet 2, e190.

28. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature 526, 68-74.

29. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., Y, A.R., H, K.F., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 48, 1443-1448.

30. Durbin, R. (2014). Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). Bioinformatics 30, 1266-1272.

31. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39, 906-913.

32. Cheng, J.B., Levine, M.A., Bell, N.H., Mangelsdorf, D.J., and Russell, D.W. (2004). Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. Proc Natl Acad Sci U S A 101, 7711-7715.

33. International Multiple Sclerosis Genetics Consortium, Beecham, A.H., Patsopoulos, N.A., Xifara, D.K., Davis, M.F., Kemppinen, A., Cotsapas, C., Shah, T.S., Spencer, C., Booth, D., et al. (2013). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet 45, 1353-1360.

34. Casella, S.J., Reiner, B.J., Chen, T.C., Holick, M.F., and Harrison, H.E. (1994). A possible genetic defect in 25-hydroxylation as a cause of rickets. J Pediatr 124, 929-932.

35. Barresi, C., Stremnitzer, C., Mlitz, V., Kezic, S., Kammeyer, A., Ghannadan, M., Posa-Markaryan, K., Selden, C., Tschachler, E., and Eckhart, L. (2011). Increased sensitivity of histidinemic mice to UVB radiation suggests a crucial role of endogenous urocanic acid in photoprotection. J Invest Dermatol 131, 188-194.

36. Suchi, M., Sano, H., Mizuno, H., and Wada, Y. (1995). Molecular cloning and structural characterization of the human histidase gene (HAL). Genomics 29, 98-104.

2.10 Tables and Figures

2.10.1 Tables

Table 1. Participating cohorts and number of DNA samples per cohort.

WGS: Whole-Genome Sequenced

Study Acronym*	Imputed	WGS	TOTAL
ALSPAC	3,679	1,606	
TUK	1,919	1,013	
Generation R	1,442		
BPROOF	2,514		
FHS	5,402		
MrOS	3,265		
RSI	3,320		
RSII	2,022		
RSIII	2,913		
CHS	1,792		
BMDCS	863		
MrOS GBG	945		
GOOD	921		
MrOS Malmo	893		
PIVUS	943		
ULSAM	1,095		
NEO	5,727		
TOTAL	39,655	2,619	42,274

*For full names of the studies see Table S6

Table 2: Association results for genome-wide significant low-frequency variants from discovery 25OHD meta-analysis, before and after conditioning on the lead common *CYP2R1* SNP, rs10741657, and the lead low-frequency *CYP2R1* variant, rs117913124.

SNV	Chr	Position	EA*	EAF#	Candidate Gene	Function	Beta$	P-value		Beta$	P -value		Beta$	P-value	N
										Conditional on rs10741657			Conditional on rs117913124		
rs117913124	11	14900931	A	0.025	*CYP2R1*	exon 4 (synonymous codon)	-0.43	1.5×10^{-88}		-0.39	2.4×10^{-78}		NA	NA	41,336
rs116970203		14876718	A	0.025	*CYP2R1 (nearest gene: PDE3B)*	Intron 11 variant	-0.43	2.2×10^{-90}		-0.40	3.3×10^{-80}		NA	NA	41,138
rs117361591		14861957	T	0.014		Intron 11 variant	-0.44	9.1×10^{-51}		-0.40	2.2×10^{-44}		-0.05	0.017	38,286
rs117621176		14861320	G	0.014		Intron 11 variant	-0.44	8.7×10^{-51}		-0.40	2.1×10^{-44}		-0.05	0.016	38,273
rs142830933		14838760	C	0.014		Intron 5 variant	-0.44	1.4×10^{-48}		-0.40	1.7×10^{-42}		-0.05	0.03	37,541
rs117672174		14746404	T	0.014		Intron 1 variant	-0.43	2.8×10^{-45}		-0.39	2.9×10^{-39}		-0.04	0.062	37,209

*****Effect allele is the 25OHD decreasing allele

Effect allele frequency

$ Betas represent changes in standard deviations of the standardized log-transformed 25OHD levels

Table 3. Summary statistics results for the *CYP2R1* low-frequency variant, rs117913124, from 19 studies.

STUDY	25OHD	N	Effect Allele A*	Beta$	Standard	P-value	Information
ALSPAC Imputed	MS	3,675	0.028	-0.59	0.07	$3.43x10^{-18}$	0.99
ALSPAC WGS	MS	1,606	0.028	-0.65	0.11	$8.23x10^{-10}$	NA
BPROOF	MS	2,512	0.027	-0.4	0.09	$4.99x10^{-6}$	0.97
BMDCS	MS	863	0.019	-0.11	0.06	0.058	0.98
CHS	MS	1,581	0.022	-0.55	0.11	$5.15x10^{-7}$	0.88
FHS	CLIA	5,402	0.021	-0.45	0.07	$2.32x10^{-10}$	0.97
GenerationR	MS	1,442	0.033	-0.66	0.1	$1.78x10^{-6}$	1
GOOD	CLIA	921	0.028	-0.14	0.14	0.31	0.96
MrOS	MS	3,265	0.018	-0.76	0.09	$5.63x10^{-16}$	0.96
MrOS Malmo	CLIA	893	0.033	-0.33	0.14	0.016	0.94
MrOS GBG	CLIA	945	0.026	-0.61	0.14	$7.87x10^{-6}$	1
NEO	MS	5,727	0.025	-0.54	0.06	$2.73x10^{-19}$	1
PIVUS	CLIA	943	0.028	-0.66	0.14	$2.56x10^{-6}$	0.99
RSI	ECLIA	3,320	0.025	-0.19	0.08	0.019	0.98
RSII	ECLIA	2,022	0.033	-0.37	0.09	$2.38x10^{-5}$	0.99
RSIII	ECLIA	2,913	0.027	-0.51	0.08	$4.61x10^{-10}$	0.98
TUK Imputed	CLIA	1,919	0.021	-0.1	0.11	0.35	0.98
TUK WGS	CLIA	1,013	0.025	-0.39	0.14	0.006	NA
ULSAM	MS	1,095	0.025	-0.33	0.14	0.02	1

*****Effect allele is the 25OHD decreasing allele # MS: mass spectrometry, CLIA: chemiluminescence immunoassay, ECLIA: electrochemiluminescence immunoassay $ Betas represent changes in standard deviations of the standardized log-transformed 25OHD levels

Table 4. Effect of different haplotype combinations of the low frequency (rs117913124) and the common (rs10741657) CYP2R1 variants on 25OHD levels.

Results are based on individuals from the Twins UK Whole Genome Sequenced cohort (the first allele in each block is the rs117913124, the second allele is the rs10741657 for both chromatids). The two "AG" blocks in bold contain the 25OHD decreasing allele (A) of the low-frequency variant, which is always inherited with the 25OHD decreasing allele (G) of the common variant.

Haplotype		Beta$	P-value	N
rare/common*	rare/common*			
GA	**GA**	-0.02	0.79	156
AG	**GA**	-0.49	0.02	23
AG	**GG**	-0.3	0.13	27
GA	**GG**	0.01	0.87	477
GG	**GG**	0.05	0.58	330

***** The first allele in each chromatid corresponds to the low-frequency variant rs117913124; the second allele corresponds to the common variant rs10741657. 25OHD decreasing alleles appear in bold for both variants.

$ Betas represent changes in standard deviations of the standardized log-transformed 25OHD level

Table 5. Main findings of the GWAS meta-analysis

SNP	Chr	Candidate Gene	Effect allele	Effect allele frequency	Beta$	P-value	N
rs117913124	11	*CYP2R1*	A	0.025	-0.43	1.5×10^{-88}	41,336
rs3819817	12	*HAL*	C	0.45	0.04	3.2×10^{-10}	41,071
rs2277458	14	*GEMIN2*	G	0.21	-0.05	6.0×10^{-09}	39,746

$ Betas represent changes in standard deviations of the standardized log-transformed 25OHD levels, while controlling for age, sex, BMI and seas

2.10.2 Figures

Figure 1: Schematic of the discovery single variant meta-analysis

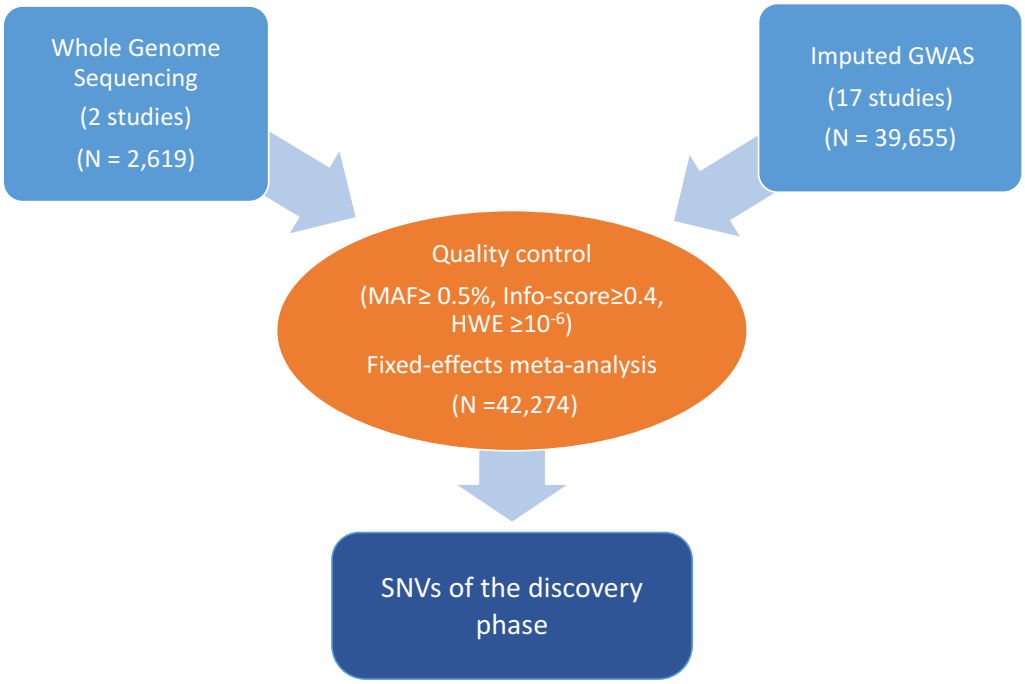


Figure 2: Discovery single-variant meta-analysis.

Legend: A. Quantile-quantile plot for the single SNV meta-analysis. B. Manhattan plot of the meta-analysis. The plot depicts variants with MAF > 0.5% across the 22 autosomes against the –log10 p-value from the meta-analysis of 19 cohorts, which included 42,274 individuals.

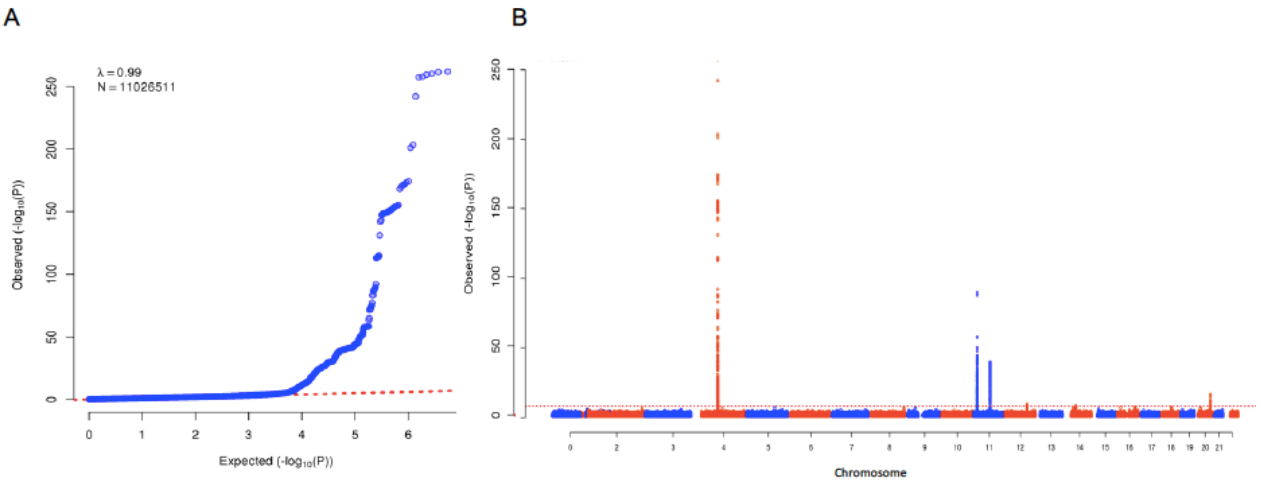


Figure 3: Forest Plot by Cohort for rs117913124 and Forest Plot of the rs117913124 and the Previously Described Common 25OHD-related Variants from Discovery Meta-analysis

Legend: A. Forest plot of estimates from all 19 studies for the low-frequency *CYP2R1* variant rs117913124 B. Forest-plot of the effect of the four common SUNLIGHT variants and of the *CYP2R1* low-frequency variant rs117913124 on log-transformed 25OHD levels.

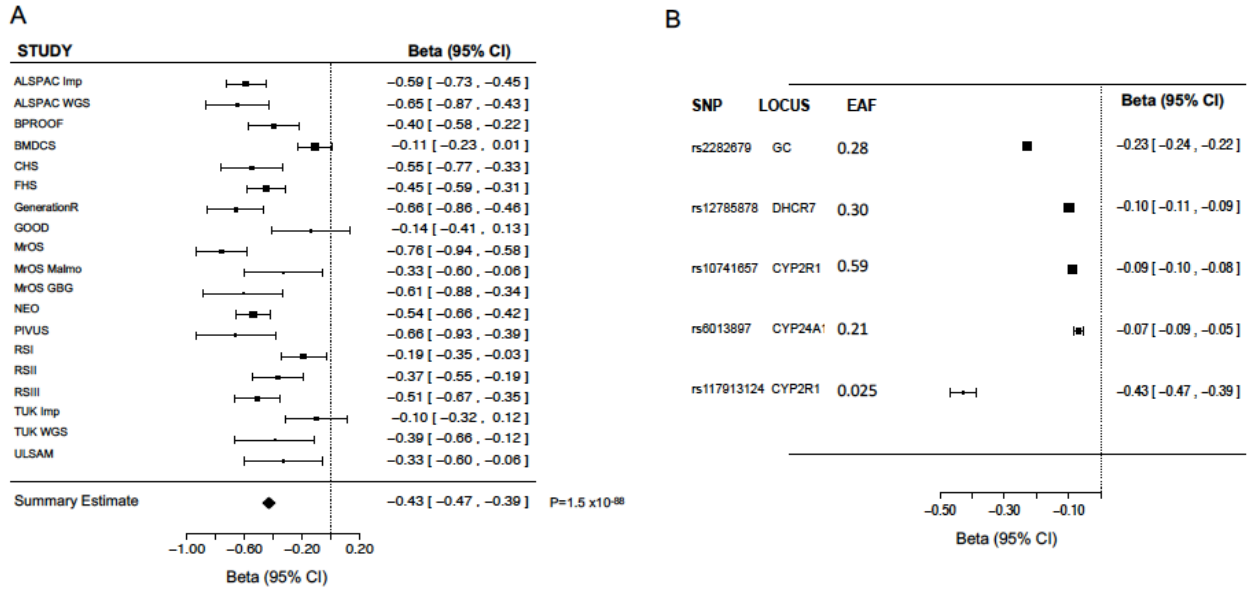


Figure 4: Association Signals from 11p.15.2

Legend: A. Snapshot from the UCSC genome-browser including the top low-frequency SNVs (see **Table 2**) and the lead common variant rs10741657 at the *CYP2R1* locus. The position of rs117913124 is highlighted in light blue. B. Regional disequilibrium plot showing the rs117913124 (purple dot), its perfect proxy rs11670203 (red dot) and the other genome-wide significant SNVs in the same locus (blue and green dots). The plot depicts SNVs within 1 Mb of a locus' lead SNV (x-axis) and their associated meta-analysis p value (-log10) (for more details see **Table S10**). SNVs are color coded according to r^2 with the lead SNV (labelled, r^2 calculated from UK10K whole genome sequencing dataset). Recombination rate (blue line), and the position of genes, their exons and the direction of transcription are also displayed (below plot). C. Linkage disequilibrium plot indicating the r^2 values between the SNVs of **Table 2** (top low-frequency variants) and between these low-frequency SNVs and the lead common variant (rs107416570) at the same *CYP2R1* locus (r^2 calculated from the 1000 Genomes dataset).

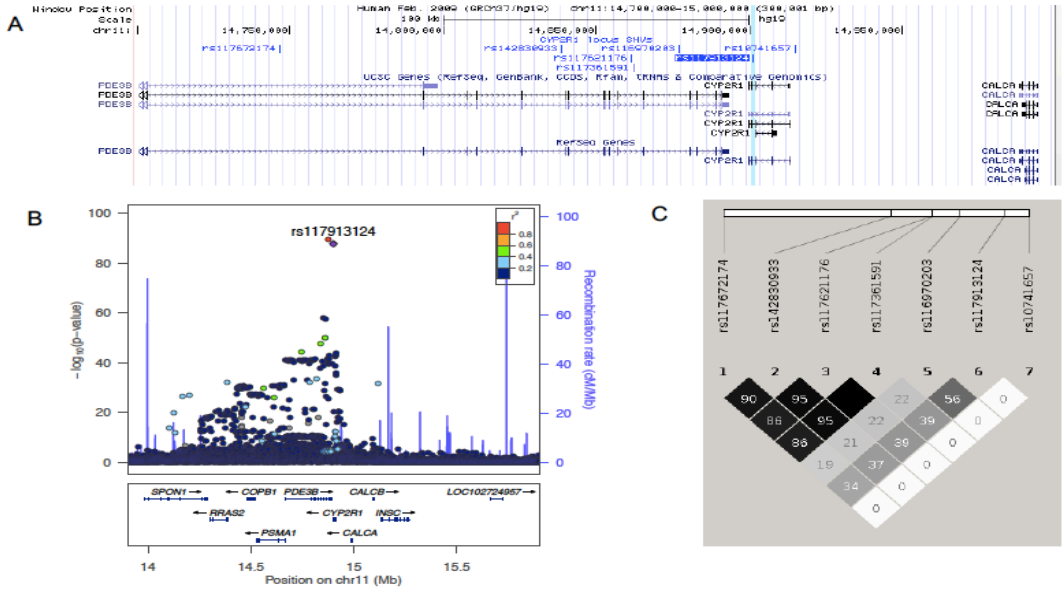


Figure 5: Effect of the rs117913124 on Vitamin D Insufficiency

Legend: Forest-plot of the effect of the low-frequency *CYP2R1* variant rs117913124 on vitamin D insufficiency in 4 studies.

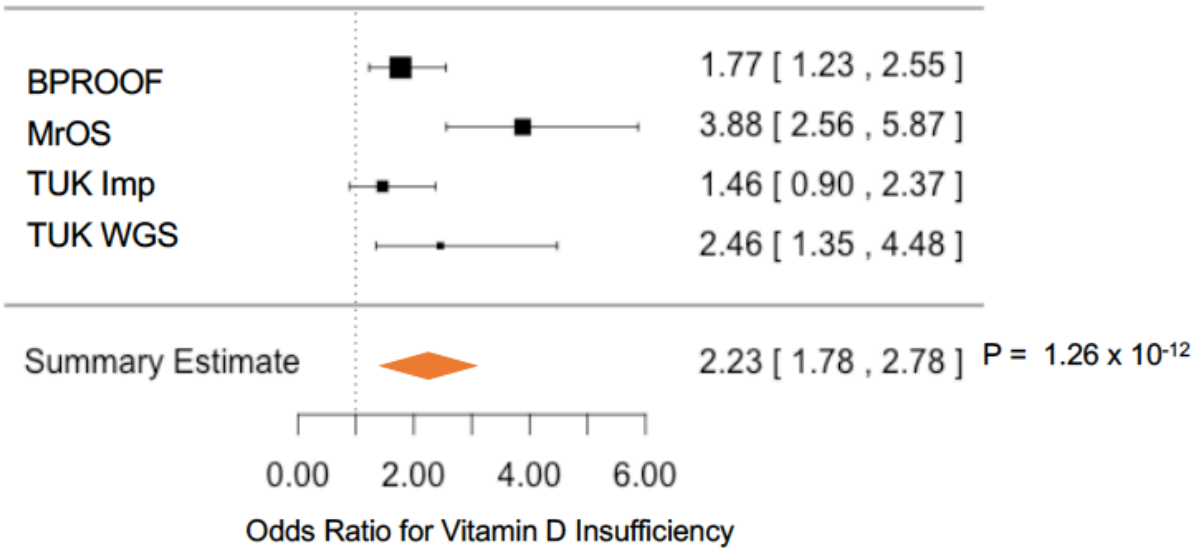


Figure 6: Association Signals from Chromosomes 12 and 14

Legend: Forest plots with A. estimates for the chromosome 12 common variant rs3819817 and B. estimates for the chromosome 14 common variant rs2277458 from all 19 studies of the meta-analysis where both variants were present.

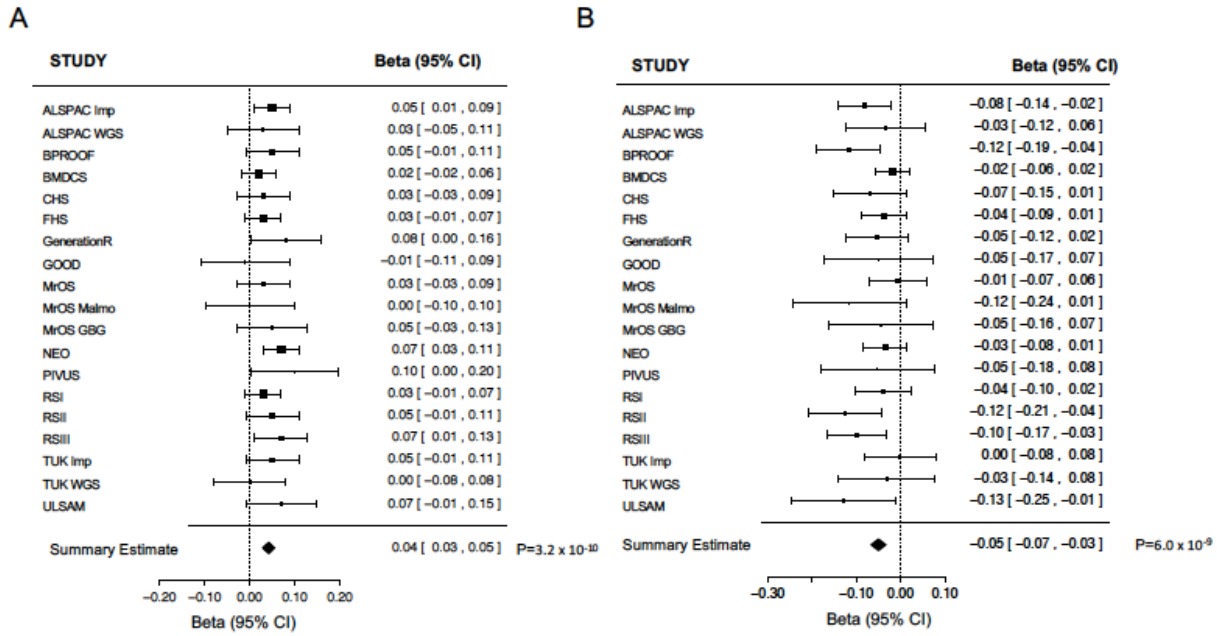
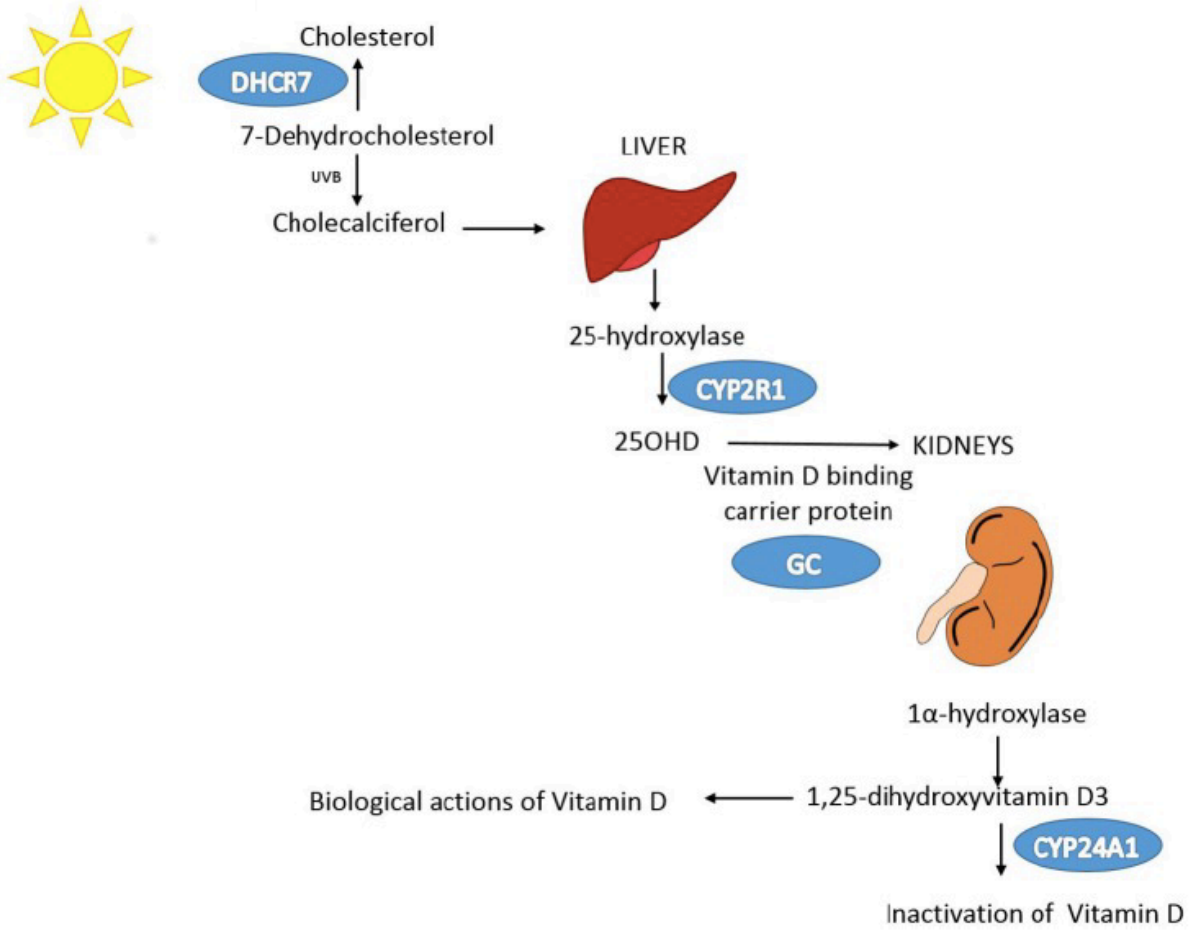


Figure 7: Schematic of the Vitamin D Metabolic Pathway

Legend: UVB: ultraviolet B rays.



2.11 Supplementary Material

Supplementary Methods, Tables and Figures can be downloaded from the open access publication Manousaki et al. in AJHG available here:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC 28757204/

Chapter 3: Mendelian Randomization Studies do not Support a Role for Vitamin D in Coronary Artery Disease

This chapter contains a manuscript published under the same title in Circulation: Cardiovascular Genetics in 2016.

Date of publication: Aug 9 2016; Epub date: Jul 14 2016

Volume 4, pages 349-56. Wolters Kluwer Health©

doi: 10.1161/CIRCGENETICS.116.001396.

PMID: 27418593

Preface: Bridge between Chapter 2 and Chapters 3 and 4

In the previous Chapter, we performed a large GWAS meta-analysis aiming to identify low-frequency and rare variants with large effects on 25OHD levels. In the next two Chapters (3 and 4), we used vitamin D SNPs identified by previous GWAS in two Mendelian randomization studies, testing causal associations between vitamin D and coronary artery disease (in Chapter 3), and atopic outcomes, such as asthma, atopic dermatitis, and IgE levels (in Chapter 4). It is important to underline that these two Mendelian randomization studies were published before the 25OHD GWASes of Chapters 2 and 5, and thus they do not include the novel variants described in these studies. Although the results of these MR studies might have changed by including an enhanced set of genetic instruments, it is important to underline that the 4 common variants from SUNLIGHT, which are used as instruments in these two MR papers, still explain a large part of the variance in 25OHD levels (2.4% vs 4.9% of all 138 conditionally independent variants

identified in Chapter 5). Also, since these SNPs map directly in or near genes involved in vitamin D synthesis or metabolism, they minimize the risk of pleiotropy, a considerable limitation of the MR study design. Thus, we believe that these studies continue to provide valid evidence refuting causal effects of low vitamin D levels in the studied outcomes. It is also important to note that the results of the MR study on coronary artery disease presented in Chapter 3 were replicated in a large randomized controlled trial published two years later (Vitamin D and Omega-3 Trial-VITAL), which reported a hazard ratio very comparable to the MR estimate provided in our Circulation: Cardiovascular Genetics paper.

Title page

Title: Mendelian Randomization Studies do not Support a Role for Vitamin D in Coronary Artery Disease

Short Running Title: Vitamin D deficiency is not a cause of CAD

AUTHORS: Despoina Manousaki, MD[1] Lauren E. Mokry, MSc[1], Stephanie Ross, PhD[1], David Goltzman, MD[2], J. Brent Richards (*), MD, MSc[1,2,3,4]

[1]Centre for Clinical Epidemiology, Department of Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, Canada

[2]Department of Medicine, McGill University Montreal, Quebec, Canada

[3]Department of Human Genetics, McGill University, Montréal, Québec, Canada

[4]Department of Twin Research and Genetic Epidemiology, King's College London, United Kingdom

Corresponding author:

Brent Richards, MD, MSc

Pavillon H-413, Jewish General Hospital

3755 Cote Ste Catherine

Montreal, QC, Canada, H3T 1E2

T: +1 514 340 8222 ext. 4362 /F: +1 514 340 7529

brent.richards@mcgill.ca

Despoina Manousaki, Post-doctoral fellow, Department of Human Genetics, McGill University despoina.manousaki@mail.mcgill.ca

Lauren E. Mokry, MSc candidate, Department of Epidemiology, McGill University lauren.mokry@mail.mcgill.ca

Stephanie Ross, Post-doctoral fellow, Department of Epidemiology, McGill University stephanie.ross@mail.mcgill.ca

David Goltzman, Professor, Departments of Medicine and Physiology, McGill University david.goltzman@mcgill.ca

Brent Richards, Associate Professor of Medicine, Departments of Medicine, Human Genetics, Epidemiology and Biostatistics McGill University brent.richards@mcgill.ca

Address of all authors except DG: Centre for Clinical Epidemiology, Department of Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, 3755 Cote Ste Catherine, Montreal, QC, Canada, H3T 1E2

Address of DG: McGill University Health Centre—Glen Site, 1001 Decarie Blvd, Room EM1.3220, Montreal, PQ, Canada H4A 3J1

Journal Subject Terms: Genetic Association Studies, Biomarkers, Coronary Artery Disease, Epidemiology, Risk Factors

Number of tables:3, number of figures:2, Supplemental Material available

3.1 Abstract

Background: Observational studies support a possible association between decreased vitamin D levels and risk of coronary artery disease (CAD), however it remains unclear whether this relationship is causal. We aimed to evaluate whether genetically lowered vitamin D levels influence the risk of CAD using a Mendelian randomization (MR) approach.

Methods and Results: In this two-stage MR study, we first identified SNPs associated with 25-hydroxyvitamin D (25OHD) levels in the SUNLIGHT consortium (n=33,996), then tested them for possible violation of MR assumptions. A count of risk alleles was tested for association with 25OHD levels in a separate cohort (n = 2,347). Alleles were weighted by their relative effect on 25OHD and tested for their combined effect on CAD in the CARDIoGRAM study (22,233 cases/64,762 controls). Four SNPs were identified to be associated with 25OHD levels, all in or near genes implicated in 25OHD synthesis, transport or metabolism. A count of these risk alleles was strongly associated with 25OHD (n= 2,347, F-test statistic = 49.7, $P = 2x10^{-12}$). None of the SNPs associated with 25OHD levels were associated with CAD (all P-values >0.6). The MR odds ratio for CAD was 0.99 (95% CI 0.84-1.17, P=0.93, I^2=0) per standard deviation decrease in log transformed 25OHD levels. These results persisted after sensitivity analyses for population stratification and pleiotropy.

Conclusions: Genetically lowered 25OHD levels were not associated with increased risk of CAD in a large, well-powered study, suggesting that previous associations between circulating 25OHD levels and CAD are possibly confounded or due to reverse causation.

Key words: Vitamin D, Coronary Artery Disease, Mendelian Randomization, Genome-Wide Association Study

3.2 Introduction

Despite increasing public awareness and major therapeutic progress, coronary artery disease (CAD) remains the international leading cause of morbidity and mortality [1]. Growing evidence suggests that vitamin D deficiency is associated with CAD development [2-4]. Specifically, the results from many[5-8], although not all [9-12], observational studies investigating this relationship suggest that low levels of vitamin D, as measured by serum total 25-hydroxy vitamin D (25OHD) concentrations, are associated with increased risk of CAD-related outcomes. These findings may have important public health impact since vitamin D insufficiency and deficiency is common (affecting one third of Americans in 2001-2006)[13], and vitamin D replacement therapy is relatively safe and inexpensive.

Such findings have motivated inquiry into the biological plausibility of this relationship and some evidence has suggested that vitamin D may influence CAD directly through the vitamin D receptor in smooth muscle cells of the cardiac vasculature, or indirectly by promoting calcium absorption at the expense of lipid absorption or excretion in the gut [14]. Other mechanistic explanations have been proposed, including endothelial dysfunction from lack of adequate vitamin D, vascular compliance impairment due to smooth muscle changes, enhanced inflammation, effects related to high levels of PTH, or the renin-angiotensin system[15-17]. Finally, some studies have suggested that vitamin D supplementation could lower vascular risk by improving glucose tolerance and/or inhibiting inflammatory components in the metabolic syndrome[18-20].

Recent meta-analyses of observational studies have reported, on balance, a significant relationship between 25OHD levels and composite cardiovascular events [21-23]. Meta-analyses of the few randomized controlled trials that have assessed the effect of vitamin D administration on risk of CAD have not produced conclusive results, possibly because of heterogeneity of outcomes and interventions[20 21 23].

Given this uncertainty, clinical practice guidelines do not support vitamin D supplementation for CAD risk reduction and conclude that additional research, particularly from randomized trials, is needed[24 25]. Given the high prevalence of low vitamin D levels such randomized controlled trials would be relevant, yet large, long-term RCTs are difficult to fund since vitamin D therapy cannot be patented.

In the absence of high-quality RCT data and given the inconclusive results from the existing observational studies, the principles of Mendelian randomization (MR) can be applied to strengthen, or refute, the causality of biomarkers in disease etiology[26]. MR analysis uses genetic associations to test the relationship between biomarkers, such as 25OHD, and risk of disease, such as CAD. This approach, which is conceptually similar to an RCT, is based on the principle that genetic variants are randomly allocated at gamete formation and consequently these genetic variants are independent of confounding factors that bias observational studies. Further, MR is free of reverse causation since genotypes are assigned prior to the onset of CAD. A recent MR study found no relationship between 25OHD levels and CAD risk, but only used two of four well-validated 25OHD genetic loci in a smaller sample size of 14,455 cases[27].

In the present study, we adopted an MR design to estimate the effect of genetically lowered 25OHD levels on CAD susceptibility combining data from several large-scale studies. We first selected genome-wide significant single nucleotide polymorphisms (SNPs) as identified by the SUNLIGHT consortium, the largest genome-wide association study (GWAS) published to date for 25OHD levels (n=33,995)[28]. Next, we used the estimates of the effect of each of these SNPs upon 25OHD levels in the Canadian Multicentre Osteoporosis Study (CaMos)[29], which we had generated in a previous MR study[30]. The CaMos cohort was used since effect sizes could not be estimated from the SUNLIGHT consortium, because of different 25OHD measurement methods used in SUNLIGHT cohorts. Finally, we applied the principles of MR to test whether a lifetime of genetically lowered 25OHD levels influence CAD risk using data from the largest meta-analysis of GWAS studies assessing the CAD risk, the Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM, n=86,995)[31].

3.3 Methods

3.3.1 SNP selection and Data Sources

We used the SUNLIGHT Consortium[28], a genome-wide association (GWAS) study consisting of 33,996 individuals of European descent from 15 cohorts, to obtain genetic variants associated with 25OHD levels at a genome-wide significant level ($P < 5x10^{-8}$). 25OHD levels in this study were measured either by radioimmunoassay, chemiluminescent assay, ELISA or mass spectrometry. Given that different cohorts used different methods to measure 25OHD levels, results were combined across cohorts in the SUNLIGHT Consortium using Z-score-weighted meta-analysis. While other GWAS for 25OHD have been reported[32-34], the SUNLIGHT consortium offers the largest sample size and no other loci have been genome-wide significant, to our knowledge.

Data from the Canadian Multicentre Osteoporosis Study (CaMos)[29] were used to estimate the effect of each genome-wide significant SNP on 25OHD levels, since the effect of each SNP upon 25OHD levels could not be used from the SUNLIGHT Consortium, due to the Z-score meta-analytic approach employed. CaMos is a large population-based cohort, and was amongst the largest included in the replication phase of the SUNLIGHT consortium. It includes 2,347 individuals who were genotyped using TaqMan genotyping at the same genome-wide significant vitamin D loci found in the SUNLIGHT consortium. Serum total 25OHD was measured using chemiluminescent immunoassay technology. We have previously reported the effects of these SNPs in CaMos[30].

To obtain precise estimates for the association of 25OHD on CAD, we tested the effect of each genome-wide significant SNP for vitamin D levels in Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM)[31], a meta-analysis of 22 GWAS studies of European descent imputed to HapMap 2 involving 22,233 cases and 64,762 controls. CAD outcomes were defined as one of the following: MI, >50% stenosis in at least one coronary vessel at angiography, history of percutaneous transluminal coronary angioplasty or coronary artery bypass graft surgery, angina or death due to CAD. Genotyping in individual discovery GWA studies was carried out on Affymetrix or Illumina platforms. Data from the more recent and larger CARDIoGRAMplusC4D Metabochip meta-analysis[35], involving 63,746 cases and 130,681 controls, could not be used in the present MR analysis, as no estimates for the SNPs of interest for vitamin D levels were found on the Metabochip, whereas all four SNPs were present in the original CARDIoGRAM GWAS. Moreover, no proxies of the four SNPs with an appropriate level of r^2 (>0.8) were found in Metabochip.

3.3.2 SNP Validation

MR analysis requires genetic variants used as instrumental variables to be evaluated for the several MR assumptions: linkage disequilibrium, population stratification and pleiotropy. For the four SUNLIGHT 25OHD SNPs, linkage disequilibrium and population stratification has been already tested in our previous MR study[30]. Our population stratification assessment showed that only the *DHCR7* SNP was strongly associated with non-European ancestry in the CaMos cohort (P=2.7 $X10^{-13}$). In the present study, we performed an assessment for pleiotropy, in order to assure that the chosen SNPs do not exert effects on CAD through biologic pathways independent of 25OHD

levels. Although findings from the 1958 British Birth Cohort [36] did not support any association between the SUNLGHT 25OHD SNPs and relevant pleiotropic pathways, such as sun exposure, time outside, physical activity, fish oil consumption, smoking, alcohol consumption and BMI, some of these factors could act at least partially through the vitamin D pathway. In order to assess for additional possible pleiotropy, we looked at the association between the vitamin D-related SNPs and major risk factors for CAD: LDL cholesterol, systolic and diastolic blood pressure, as well as diagnosis of type 2 Diabetes and BMI in four large GWAS consortia (the Global Lipids Genetic Consortium[37], the International Consortium for Blood Pressure consortium (ICBP)[38], the DIabetes Genetics Replication And Meta-analysis consortium (DIAGRAM)[39] and the The Genetic Investigation of ANthropometric Traits (GIANT)[40] respectively). These CAD risk factors may represent possible pleiotropic pathways if the 25OHD-associated SNPs influence CAD through these risk factors, independently of 25OHD. We further explored for pleiotropy by conducting a literature search of gene name and gene mutation to identify published possible pleiotropic mechanisms for any of the selected SNPs and CAD.

3.3.3 Association of 25OHD-Associated SNPs with CAD Susceptibility

In order to increase study power and obtain the most precise estimates of the association of 25OHD-associated SNPs upon risk of CAD, we used summary-level data from the CARDIoGRAM study (as described above). We assessed whether each SNP was associated with risk of CAD, applying a Bonferroni correction, where statistical significance was declared at $P \leq 0.05/4$ since four SNPs were associated with 25OHD levels from the SUNLIGHT consortium.

3.3.4 Mendelian Randomization Estimates

We assessed the effects of the SNPs upon risk of CAD, weighting the effect of each SNP by the magnitude of its effect upon 25OHD levels using a two-sample MR approach[41]. According to this study design, the independent SNPs were used to evaluate the association of exposure to genetically lowered 25OHD based on data from one study (CaMos) with CAD risk in another study (CARDIOGRAM). The SNP alleles were aligned in the two studies, to ensure that the estimates represent the effects of 25OHD decreasing alleles for each SNP. These individual estimates were then pooled using statistically efficient estimators formally analogous to those of inverse-variance weighted meta-analysis [42]. We next meta-analyzed the estimates obtained from individual 25OHD decreasing alleles, using a fixed-effects model with an I^2 estimate to account for heterogeneity in the effect size[43 44.] The effect-size for the meta-analysis is reported in the main results as the effect of a standard deviation (SD) change in natural log-transformed 25OHD levels, since this metric is more interpretable than an arbitrary difference. Next, we undertook power calculations[45] to test whether our study was adequately powered to detect clinically relevant changes in CAD risk.

To better understand the meaning of a one standard deviation change in natural log-transformed 25OHD levels, we report data from a previous MR study published from our group[30]. In this study[30], the effect of 1 SD increase in log-transformed 25OHD levels on 25OHD levels in vitamin D sufficient individuals (defined as individuals with 25OHD levels between 50 and 75nmol/l) was 35.6 nmol/L. The same effect on vitamin D insufficient individuals (25OHD levels between 25 and 50nmol/l) was 23.72 nmol/L and on vitamin D deficient individuals (25OHD levels

<25nmol/l) was 11.86 nmol/l. Therefore, in the vitamin D insufficient and sufficient groups, a 1 SD increase in log-transformed 25OHD levels results in normalization of 25OHD levels, effect comparable to that achieved with supplementation[46].

3.3.5 Sensitivity Analyses

We then recalculated our MR estimates after exclusion of SNPs potentially influenced by pleiotropy or population stratification. Since SNPs associated with 25OHD levels in the SUNLIGHT consortium influence either 25OHD synthesis or 25OHD metabolism[47], we elected to perform a stratified MR analysis where SNPs involved in either 25OHD synthesis or metabolism were analyzed separately.

3.4 Results

3.4.1 SNP Selection and Validation

SNP selection

All the data sources used in the present study are generated from populations of European descent (**Fig. 1)**. The SUNLIGHT Consortium identified four SNPs as genome-wide significant for 25OHD levels [28]: rs2282679 in *GC* (association with 25OHD P = $1.9x10^{-109}$), rs12785878 near *DHCR7* (P = $2.1x10^{-27}$), rs10741657 near *CYP2R1* (P = $3.3x10^{-20}$) and rs6013897 in *CYP24A1* (P = 6.0×10^{-10}). In addition to this strong statistical evidence of association, all SNPs map to genes implicated in the modulation of 25OHD levels through distinct mechanisms, and more specifically transport (*GC*), synthesis (*DHCR7*), hepatic hydroxylation (*CYP2R1*) and catabolism (*CYP24A1*)[47]. Notably, all 4 SNPs lie in intergenic or intronic regions, yet the exact effect of each SNP on these enzymes is unknown. Nevertheless, all SNPs reside near genes strongly implicated in vitamin D synthesis or metabolism[47].

Linkage Disequilibrium and Pleiotropy Assessment

We found no evidence of LD between any of these SNPs (all pairwise $r^2 \leq 0.01$) in the 1000 Genomes Project CEU samples [48]. Of note, only two of our SNPs, rs10741657 and rs12785878 were located on the same chromosome, which greatly decreases risk of confounding by LD. Among the five vitamin D- associated SNPs reported in the GWAS catalog, only the rs3829251 on chromosome 11 was not reported as genome-wide significant in the SUNLIGHT study and thus not included as instrumental variable in our MR analysis. We did not find evidence of LD between this SNP and our two chromosome 11 SNPs rs10741657 and rs12785878 (r^2 <0.2 for both).

We undertook a literature review for possible pleiotropic pathways influencing cardiometabolic traits, assessing associations between the four SNPs with known CAD risk factors such as hypertension, hyperlipidemia, type 2 diabetes and obesity. We found no evidence for pleiotropic mechanisms for the vitamin D metabolism SNPs: rs10741657 (*CYP2R1)* and rs6013897 (*CYP24A1)*. Interestingly, a GWAS study by Shen et al[49] has demonstrated a marginal association between another SNP (rs2762939) in the *CYP24A1* gene and coronary artery calcification with a p-value of 2.9×10^{-6}, but there was no correlation between this SNP and 25OHD levels in the SUNLIGHT Consortium[50]. Also, we did not detect any LD between the rs2762939 SNP and the 25OHD-associated *CYP24A1* SNP, rs10741657. A recent MR study using as instrumental variables the rs12785878 (*DHCR7*) and rs10741657 (*CYP24A1*), along with two other SNPs on the *GC* and *CYP24A1* genes*,* did not demonstrate any association with type 2 diabetes [51].

Although it has been argued that vitamin D binding protein, encoded by *GC*, can act independently of vitamin D to produce clinical phenotypes, this does not appear to be the case for CAD[52]. For rs12785878 (*DHCR7*), a large MR study on vitamin D levels and blood pressure showed a small but significant association with hypertension[53] (OR 0.92, 95% CI 0.87-0.97, p=0.001), a major risk factor for CAD[54]. By querying the association of rs12785878 in the International Consortium for Blood Pressure consortium (ICBP)[38], a large GWAS meta-analysis of 200,000 subjects of European descent, we found that this SNP was not associated with systolic or diastolic blood pressure (p=0.703 and p=0.121 respectively). Genetic variation in *DHCR7* also appears to cause Smith-Lemli-Opitz syndrome, a clinical phenotype relating to cholesterol deficiency. Given that

several studies suggest an inter-dependence of cholesterol and vitamin D pathways in the etiology of CAD[55], we queried the association of rs12785878, in the largest publically available GWAS consortium results for lipids, the Global Lipids Genetics Consortium[37], and found that this SNP was associated with a minimum p-value of 0.043 across all lipid traits, suggesting that the SNP is weakly associated with cholesterol.

Finally, as part of our assessment for pleiotropy we queried all four SNPs from the SUNLIGHT consortium in the Global Lipids Genetics Consortium[37], the International Consortium for Blood Pressure (ICBP)[38], the DIabetes Genetics Replication And Meta-analysis consortium (DIAGRAM)[39] and the The Genetic Investigation of ANthropometric Traits (GIANT)[40] consortium (**Table 1**). Given the threshold of a P-value of ≤ 0.0125 (0.05/4) set in the context of our study , we found no significant association between the *GC, DHCR7, CYP24A1* and *CYP2R1* SNPs and the clinical outcomes of the four consortia (LDL cholesterol, systolic/diastolic blood pressure, type 2 diabetes and BMI).

Population Stratification Assessment

Given that CAD varies by geographical location and such location might be a surrogate for ancestry, we assessed whether any of the 25-OHD SNPs was associated with non-European ancestry. We have previously demonstrated that only rs1278578 may be associated with non-European ancestry[30] and thus undertook sensitivity analyses excluding this SNP.

3.4.2 Association of SUNLIGHT SNPs with 25OHD Levels

The association of the four SNPs that achieved genome-wide significance for 25OHD levels in the SUNLIGHT consortium with 25OHD levels is described in **Table 2**. Each of these SNPs explained an important proportion of the population-level variance in 25OHD levels, as reflected by the F-statistics. As already shown in our previous study[30], the count of 25OHD decreasing alleles across these four SNPs was strongly associated with lower total 25OHD levels in the 2,347 CaMos participants (**Table 2**).

3.4.3 Association of SUNLIGHT SNPs with CAD Susceptibility

Summary statistics for the four 25OHD-associated SNPs, (rs2282679 at *GC*, rs10741657 at *CYP2R1,* rs12785878 at *DHCR7,* rs6013897 at *CYP24A1)* and their association with CAD was taken from the CARDIoGRAM study. All four 25OHD-decreasing alleles were not associated with risk of CAD (**Table 2**) and the 95% confidence intervals were tight around the null.

3.4.4 Mendelian Randomization Analysis for the Association of 25OHD with CAD Risk

In order to estimate the association of genetically lowered 25OHD upon CAD, we used a fixed-effects model in which all four 25OHD-decreasing alleles of the MR set were included. A decrease in 25OHD levels by one standard deviation on the natural log scale was not associated with CAD and the 95% confidence interval limits were close to the null (OR = 0.99, 95% CI: 0.84-1.17; P= 0.93 $I^2 = 0$) (**Table 3** and **Fig. 2).** We note that since our model included only 4 SNPs, the 95% CIs of the I^2 statistic are wide and consequently heterogeneity cannot be accurately measured using

this parameter. In addition, due to potential effects of population stratification, we undertook a sensitivity analysis by excluding the rs12785878 SNP (*DHCR7)*. Despite removal of this variant, we again observed no association of genetically lowered 25OHD levels with risk of CAD (OR= 1.04, 95% CI: 0.85-1.27; P=0.724; I^2 = 0%, 95% CI: 0%-84.7%) (**Table 3** and Supplemental **Fig. 1**), To further assess the effect of the independent vitamin D pathways on risk of CAD, we analyzed SNPs near genes implicated in 25OHD synthesis (*DHCR7* and *CYP2R1*) and metabolism (*GC*, and *CYP24A1*) separately and found that both were again not associated with increased risk of CAD (Supplemental **Table 1**). Lastly, since the *CYP24A1* SNP has a low F-statistic for 25OHD and may lead to weak instrument bias, and therefore mask a true causal effect[56], we undertook further sensitivity analyses after removing this variant. The results were again consistent (**Table 3**). Given these null results, we undertook a power calculation[45]. Based on our sample size of 86,995 individuals (22,233 cases and 64,762 controls from the CARDIoGRAM study) and setting alpha to 0.05, our study had 98% power to detect an OR of 1.02 for the effect of 25OHD on CAD risk and 100% power to detect and OR of 1.2.

3.5 Discussion

Using summary level data for CAD and total 25OHD levels from large populations of European descent, our study provides evidence against a causal role for vitamin D in CAD susceptibility. These findings suggest that previous observational epidemiologic associations may have been influenced by confounding or reverse causation. The 95% confidence intervals of our summary estimates do not include clinically relevant effects of vitamin D on CAD, despite the large change in vitamin D levels. These results provide no rationale for the use of vitamin D to prevent CAD.

The discrepancy between the findings of many observational studies and of our MR study is likely due to residual confounding. Adiposity predisposes to CAD and lowers 25OHD levels[57]. Indeed, while most observational studies adjust for BMI, very few studies adjust for DXA measured percent fat mass; contrary to body fat, BMI is not specific for adiposity and thus residual confounding cannot be eliminated[58]. This concept is supported by a recent observational study in a lean and physically active population[59], where although a strong association between serum 25OHD and DXA-assessed body fat in both sexes was found, no association between 25OHD concentrations and CV indexes was demonstrated after adjusting for body fat. Physical activity might be another strong confounder in observational studies, because it is associated with both CAD risk and sunlight exposure, which in-turn influences vitamin D status. Most of the existing observational studies accounted for self-reported physical activity, which is, in general, a poor measure of physical activity[60].

Our results are in accordance with the most recent meta-analysis of randomized trials[23], which showed no association between 25OHD and three different cardiovascular outcomes (death, MI and stroke). They are also in agreement with an MR analysis published by our group studying the association of vitamin D binding protein (DBP), a key determinant of 25OHD levels, with the risk of CAD[52]. Using the single polymorphism rs2282679 near the *GC* gene as instrumental variable (whose effect allele was associated with an age- and sex-adjusted decrease in DBP level of 27.4 mg/l), this study investigated the relationship of DBP with multiple cardiometabolic outcomes in the CARDIoGRAM consortium and found no association (OR = 1.02; 95%CI 0.99-1.05; p=0.31). Using an approach related to MR, Jorde et al.[61] examined causal associations of 25OHD with risk of CAD on 9,528 subjects, but could not establish or exclude any causal relationship, possibly because of their relatively small sample size. In agreement with our findings, a recent MR study using 14,455 ischemic heart disease Danish cases[27] showed that genetically lowered 25OHD concentrations were not associated with increased myocardial infarction. Another MR study used 3,231 cases of death from CVD[62] from the same Danish population and reported no association between low 25OHD levels and cardiovascular mortality[62]. The instrumental variables used for both studies were only SNPs near the *DHCR7* and *CYP2R1* genes, involved only in vitamin D synthesis, but did not include either *GC* SNPs, which have been shown to have the largest impact on 25OHD levels[28], or those at *CYPR24A1*. Our study thus provides a more thorough examination of the effects of 25OHD on CAD risk by using a substantially larger sample size, including all 25OHD-associated loci, in a general European-ancestry population.

MR analyses assessing the effect of 25OHD on cardiometabolic outcomes, other than CAD have also been described. We and others have recently provided evidence from MR that low vitamin D levels do not increase insulin resistance or the risk of type 2 diabetes[51], but do increase the risk of type 1 diabetes[63] and blood pressure[53]. Interestingly MR evidence has shown that 25OHD levels are directly influenced by body mass index (BMI), and converse effects are likely to be small[57].

Our analysis has several strengths. First, using data from a large genetic consortium for 25OHD levels (n = 33,996) and CAD risk (up to 22,233 cases and 64,762 controls) has enabled us to more precisely test our study hypothesis than if we had used individual-level data from a small study. The null association between 25OHD and CAD, as reflected in an OR close to 1 and with a tight confidence interval, indicates that our study had enough power to rule out clinically relevant effects of a large change in 25OHD levels. This observation is also supported by the high statistical power of our study to detect a potential effect. Second, previous work has shown that the use of estimates from meta-analytic data for uncorrelated genetic variants are similarly efficient to individual-level data in MR studies[41]. Lastly, the findings from this study represent the association of a life-long exposure to reduced vitamin D levels in the general European population and in the absence of large-scale, long-term RCT data, our findings provide strong evidence against a causal role for low vitamin D levels in CAD susceptibility.

Our study also has limitations. Pleiotropy is difficult to exclude in any MR study; however, our sensitivity analyses demonstrated no evidence of pleiotropic effects. The null result could also be explained by canalization, which is defined as compensatory feedback interactions[26 64 65]. Similar

to previous studies, our MR analysis might be limited in its ability to elucidate the causal role of biologically active vitamin D, reflected by the levels of the active metabolite 1,25-dihydroxyvitamin D ($1,25[OH]_2D$). Thus, although genetically lowered total 25OHD levels do not appear to be associated with increased risk of CAD, our study still leaves open the possibility that reduced lifelong total 25OHD is not associated with reduced production of $1,25[OH]_2D$. In this respect, concentrations of total 25OHD and $1,25[OH]_2D$ are weakly correlated ($r<0.3$)[66]. However, $1,25[OH]_2D$ remains understudied because of its short half-life and its low concentration in blood[66]. Furthermore, although developmental differences in prenatal and postnatal expression of the vitamin D receptor through which $1,25[OH]_2D$ acts has been well demonstrated and reflect developmental differences in vitamin D function[67], developmental differences in synthesis of $1,25[OH]_2D$ have not been well studied. Consequently, it is also possible that in lifelong low total 25OHD states, a compensatory increase in conversion of free 25OHD to $1,25[OH]_2D$ could result in a new steady state with low free 25OHD but normal $1,25[OH]_2D$; however, the same compensation might not occur with acquired 25OHD deficiency later in life or the mechanism of this compensation may be different and more deleterious to the vasculature. Finally, the results of our MR study are generalizable only in populations of European ancestry and only in generally healthy adults.

In conclusion, evidence from the largest existing genetic consortia, provides no support for a causal role for 25OHD levels in risk of CAD in individuals of European descent. Instead, association of 25OHD levels with CAD may be attributable to confounding by lifestyle factors such as obesity

and physical inactivity, which may provide more fruitful targets for cardiovascular disease prevention than vitamin D supplementation.

3.6 Acknowledgements

We wish to kindly thank the SUNLIGHT Consortium, the CaMos study and the CARDIoGRAM Consortium for access to their data. "Data on coronary artery disease / myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG''

3.7 Author contributions

JBR conceived the experiment, with input from DG. JBR. LEM, SR & DM conducted the analyses. JBR & DM wrote the first draft of the manuscript. JBR is the guarantor of the manuscript.

3.8 Funding

This work was supported by the Canadian Institute of Health Research, the Canadian Foundation for Innovation, The Fonds de la Recherche en Santé Québec, the Lady Davis Institute and the Jewish General Hospital. URL: http://www.cihr-irsc.gc.ca/e/193.html.The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

3.9 Disclosures

None

3.10 References

1. Eyre H, Kahn R, Robertson RM, Clark NG, Doyle C, Hong Y, et al. Preventing cancer, cardiovascular disease, and diabetes: a common agenda for the American Cancer Society, the American Diabetes Association, and the American Heart Association. *Stroke*. 2004;35:1999-2010.

2. Artaza JN, Mehrotra R, Norris KC. Vitamin D and the cardiovascular system. *Clin J Am Soc Nephrol*. 2009;4:1515-22.

3. Giallauria F, Milaneschi Y, Tanaka T, Maggio M, Canepa M, Elango P, et al. Arterial stiffness and vitamin D levels: the Baltimore longitudinal study of aging. *J Clin Endocrinol Metab*. 2012;97:3717-3723.

4. Lai H, Fishman EK, Gerstenblith G, Brinker JA, Tong W, Bhatia S, et al. Vitamin D deficiency is associated with significant coronary stenoses in asymptomatic African American chronic cocaine users. *Int J Cardiol*. 2012;158:211-216.

5. Wang TJ, Pencina MJ, Booth SL, Jacques PF, Ingelsson E, Lanier K, et al. Vitamin D deficiency and risk of cardiovascular disease. *Circulation*. 2008;117:503-511.

6. Giovannucci E, Liu Y, Hollis BW, Rimm EB. 25-hydroxyvitamin D and risk of myocardial infarction in men: a prospective study. *Arch Intern Med*. 2008;168:1174-1180.

7. Dobnig H, Pilz S, Scharnagl H, Renner W, Seelhorst U, Wellnitz B, et al. Independent association of low serum 25-hydroxyvitamin d and 1,25-dihydroxyvitamin d levels with all-cause and cardiovascular mortality. *Arch Intern Med*. 2008;168:1340-1349.

8. Melamed ML, Michos ED, Post W, Astor B. 25-hydroxyvitamin D levels and the risk of mortality in the general population. *Arch Intern Med*. 2008;168:1629-1637.

9. Marniemi J, Alanen E, Impivaara O, Seppanen R, Hakala P, Rajala T, et al. Dietary and serum vitamins and minerals as predictors of myocardial infarction and stroke in elderly subjects. *Nutr Metab Cardiovasc Dis*. 2005;15:188-197.

10. Cawthon PM, Parimi N, Barrett-Connor E, Laughlin GA, Ensrud KE, Hoffman AR, et al. Serum 25-hydroxyvitamin D, parathyroid hormone, and mortality in older men. *J Clin Endocrinol Metab*. 2010;95:4625-4634.

11. Kendrick J, Targher G, Smits G, Chonchol M. 25-Hydroxyvitamin D deficiency and inflammation and their association with hemoglobin levels in chronic kidney disease. *Am J Nephrol*. 2009;30:64-72.

12. Ross AC, Manson JE, Abrams SA, Aloia JF, Brannon PM, Clinton SK, et al. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. *J Clin Endocrinol Metab*. 2011;96:53-58.

13. Looker AC, Johnson CL, Lacher DA, Pfeiffer CM, Schleicher RL, Sempos CT. Vitamin D status: United States, 2001-2006. *NCHS Data Brief*. 2011:1-8.

14. Chen S, Law CS, Grigsby CL, Olsen K, Hong TT, Zhang Y, et al. Cardiomyocyte-specific deletion of the vitamin D receptor gene results in cardiac hypertrophy. *Circulation*. 2011;124:1838-1847.

15. Sepulveda JL, Mehta JL. C-reactive protein and cardiovascular disease: a critical appraisal. *Curr Opin Cardiol*. 2005;20:407-416.

16. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9:356-369.

17. Li YC, Kong J, Wei M, Chen ZF, Liu SQ, Cao LP. 1,25-Dihydroxyvitamin D(3) is a negative endocrine regulator of the renin-angiotensin system. *J Clin Invest*. 2002;110:229-238.

18. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet*. 2009;2:73-80.

19. Oh JH, Kim SH, Kim JH, Shin YH, Yoon JP, Oh CH. The level of vitamin D in the serum correlates with fatty degeneration of the muscles of the rotator cuff. *J Bone Joint Surg Br*. 2009;91:1587-1593.

20. Bolland MJ, Grey A, Gamble GD, Reid IR. The effect of vitamin D supplementation on skeletal, vascular, or cancer outcomes: a trial sequential meta-analysis. *Lancet Diabetes Endocrinol*. 2014;2:307-320.

21. Grandi NC, Breitling LP, Brenner H. Vitamin D and cardiovascular disease: systematic review and meta-analysis of prospective studies. *Prev Med*. 2010;51:228-233.

22. Pittas AG, Chung M, Trikalinos T, Mitri J, Brendel M, Patel K, et al. Systematic review: Vitamin D and cardiometabolic outcomes. *Ann Intern Med*. 2010;152:307-314.

23. Elamin MB, Abu Elnour NO, Elamin KB, Fatourechi MM, Alkatib AA, Almandoz JP, et al. Vitamin D and cardiovascular outcomes: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2011;96:1931-1942.

24. Rosen CJ, Adams JS, Bikle DD, Black DM, Demay MB, Manson JE, et al. The nonskeletal effects of vitamin D: an Endocrine Society scientific statement. *Endocr Rev*. 2012;33:456-492.

25. Pilz S, Verheyen N, Grubler MR, Tomaschitz A, Marz W. Vitamin D and cardiovascular disease prevention. *Nat Rev Cardiol*. 2016;13:404-417

26. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27:1133-1163.

27. Brondum-Jacobsen P, Benn M, Afzal S, Nordestgaard BG. No evidence that genetically reduced 25-hydroxyvitamin D is associated with increased risk of ischaemic heart disease or myocardial infarction: a Mendelian randomization study. *Int J Epidemiol*. 2015;44:651-661.

28. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet*. 2010;376:180-188.

29. Faye LL, Sun L, Dimitromanolakis A, Bull SB. A flexible genome-wide bootstrap method that accounts for ranking and threshold-selection bias in GWAS interpretation and replication study design. *Stat Med*. 2011;30:1898-1912.

30. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Leong A, et al. Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. *PLoS Med*. 2015;12:e1001866.

31. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333-338.

32. Engelman CD, Meyers KJ, Ziegler JT, Taylor KD, Palmer ND, Haffner SM, et al. Genome-wide association study of vitamin D concentrations in Hispanic Americans: the IRAS family study. *J Steroid Biochem Mol Biol*. 2010;122:186-192.

33. Lasky-Su J, Lange N, Brehm JM, Damask A, Soto-Quiros M, Avila L, et al. Genome-wide association analysis of circulating vitamin D levels in children with asthma. *Hum Genet*. 2012;131:1495-1505.

34. Ahn J, Yu K, Stolzenberg-Solomon R, Simon KC, McCullough ML, Gallicchio L, et al. Genome-wide association study of circulating vitamin D levels. *Hum Mol Genet*. 2010;19:2739-2745.

35. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25-33.

36. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol*. 2006;35:34-41.

37. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274-1283.

38. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478:103-109.

39. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010;42:937-948.

41. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37:658-665.

42. Patsopoulos NA, Evangelou E, Ioannidis JP. Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. *Int J Epidemiol*. 2008;37:1148-57.

43. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21:1539-1558.

44. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-560.

45. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol*. 2013;42:1497-1501.

46. Aspray TJ, Bowring C, Fraser W, Gittoes N, Javaid MK, Macdonald H, et al. National Osteoporosis Society vitamin D guideline summary. *Age Ageing*. 2014;43:592-595.

47. Dastani Z, Li R, Richards B. Genetic regulation of vitamin D levels. *Calcif Tissue Int*. 2013;92:106-117.

48. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061-1073.

49. Shen H, Bielak LF, Ferguson JF, Streeten EA, Yerges-Armstrong LM, Liu J, et al. Association of the vitamin D metabolism gene CYP24A1 with coronary artery calcification. *Arterioscler Thromb Vasc Biol*. 2010;30:2648-2654.

50. Dastani Z, Richards JB. Is coronary artery calcification at the intersection of vitamin D and coronary artery disease? *Arterioscler Thromb Vasc Biol*. 2010;30:2329-2330.

51. Ye Z, Sharp SJ, Burgess S, Scott RA, Imamura F, InterAct C, et al. Association between circulating 25-hydroxyvitamin D and incident type 2 diabetes: a mendelian randomisation study. *Lancet Diabetes Endocrinol*. 2015;3:35-42.

52. Leong A, Rehman W, Dastani Z, Greenwood C, Timpson N, Langsetmo L, et al. The causal effect of vitamin D binding protein (DBP) levels on calcemic and cardiometabolic diseases: a Mendelian randomization study. *PLoS Med*. 2014;11:e1001751.

53. Vimaleswaran KS, Cavadino A, Berry DJ, LifeLines Cohort Study i, Jorde R, Dieffenbach AK, et al. Association of vitamin D status with arterial blood pressure and hypertension risk: a mendelian randomisation study. *Lancet Diabetes Endocrinol*. 2014;2:719-729.

54. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies C. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002;360:1903-1913.

55. Browne RW, Weinstock-Guttman B, Zivadinov R, Horakova D, Bodziak ML, Tamano-Blanco M, et al. Serum lipoprotein composition and vitamin D metabolite levels in clinically isolated syndromes: Results from a multi-center study. *J Steroid Biochem Mol Biol*. 2014;143:424-433.

56. Inoue A SG. Two-Sample Instrumental Variables Estimators. *The Review of Economics and Statstics*. 2010:557-561.

57. Vimaleswaran KS, Berry DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, et al. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. *PLoS Med*. 2013;10:e1001383.

58. Prentice AM, Jebb SA. Beyond body mass index. *Obes Rev*. 2001;2:141-147.

59. Baker CP, Kulkarni B, Radhakrishna KV, Charyulu MS, Gregson J, Matsuzaki M, et al. Is the Association between Vitamin D and Cardiovascular Disease Risk Confounded by Obesity? Evidence from the Andhra Pradesh Children and Parents Study (APCAPS). *PLoS One*. 2015;10:e0129468.

60. Wareham NJ, Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? *Int J Obes Relat Metab Disord*. 1998;22 Suppl 2:S30-38.

61. Jorde R, Schirmer H, Wilsgaard T, Joakimsen RM, Mathiesen EB, Njolstad I, et al. Polymorphisms related to the serum 25-hydroxyvitamin D level and risk of myocardial infarction, diabetes, cancer and mortality. The Tromso Study. *PLoS One*. 2012;7:e37295.

62. Afzal S, Brondum-Jacobsen P, Bojesen SE, Nordestgaard BG. Genetically low vitamin D concentrations and increased mortality: Mendelian randomisation analysis in three large cohorts. *BMJ*. 2014;349:g6330.

63. Cooper JD, Smyth DJ, Walker NM, Stevens H, Burren OS, Wallace C, et al. Inherited variation in vitamin D genes is associated with predisposition to autoimmune disease type 1 diabetes. *Diabetes*. 2011;60:1624-1631.

64. Mokry LE, Ahmad O, Forgetta V, Thanassoulis G, Richards JB. Mendelian randomisation applied to drug development in cardiovascular disease: a review. *J Med Genet*. 2015;52:71-79.

65. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32:1-22.

66. Zittermann A, Schleithoff SS, Frisch S, Gotting C, Kuhn J, Koertke H, et al. Circulating calcitriol concentrations and total mortality. *Clin Chem*. 2009;55:1163-1170.

67. Goltzman D, Hendy GN, White JH. Vitamin D and its receptor during late development. *Biochim Biophys Acta*. 2015;1849:171-180.

3.11 Tables and Figures

3.11.1 Tables

Table 1: P-values of the association of the SNPs used as instrumental variables with the cardiometabolic outcomes

SNP (gene)	LDL cholesterol (Global Lipids Consortium[37])	Systolic/diastolic blood pressure (ICBP[38])	BMI (GIANT[40])	Type 2 diabetes (DIAGRAM[39])
rs2282679 (*GC*)	0.29	0.47/0.64	0.91	0.76
rs6013897 (*CYP24A1*)	0.53	0.05/0.02	0.61	0.06
rs10741657 (*CYP2R1*)	0.16	0.99/0.59	0.29	0.23
rs12785878 (*DHCR7*)	0.04	0.70/0.12	0.78	0.14

Table 2: Characteristics of SNPs used as instrumental variables

Vitamin D Results										CAD Results (CARDIOGRAM)			
Locus	**Chr**	**25OHD Associated SNP**	**25OHD Decreasing Allele**	**Allele Frequency**	**Effect on 25OHD***	**Standard error of the effect on 25OHD***	**P-value for Association with 25OHD ***	**F-Statistic for 25OHD***	**SUNLIGHT P-value for Association with 25OHD†**	**OR**	**95% CI**	**P-Value for Association with CAD**	**Sample size (cases)**
CYP2R1	11	rs10741657	C	0.62	-0.052	0.012	$1.6x10^{-5}$	18.78	$3.3x10^{-20}$	1.00	0.97-1.03	0.90	80,677 (19,739)
DHCR7	11	rs12785878	G	0.27	-0.056	0.013	$2.0x10^{-5}$	18.29	$2.1x10^{-27}$	0.99	0.96-1.02	0.53	83,295 (21,369)
GC	4	rs2282679	C	0.30	-0.047	0.013	$2.6x10^{-4}$	13.38	$1.9x10^{-109}$	0.98	0.96-1.01	0.31	82,323 (20,728)
CYP24A1	20	rs6013897	A	0.19	-0.027	0.015	$7.7x10^{-2}$	3.13	$6.0x10^{-10}$	0.99	0.96-1.02	0.60	84,099 (21,840)

*Effect on Multiply Adjusted Natural Log-Transformed 25OHD levels, standard error, p-value and F-statistic of the association in the CaMos Cohort (n=2,347)

†P-values derived from the SUNLIGHT Consortium (n=33,996)

Table 3: MR estimate of the association of decreased 25OHD on the risk of CAD

Model	OR (95%CI)*	P-Value	I^2 (95% CI)
Fixed Effects (including the *DHCR7* locus)	0.99 (0.84-1.17)	0.93	0 (0-84.7)
Fixed Effects (excluding the *DHCR7* locus)	1.04 (0.85-1.27)	0.72	0 (0-84.7)
Fixed Effects (excluding the *CYP24A1* locus)	1.00 (0.85-1.19)	0.96	0 (0-84.7)

*OR is expressed as the odds of CAD for a one standard deviation decrease in natural log-transformed 25OHD levels. Estimates of the effect of the SNPs on 25OHD were derived from the CaMos Cohort, and estimates of the effect of the SNPs on CAD were derived from the CARDIOGRAM consortium.

3.11.2 Figures

Fig. 1: Schematic representation of Mendelian Randomization analysis.

The box on the left describes SNPs which were genome-wide significant for 25OHD levels in the SUNLIGHT Consortium. The blue arrow represents the effect of the SNPs on multiply-adjusted, natural log-transformed 25OHD levels using data from the Canadian Multicentre Osteoporosis Study (CaMos). The green arrow represents the causal effect of decreased 25OHD levels on the risk of CAD using data from the CARDIoGRAM consortium.

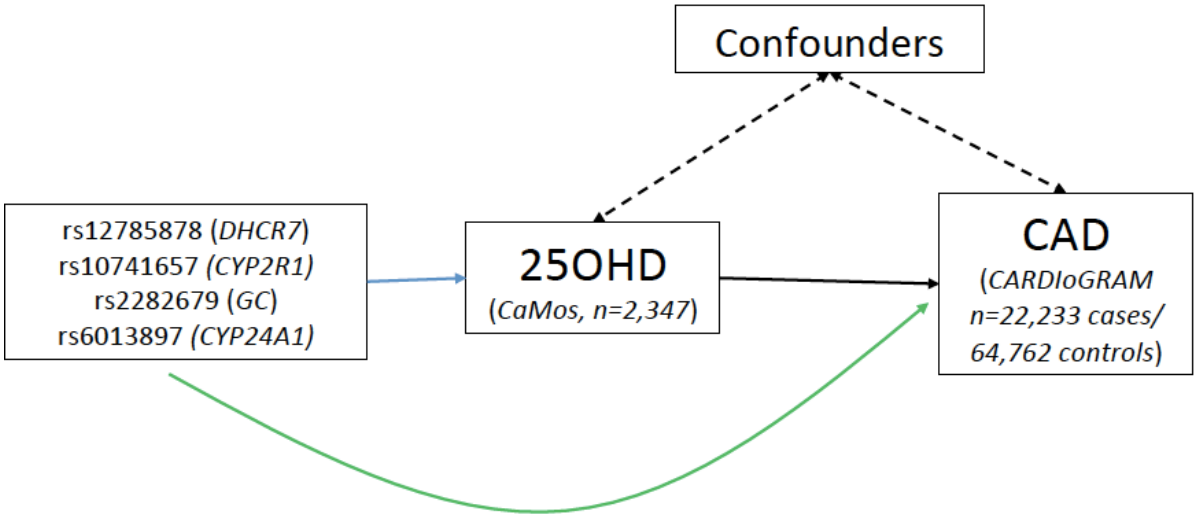
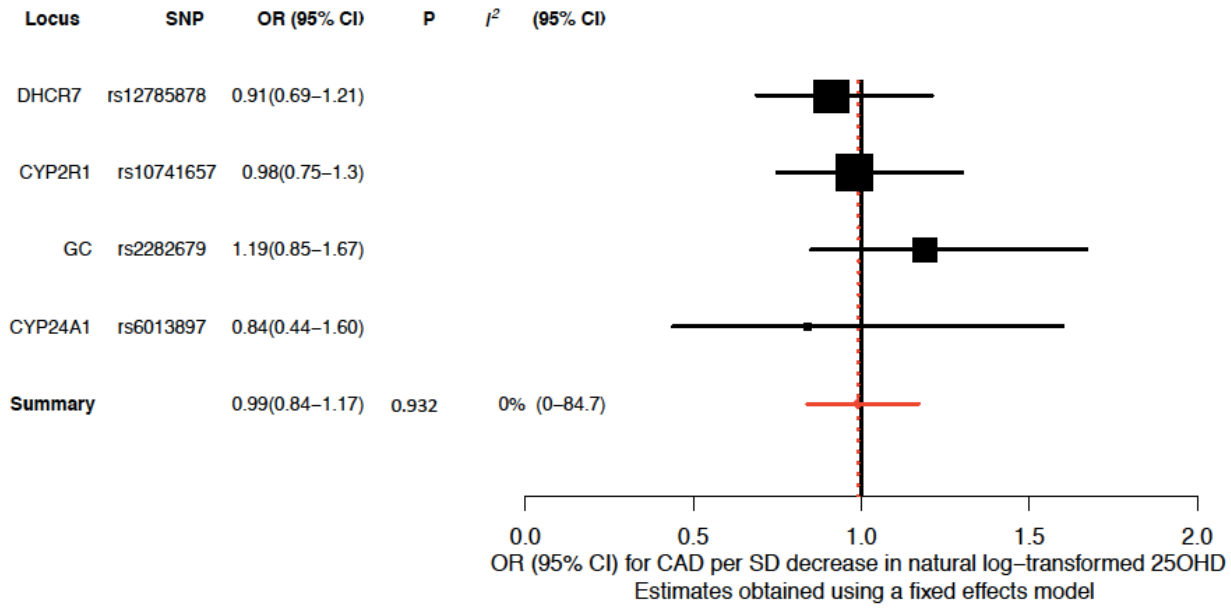


Fig. 2: Mendelian Randomization Estimate of the Association of 25OHD Levels with Risk of CAD.

Estimates obtained from using a fixed-effects model. Estimates of the effect of the SNPs on 25OHD were derived from the CaMos Cohort, and estimates of the effect of the SNPs on CAD were derived from the CARDIOGRAM consortium.



Chapter 4: Vitamin D levels and susceptibility to asthma, elevated IgE levels and atopic dermatitis: a Mendelian Randomization Study.

This chapter contains a manuscript published under the same title in PLoS Medicine in 2017.

Date of publication: May 9 2017

Volume 14 (issue 5): e1002294.

doi: 10.1371/journal.pmed.1002294

PMID:28486474

Preface: Bridge Between Chapter 3 and Chapter 4

See bridge between Chapter 2 and Chapter 3.

Title page

Title: Vitamin D levels and susceptibility to asthma, elevated IgE levels and atopic dermatitis: a Mendelian Randomization Study.

Short Title: Genetically Lowered Vitamin D Levels and Susceptibility to Atopic Disease

Despoina Manousaki,[1] Lavinia Paternoster[2], Marie Standl[3], Miriam F. Moffatt,[4] Martin Farrall[5,6], Emmanuelle Bouzigon[7], David P Strachan[8], Florence Demenais[7], Mark Lathrop[9], William O.C.M. Cookson[4,10], J. Brent Richards (*)[1,11,12]

[1]Centre for Clinical Epidemiology, Department of Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, Canada

[2]MRC Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, UK

[3]Institute of Epidemiology I, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany

[4]National Heart and Lung Institute, Imperial College, UK

[5]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

[6]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

[7]Genetic Variation and Human Diseases Unit, UMR-946, INSERM, Université Paris Diderot, Université Sorbonne Paris Cité, Paris, France

[8]Population Health Research Institute, St George's University of London, London, UK

[9]McGill University and Genome Québec Innovation Centre, Montréal, Québec, Canada.

[10]Royal Brompton and Harefield NHS Foundation Trust, London, UK

[11]Departments of Medicine and Human Genetics, McGill University Montreal, Quebec, Canada

[12]Department of Twin Research and Genetic Epidemiology, King's College London, United Kingdom

*brent.richards@mcgill.ca

Key words: Vitamin D, Asthma, Atopic Dermatitis, IgE level, Mendelian Randomization, Genome-Wide Association Study

Number of tables: 3, number of figures: 6. This article has an Appendix

4.1 Abstract

Introduction: Low circulating vitamin D levels have been associated with risk of asthma, atopic dermatitis and elevated total IgE. These epidemiological associations, if true, would have public health importance, since vitamin D insufficiency is common and correctable.

Methods and Results: To control bias due to confounding and reverse causation we applied Mendelian randomization (MR) to test whether genetically lowered vitamin D levels were associated with risk of asthma, atopic dermatitis or serum IgE levels. Using four single nucleotide polymorphisms (SNPs) strongly associated with 25-hydroxyvitamin D (25OHD) levels in 33,996 individuals, we conducted MR studies to estimate the effect of lowered 25OHD on the risk of asthma (n=146,761), childhood onset asthma (n=15,008), atopic dermatitis (n=40,835) and IgE level (n=12,853) and tested MR assumptions in sensitivity analyses. None of the four 25OHD-lowering alleles were associated with asthma, atopic dermatitis or IgE levels (P-values ≥0.2). The MR odds ratio per standard deviation decrease in log transformed 25OHD was 1.03 (95% CI 0.90,1.19, P=0.63) for asthma, 0.95 (95% CI 0.69,1.31, P=0.76) for childhood-onset asthma, 1.12 (95% CI 0.92,1.37, P=0.27) for atopic dermatitis and the effect size on log-transformed IgE levels was -0.40 (95% CI -1.65, 0.85, P =0.53). These results persisted in sensitivity analyses assessing population stratification and pleiotropy and vitamin D synthesis and metabolism pathways. The main limitations of this study are that the above results do not exclude an association between the studied outcomes and 1,25-dihydoxyvitamin D, the active form of vitamin D, the study was underpowered to detect effects smaller that an OR of 1.33 for childhood asthma, and the analyses were restricted to white populations of European ancestry.

Conclusions: In this study, we found no evidence that genetically-determined reduction in 25OHD levels conferred an increased risk of asthma, atopic dermatitis or elevated total serum IgE, suggesting that efforts to increase vitamin D are unlikely to reduce risks of atopic disease.

4.2 Author Summary

Why Was This Study Done?

1. Observational epidemiological studies have associated low vitamin D levels with risk of asthma, atopic dermatitis and elevated IgE levels, however these studies are susceptible to confounding and reverse causation and thus it remains unclear whether these associations are true.
2. The randomized controlled trials published to date on this topic have been inconclusive.
3. If vitamin D insufficiency did cause atopic diseases, this would be of clinical relevance, since vitamin D insufficiency is common and safely correctable.

What Did Researchers Do and Find?

1. We applied a Mendelian randomization study design, which greatly limits bias due to confounding and prevents bias due to reverse causation, to understand if vitamin D levels are associated with a higher risk of adult and pediatric asthma, atopic dermatitis and elevated IgE levels.
2. Despite high statistical power, our study showed no evidence of an unconfounded association between vitamin D and the studied outcomes.

What Do These Findings Mean?

1. Our findings suggest that the previous epidemiological associations between vitamin D and atopic diseases could be due to confounding.
2. Efforts to increase vitamin D levels will probably not result in decreased risk of adult and pediatric asthma, atopic dermatitis, or elevated IgE levels.

4.3 Introduction

Atopy refers to the shared predisposition to develop allergic diseases, such as asthma and atopic dermatitis, and is characterized by increased serum IgE levels. Observational studies have identified a controversial association between low vitamin D status (as measured by serum 25 hydroxyvitamin D (25OHD) levels) and risk of asthma, atopic dermatitis and elevated serum IgE levels [1-9]. If low vitamin D were a causal risk factor for atopic diseases, this would be important for public health because vitamin D insufficiency is common, affecting 42% of Americans [10] and correctable using vitamin D supplementation.

The immune system appears to play an important role in atopy pathogenesis, and vitamin D, a proposed modulator of the immune system response, may influence the development of atopic susceptibility [11]. Although two randomized controlled trials (RCTs) [12,13] and a recent Cochrane meta-analysis of RCTs for asthma [14] showed a role for vitamin D supplementation in the reduction of atopic exacerbations, other recent data do not support its benefits for asthma, [12,13,15,16] atopic dermatitis or IgE levels [17-19]. Given this lingering controversy, clinical practice guidelines do not support vitamin D supplementation to prevent atopic disease [1,4]. However, RCT evidence is typically of low quality in vitamin D trials, because of their small sample size and limited duration; since vitamin D therapy cannot be patented, large-scale trials would rely upon the limited means of the public purse.

Given the inconclusive results from the existing RCTs and observational studies, the principles of Mendelian randomization (MR) can be applied to test the role of biomarkers in disease etiology [20]. MR uses genetic data to ascertain whether a given biomarker, such as 25OHD, is implicated in disease etiology, relying on a simple tenet: if a biomarker is etiologically involved in a disease process, then the genetic factors which influence the biomarker will influence disease risk. This established technique greatly limits confounding, since genotypes are randomly assorted at conception; further, it is free of reverse causation since genotypes are always assigned prior to the onset of disease. Thus, MR studies overcome the above limitations of observational studies and are conceptually similar to RCTs, but provide a life-long assessment of exposure to a biomarker, such as low 25OHD levels. Further, recent advances in genotyping enable the application of MR methods in sample sizes that are not realistic for RCTs of vitamin D therapy.

In the present study, we adopted an MR design to estimate the effect of genetically lowered 25OHD levels on atopic susceptibility combining retrospective data from several large-scale studies in Europeans. Our MR instruments were single nucleotide polymorphisms (SNPs) identified by the SUNLIGHT consortium, the largest genome-wide association study (GWAS) published to date for 25OHD levels (n=33,995) [21]. We then applied MR to test whether genetically lowered 25OHD levels influence asthma and atopic dermatitis susceptibility, and total serum IgE levels, using data from the largest GWAS meta-analyses to date: the asthma GABRIEL consortium [22], combined with the UK Biobank [23], for a total sample size of 25,471 cases and 121,290 controls for asthma; childhood asthma in 7,047 cases and 7,961 controls from the GABRIEL consortium; 10,788 cases and 30,047 controls for atopic dermatitis in the The EArly

Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium [24]; and the GABRIEL consortium for IgE levels (5,888 asthma cases and 6,965 controls).

4.4 Methods

An ethics statement was not required for this work.

4.4.1 SNP selection and Data Sources

This study did not have a prospective analysis plan, since it was based on already available data from large GWAS meta-analyses. Specifically, we selected the lead genome-wide significant SNPs (P values < 5 x 10-8) associated with 25OHD from the SUNLIGHT Consortium [21] as instruments in the present MR analysis.

The estimates of the effect of each SNP on 25OHD levels were obtained using data from the Canadian Multicentre Osteoporosis Study (CaMos) [25], since the effect of each SNP on 25OHD levels was not reported in the SUNLIGHT, due to the different 25OHD measurement methods used among individual cohorts, and because CaMos was among the largest replication cohorts in the SUNLIGHT Consortium, thereby providing more accurate effects of SNPs on 25OHD. The effects of these SNPs in CaMos have been previously reported [26].

To obtain precise estimates of the genome-wide significant SNP for vitamin D levels on asthma, we tested the effect of each of these SNPs in a meta-analysis of the UK Biobank Asthma GWAS [27] with the GABRIEL consortium [22] (**S1 Text**).

The association between vitamin D levels and asthma may be stronger in children [28] than adults [29]. Therefore, we conducted a separate analysis in a subsample of the GABRIEL consortium, including 15,008 individuals (7,047 cases / 7,961 controls) with childhood-onset asthma.

For atopic dermatitis, we obtained the effects of the selected SNPs from The EAGLE Eczema Consortium [24]. For naturally log-transformed total serum IgE levels, the same estimates were obtained from the GABRIEL consortium. A description of the participating studies and definitions of the asthma, atopic dermatitis and IgE phenotypes appear in the **S1 Text**.

4.4.4 SNP Validation

To validate the four SUNLIGHT 25OHD SNPs as instruments for our MR analysis, we tested them for the three MR assumptions: strong association with the exposure (25OHD), absence of association with known confounders of the exposure–outcome association, and absence of pleiotropy, where the genetic variant influences the atopic outcome through mechanisms that are independent of the vitamin D. Bias due to linkage disequilibrium (LD) and population stratification has been previously tested for the four 25OHD-associated SNPs [26].

Due to randomization of alleles at conception, confounding is greatly minimized in MR studies, however, we have examined if 25OHD-associated SNPs may influence important known confounders that may link vitamin D to common disease [30]. Specifically, these SNPs were not associated with sun exposure, time outside, physical activity, smoking and BMI [30].

Pleiotropy may bias results if the chosen SNPs exert effects on asthma, atopic dermatitis and IgE levels independently of 25OHD levels. In this study, pleiotropy is less likely since all 25OHD-associated SNPs map to genes strongly implicated in 25OHD physiology. Nonetheless, we conducted a PubMed literature search to identify possible pleiotropic mechanisms (**S1 Text**).

4.4.5 Statistical analysis

Association of SUNLIGHT SNPs with Asthma, Atopic Dermatitis and IgE Level

We first assessed whether each SNP was associated with risk of asthma, atopic dermatitis or IgE level, applying a Bonferroni correction, where statistical significance was declared at $P \leq 0.05/4$ since four SNPs were used as instruments.

Mendelian Randomization Estimates

We assessed the effects of the SNPs upon the four outcomes, weighting the effect of each SNP by the magnitude of its effect upon 25OHD levels. In the absence of available data on 25OHD levels in the GWAS assessing the four outcomes, the instrumental variable estimates of genetically

determined odds ratios and betas were calculated by using the two sample MR approach [31]. To provide a summary measure for the effect including all SNPs genome-wide significant for 25OHD, we combined weighted estimates using fixed effects models, and used I2 estimate as a measure of heterogeneity [32,33]. The effect-size for the meta-analysis is reported in our results as the effect of a standard deviation (SD) change in natural log-transformed 25OHD levels, since this metric is more interpretable than an arbitrary difference. Finally, we undertook power calculations [34] to test whether our study was adequately powered to detect clinically relevant change in the outcomes.

Sensitivity Analyses

Our MR estimates were recalculated after exclusion of SNPs potentially influenced by pleiotropy or population stratification. We also performed a stratified MR analysis where SNPs involved in either 25OHD synthesis or metabolism were analyzed separately [30].

4.4.5 Ethical Approval

All human studies were approved by their institutional ethics review committees, and all participants provided written consent.

4.5 Results

4.5.1 SNP Selection and Validation

SNP selection

The SUNLIGHT Consortium identified four genome-wide significant vitamin-D associated SNPs[21]: rs2282679 in *GC* (vitamin D binding protein), rs12785878 near *DHCR7* (7-dehydrocholesterol reductase), rs10741657 near *CYP2R1* (cytochrome P450 family 2 subfamily R member 1) and rs6013897 in *CYP24A1* (cytochrome P450 family 24 subfamily A member 1) (Table 1). All four SNPs map near or in genes implicated in mechanisms modulating 25OHD levels, and more specifically transport (*GC*), synthesis (*DHCR7*), hepatic hydroxylation (*CYP2R1*) and catabolism (*CYP24A1*) [35].

LD, Confounding and Pleiotropy Assessment

We found no evidence of LD between any of these SNPs (all pairwise $r2 \leq 0.01$).

In our literature search for potential confounders, obesity and smoking were identified as risk factors for asthma [36] that have been associated with vitamin D levels [30]. We found no association between these SNPs and BMI (all P-values$\geq$ 0.29) in The Genetic Investigation of ANthropometric Traits (GIANT) [37] consortium, or with smoking in the Tobacco and Genetics Consortium [38] (all P-values$\geq$ 0.18) (**S1 Table**).

The 25OHD-associated SNPs may also influence risk of atopic disease, independently of 25OHD, through pleiotropy (Fig. 1, Fig. 2, Fig. 3). Two *CYP2R1* SNPs (rs2060793 and rs1933064) have been associated with increased eosinophil counts, while the *GC* SNP rs7041 and the *CYP2R1* SNP rs7935792 may modulate total IgE [39]. Therefore, to assess for possible pleiotropy, we tested the *CYP2R1* SNP for LD with the aforementioned *CYP2R1* SNPs. We found evidence for strong LD

between the SUNLIGHT SNP rs10741657 and the eosinophil-related rs2060793 (r^2=0.96), but no evidence for linkage between the rs10741657 and the two other SNPs (r^2<0.2). We also found weak LD between the SUNLIGHT *GC* SNP and the rs7041 influencing IgE levels (r^2=0.5). Additionally, our literature review showed an association in children between *CYP24A1* mRNA and LL-37, an immuno-modulating peptide potentially related to asthma [29]. Therefore, we performed sensitivity analyses excluding the *CYP2R1* and CYP24A1 SNPs from our MR instruments in our asthma analysis, and excluding the *CYP2R1* SNP in our atopic dermatitis and IgE analysis.

Population Stratification Assessment

Based on our previously published results[26], only rs12785878 at DHCR7 was strongly associated with non-European ancestry. Given that the prevalence of both asthma and atopic dermatitis is increased in the individuals of African ancestry [40] we undertook sensitivity analyses excluding this SNP.

4.5.2 Association of SUNLIGHT SNPs with 25OHD Levels

The association of the four genome-wide significant SNPs from SUNLIGHT with 25OHD levels is described in Table 1. The proportion of the population-level variance in 25OHD levels explained by the four SNPs is reflected by the F-statistics. Although the low F-statistic of the *CYP24A1* SNP suggests that this might be a rather "weak" MR instrument, it is important to note that these F-statistics are derived from CaMos, a subsample of the SUNLIGHT study, and consequently the F-statistics would tend to increase if they were tested in the entire SUNLIGHT study. We have

previously shown [41] that the count of 25OHD decreasing alleles across these four SNPs was strongly associated with lower total 25OHD levels in CaMos ($P = 2.4 \times 10^{-12}$) (**Table 1**).

4.5.3 Association of SUNLIGHT SNPs with Asthma and Atopic Dermatitis Susceptibility and IgE levels

Summary statistics for the four 25OHD-associated SNPs and their associations with asthma were taken from the fixed-effects meta-analysis of the UK Biobank and GABRIEL studies. Since the *CYP24A1* SNP was absent in the UK Biobank genotypic dataset, we used the estimate of its perfect proxy rs17217119 ($r^2 = 1.0$). All four 25OHD-decreasing alleles were not associated with risk of asthma, childhood asthma, atopic dermatitis or IgE levels (**Table 1**) and the 95% confidence intervals were generally tight around the null.

4.5.4 Mendelian Randomization Analysis for the Association of 25OHD with Asthma Risk

In order to estimate the association of genetically lowered 25OHD upon asthma, we used a fixed-effects model including all four 25OHD-decreasing alleles. A decrease in 25OHD levels by one SD on the natural log scale was not associated with asthma (OR = 1.03, 95% CI: 0.90,1.19; P= 0.63 $I^2 = 0\%$) (**Table 2** and **Fig. 4**). Since our model included only 4 SNPs, the 95% CIs of the I^2 statistic were wide (0%-85%) and consequently heterogeneity could not be accurately measured using this parameter. In addition, due to potential population stratification, we undertook sensitivity analyses by excluding the *DHCR7* SNP and again observed no association with asthma (**Table 3**), To assess the effect of the independent vitamin D pathways on risk of asthma, we

analyzed SNPs near genes implicated in 25OHD synthesis (*DHCR7* and *CYP2R1*) and metabolism (*GC*, and *CYP24A1*) separately and found that both were again not associated with increased risk of asthma (**Table 3**). Also, because of evidence of possible pleiotropy for the *CYP2R1* and *CYP24A1* SNPs, we performed a sensitivity analysis after removing these two variants. The results again showed no evidence of an effect (**Table 3**). Testing the effect of genetically reduced 25OHD on risk of childhood asthma, we found that each SD increase in natural log-transformed 25OHD was not associated with risk of asthma (OR = 0.95, 95% CI: 0.69,1.31; P= 0.76 I^2 = 0%).

Given these null results, we undertook a power calculation [34]. Based on a clinically relevant effect of an OR of 1.6 for asthma [1], a sample size of 144,243 individuals and setting alpha to 0.05, our study had a power of 100% to detect an OR of 1.6 for a one SD change in log transformed 25OHD levels on asthma risk, and a 80% power to exclude effects as small as an OR of 1.12. The same power calculation for a sample size of 15,008 individuals for childhood asthma gave to our study 100% power to detect an OR of 1.6 for a one SD change in log transformed 25OHD levels on childhood asthma risk and 80% power to exclude effects as small as an OR of 1.33.

4.5.5 Mendelian Randomization Analysis of the Association of 25OHD with Atopic Dermatitis

A decrease in 25OHD levels by one SD on the natural log scale was not associated with atopic dermatitis (OR = 1.12, 95% CI: 0.92,1.37; P= 0.27 I^2 = 15%) (**Table 3 and Fig. 5**). Similar to the previous analysis for asthma, we undertook sensitivity analyses by excluding the *DHCR7* SNP to

control for possible population stratification, and by removing the *CYP2R1* SNP, because of potential pleiotropy, and the results were similar (**Table 3**). Analyzing SNPs near genes implicated in 25OHD synthesis and metabolism separately, again no association was found (**Table 3**). Based on a previously reported OR of 1.5 for atopic dermatitis in vitamin D insufficient individuals [5], a sample size of 40,835 individuals and setting alpha to 0.05, our study had a power of 100% to detect an OR of 1.5 for one SD decrease in log-transformed 25OHD levels on atopic dermatitis risk, and a 80% power to observe effects down to an OR of 1.21.

4.5.6 Mendelian Randomization Analysis of the Association of 25OHD with IgE levels

MR analyses for IgE levels showed that a decrease in 25OHD levels by one SD on the natural log scale was not associated with naturally log-transformed total serum IgE levels (beta = -0.40 natural log-transformed units, 95% CI -1.65,0.85, P=0.54, I2=0) (**Table 2 and Fig. 6**). After sensitivity analyses excluding the *DHCR7* or the *CYP2R1* SNP, and analyzing separately 25OHD synthesis and metabolism SNPs, the results still included the null (**Table 3**). Based on a previously reported beta of -0.43 for log-transformed total IgE levels per SD increase in log-transformed 25OHD levels[42], our sample size of 12,853 individuals, and setting alpha to 0.05, our study had a power of 86%.

4.6 Discussion

Using large populations of individuals of European descent, our study failed to provide evidence supporting a role for vitamin D in adult and childhood onset asthma, atopic dermatitis and IgE levels, although small effects cannot be excluded. These findings provide no rationale for the use of vitamin D for prevention of these conditions.

Residual confounding may account for a large part of the discrepancy between our findings and these of observational studies. For instance, adiposity predisposes to asthma [43] and lowers 25OHD levels [44]. While previous studies have controlled for BMI, it is a poor measure of adiposity [45]. Physical activity might also be a strong confounder in observational studies, because it is associated with asthma [46] and sunlight exposure, which in-turn influences vitamin D status. Another possible explanation of the discrepancy between MR results and observational studies could be reverse causation, since asthmatic individuals tend to be less active [46] and are therefore less exposed to sunlight. Further, steroid therapies used in asthma may result in low vitamin D levels [47]. Last, individuals with darker skin are at increased risk of atopic dermatitis and asthma [40], while they are also more susceptible to develop vitamin D insufficiency [48].

Our results are in accordance with two recent meta-analyses of RCTs [15,16], which conclude that evidence is lacking to support a regular use of vitamin D supplements for prevention of asthma exacerbations, but are in contrast with the findings of a recent Cochrane meta-analysis of RCTs [14]. Our findings for atopic dermatitis and IgE agree with recent observational studies and RCTs

[3,4,9]. In contrast, we have used the same methods to provide strong evidence supporting a causal role for 25OHD in risk of multiple sclerosis [41]. Thus, it would appear that immuno-modulatory effects of 25OHD do not uniformly influence immune-mediated diseases.

Previous studies have explored effects of vitamin D-related genes and risk on atopic phenotypes, without applying an MR approach [7,39,49-52]. SNPs in the vitamin D pathway (*CYP27A1*, *CYP27B1, CYP2R1, CYP24A1, GC*) affecting 25OHD levels demonstrated moderate effects on risk of asthma in a prior adult study [49], but the role of these variants on asthma risk later in life is unknown. A recent study provided evidence of an association between the vitamin D receptor genes and asthma in adolescents with normal 25OHD levels [50]. Another study reported an association between a *CYP2R1* variant and FEV1 in children and between specific haplotypes on *CYP2R1* and *CYP27A1* and asthma phenotypes [51]. With regards to atopic dermatitis, a polymorphism in the *CYP24A1* gene has been associated with severe atopic dermatitis in adults [52]. A vitamin D-related SNP on *CYP27A1* was also found to be protective against eczema, whereas *CYP2R1* and *VDR* haplotypes appear to alter eczema susceptibility. In regards to IgE levels, carrying a rare variant on CYP27A1 appears to increase the risk of elevated total serum IgE levels (above 1000IU/ml) and a *CYP27B1* allele has also been associated to IgE levels [53]. Nevertheless, other than testing genetic associations, the above studies were not designed to test the causal relationship between 25OHD levels and atopy.

Other MR studies have been carried out in asthma and have provided evidence supporting a causal role for increased BMI in asthma [43], and evidence against a causal role of prenatal alcohol exposure in asthma and atopy in childhood [54]. A recent MR study used a smaller sample of 1,208 cases and 3,877 controls for childhood asthma in individuals of European and non-European ancestry [55] and did not find any evidence for a causal role of vitamin D in asthma. The instruments used in this childhood MR asthma study were only the *CYP2R1* and GC SNPs, and did not include the SNPs near *DHCR7* and *CYP24A1*. Our study thus provides a more thorough examination of the effects of 25OHD on asthma risk by using a substantially larger sample size, including all 25OHD-associated loci, in both adult and childhood asthma.

Strengths of this study include the large sample size of the adult cohorts, which enabled us to more precisely test our study hypothesis than if we had used individual-level data from small studies. Although the findings from this study were null, the high statistical power and tight confidence intervals exclude most clinically relevant effects of 25OHD on risk of asthma, atopic dermatitis and IgE levels. Importantly, our findings represent the association of a life-long exposure to reduced vitamin D levels in the general adult population.

This study also has limitations. While we controlled for pleiotropy, residual bias is possible since the exact function of these SNPs is unknown. However, all SNPs lie in, or near, genes well validated for their role in vitamin D physiology. The null result could also be explained by canalization, which is a phenomenon resulting in compensatory feedback mechanisms [20]. Our

childhood asthma MR study was underpowered to exclude effects of vitamin D on pediatric asthma smaller than an OR less than 1.33 per SD change in log vitamin D. There was heterogeneity in the definition of the different atopic outcomes, since in some GWAS used for this MR study the outcomes were self-reported and in others physician- diagnosed **(see S1 Text)**. Our MR analysis might also be limited in its ability to elucidate a possible role of biologically active vitamin D, reflected by the levels of the active metabolite 1,25-dihydroxyvitamin D. Although genetically lowered total 25OHD levels do not appear to be associated with increased risk of the studied atopic phenotypes, we have not assessed whether reduced lifelong 1,25-dihydroxyvitamin D could influence these outcomes, since levels of total 25OHD and 1,25-dihydroxyvitamin D are weakly correlated [56]. As well, our study can only comment on the role of circulating 25OHD level, and not on the action of 25OHD at the cellular level. Our analyses were restricted to white populations of European ancestry, and further work will be required to investigate their relevance in populations of different ethnicity, or in those with frank vitamin D deficiency.

In conclusion, our MR study provides no support for an unconfounded relationship between 25OHD levels and risk of atopic disease in individuals of European descent. Instead, association of 25OHD levels with atopic diseases in the general population is more likely to be attributable to confounding by lifestyle factors such as obesity and physical inactivity.

4.7 Acknowledgements

We wish to kindly thank the SUNLIGHT Consortium, the CaMos study, the UKBIOBANK study and the GABRIEL and the EAGLE Eczema Consortia for access to their data.

4.8 References

1. Man L, Zhang Z, Zhang M, Zhang Y, Li J, Zheng N, et al. Association between vitamin D deficiency and insufficiency and the risk of childhood asthma: evidence from a meta-analysis. Int J Clin Exp Med. 2015;8(4):5699-706. PubMed PMID: 26131154; PubMed Central PMCID: PMCPMC4483990.

2. Cassim R, Russell MA, Lodge CJ, Lowe AJ, Koplin JJ, Dharmage SC. The role of circulating 25 hydroxyvitamin D in asthma: a systematic review. Allergy. 2015;70(4):339-54. doi: 10.1111/all.12583. PubMed PMID: 25631639.

3. Robl R, Uber M, Abagge KT, Lima MN, Carvalho VO. Serum Vitamin D Levels Not Associated with Atopic Dermatitis Severity. Pediatr Dermatol. 2016. doi: 10.1111/pde.12795. PubMed PMID: 26862046.

4. Debinska A, Sikorska-Szaflik H, Urbanik M, Boznanski A. The role of vitamin D in atopic dermatitis. Dermatitis. 2015;26(4):155-61. doi: 10.1097/DER.0000000000000128. PubMed PMID: 26172483.

5. Cheng HM, Kim S, Park GH, Chang SE, Bang S, Won CH, et al. Low vitamin D levels are associated with atopic dermatitis, but not allergic rhinitis, asthma, or IgE sensitization, in the adult Korean population. J Allergy Clin Immunol. 2014;133(4):1048-55. doi: 10.1016/j.jaci.2013.10.055. PubMed PMID: 24388009.

6. Wang SS, Hon KL, Kong AP, Pong HN, Wong GW, Leung TF. Vitamin D deficiency is associated with diagnosis and severity of childhood atopic dermatitis. Pediatr Allergy Immunol. 2014;25(1):30-5. doi: 10.1111/pai.12167. PubMed PMID: 24383670.

7. Hypponen E, Berry DJ, Wjst M, Power C. Serum 25-hydroxyvitamin D and IgE - a significant but nonlinear relationship. Allergy. 2009;64(4):613-20. doi: 10.1111/j.1398-9995.2008.01865.x. PubMed PMID: 19154546.

8. Kang JW, Kim JH, Kim HJ, Lee JG, Yoon JH, Kim CH. Association of serum 25-hydroxyvitamin D with serum IgE levels in Korean adults. Auris Nasus Larynx. 2016;43(1):84-8. doi: 10.1016/j.anl.2015.06.010. PubMed PMID: 26209260.

9. Dogru M, Kirmizibekmez H, Yesiltepe Mutlu RG, Aktas A, Ozturkmen S. Clinical effects of vitamin D in children with asthma. Int Arch Allergy Immunol. 2014;164(4):319-25. doi: 10.1159/000366279. PubMed PMID: 25277142.

10. Forrest KY, Stuhldreher WL. Prevalence and correlates of vitamin D deficiency in US adults. Nutr Res. 2011;31(1):48-54. doi: 10.1016/j.nutres.2010.12.001. PubMed PMID: 21310306.

11. Searing DA, Leung DY. Vitamin D in atopic dermatitis, asthma and allergic diseases. Immunol Allergy Clin North Am. 2010;30(3):397-409. doi: 10.1016/j.iac.2010.05.005. PubMed PMID: 20670821; PubMed Central PMCID: PMCPMC2914320.

12. Arshi S, Fallahpour M, Nabavi M, Bemanian MH, Javad-Mousavi SA, Nojomi M, et al. The effects of vitamin D supplementation on airway functions in mild to moderate persistent

asthma. Ann Allergy Asthma Immunol. 2014;113(4):404-9. doi: 10.1016/j.anai.2014.07.005. PubMed PMID: 25091714.

13. Castro M, King TS, Kunselman SJ, Cabana MD, Denlinger L, Holguin F, et al. Effect of vitamin D3 on asthma treatment failures in adults with symptomatic asthma and lower vitamin D levels: the VIDA randomized clinical trial. JAMA. 2014;311(20):2083-91. doi: 10.1001/jama.2014.5052. PubMed PMID: 24838406; PubMed Central PMCID: PMCPMC4217655.

14. Martineau AR, Cates CJ, Urashima M, Jensen M, Griffiths AP, Nurmatov U, et al. Vitamin D for the management of asthma. Cochrane Database Syst Rev. 2016;9:CD011511. doi: 10.1002/14651858.CD011511.pub2. PubMed PMID: 27595415.

15. Riverin BD, Maguire JL, Li P. Vitamin D Supplementation for Childhood Asthma: A Systematic Review and Meta-Analysis. PLoS One. 2015;10(8):e0136841. doi: 10.1371/journal.pone.0136841. PubMed PMID: 26322509; PubMed Central PMCID: PMCPMC4556456.

16. Luo J, Liu D, Liu CT. Can Vitamin D Supplementation in Addition to Asthma Controllers Improve Clinical Outcomes in Patients With Asthma?: A Meta-Analysis. Medicine (Baltimore). 2015;94(50):e2185. doi: 10.1097/MD.0000000000002185. PubMed PMID: 26683927.

17. Kim G, Bae JH. Vitamin D and atopic dermatitis: A systematic review and meta-analysis. Nutrition. 2016. doi: 10.1016/j.nut.2016.01.023. PubMed PMID: 27061361.

18. Bath-Hextall FJ, Jenkinson C, Humphreys R, Williams HC. Dietary supplements for established atopic eczema. Cochrane Database Syst Rev. 2012;2:CD005205. doi: 10.1002/14651858.CD005205.pub3. PubMed PMID: 22336810.

19. Hata TR, Audish D, Kotol P, Coda A, Kabigting F, Miller J, et al. A randomized controlled double-blind investigation of the effects of vitamin D dietary supplementation in subjects with atopic dermatitis. J Eur Acad Dermatol Venereol. 2014;28(6):781-9. doi: 10.1111/jdv.12176. PubMed PMID: 23638978; PubMed Central PMCID: PMCPMC3769441.

20. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133-63. doi: 10.1002/sim.3034. PubMed PMID: 17886233.

21. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet. 2010;376(9736):180-8. doi: 10.1016/S0140-6736(10)60588-0. PubMed PMID: 20541252; PubMed Central PMCID: PMC3086761.

22. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010;363(13):1211-21. doi: 10.1056/NEJMoa0906312. PubMed PMID: 20860503; PubMed Central PMCID: PMCPMC4260321.

23. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, 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. 2015;12(3):e1001779. doi: 10.1371/journal.pmed.1001779. PubMed PMID: 25826379; PubMed Central PMCID: PMCPMC4380465.

24. Paternoster L, Standl M, Chen CM, Ramasamy A, Bonnelykke K, Duijts L, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet. 2012;44(2):187-92. doi: 10.1038/ng.1017. PubMed PMID: 22197932; PubMed Central PMCID: PMCPMC3272375.

25. Faye LL, Sun L, Dimitromanolakis A, Bull SB. A flexible genome-wide bootstrap method that accounts for ranking and threshold-selection bias in GWAS interpretation and replication study design. Stat Med. 2011;30(15):1898-912. doi: 10.1002/sim.4228. PubMed PMID: 21538984.

26. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Goltzman D, et al. Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLoS Med. 2015;12(8):e1001866. doi: 10.1371/journal.pmed.1001866. PubMed PMID: 26305103; PubMed Central PMCID: PMCPMC4549308.

27. Ollier W, Sprosen T, Peakman T. UK Biobank: from concept to reality. Pharmacogenomics. 2005;6(6):639-46. doi: 10.2217/14622416.6.6.639. PubMed PMID: 16143003.

28. Hollams EM, Hart PH, Holt BJ, Serralha M, Parsons F, de Klerk NH, et al. Vitamin D and atopy and asthma phenotypes in children: a longitudinal cohort study. Eur Respir J. 2011;38(6):1320-7. doi: 10.1183/09031936.00029011. PubMed PMID: 21565922.

29. Goleva E, Searing DA, Jackson LP, Richers BN, Leung DY. Steroid requirements and immune associations with vitamin D are stronger in children than adults with asthma. J Allergy Clin Immunol. 2012;129(5):1243-51. doi: 10.1016/j.jaci.2012.01.044. PubMed PMID: 22330698; PubMed Central PMCID: PMCPMC3340468.

30. Vimaleswaran KS, Berry DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, et al. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. PLoS Med. 2013;10(2):e1001383. doi: 10.1371/journal.pmed.1001383. PubMed PMID: 23393431; PubMed Central PMCID: PMC3564800.

31. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65. doi: 10.1002/gepi.21758. PubMed PMID: 24114802; PubMed Central PMCID: PMC4377079.

32. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60. doi: 10.1136/bmj.327.7414.557. PubMed PMID: 12958120; PubMed Central PMCID: PMC192859.

33. Greco MF, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat Med. 2015;34(21):2926-40. doi: 10.1002/sim.6522. PubMed PMID: 25950993.

34. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42(5):1497-501. doi: 10.1093/ije/dyt179. PubMed PMID: 24159078; PubMed Central PMCID: PMCPMC3807619.

35. Dastani Z, Li R, Richards B. Genetic regulation of vitamin D levels. Calcif Tissue Int. 2013;92(2):106-17. doi: 10.1007/s00223-012-9660-z. PubMed PMID: 23114382.

36. Thomsen SF. Epidemiology and natural history of atopic diseases. Eur Clin Respir J. 2015;2. doi: 10.3402/ecrj.v2.24642. PubMed PMID: 26557262; PubMed Central PMCID: PMCPMC4629767.

37. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012;44(6):659-69. doi: 10.1038/ng.2274. PubMed PMID: 22581228; PubMed Central PMCID: PMCPMC3613127.

38. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet. 2010;42(5):441-7. doi: 10.1038/ng.571. PubMed PMID: 20418890; PubMed Central PMCID: PMCPMC2914600.

39. Wang SS, Hon KL, Kong AP, Tang MF, Sy HY, Chan JC, et al. Eczema phenotypes are associated with multiple vitamin D pathway genes in Chinese children. Allergy. 2014;69(1):118-24. doi: 10.1111/all.12337. PubMed PMID: 24730053.

40. Silverberg JI. Atopic dermatitis. JAMA Dermatol. 2014;150(12):1380. doi: 10.1001/jamadermatol.2014.2757. PubMed PMID: 25493472.

41. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Leong A, et al. Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLoS Med. 2015;12(8):e1001866. doi: 10.1371/journal.pmed.1001866. PubMed PMID: 26305103; PubMed Central PMCID: PMCPMC4549308.

42. Brehm JM, Celedon JC, Soto-Quiros ME, Avila L, Hunninghake GM, Forno E, et al. Serum vitamin D levels and markers of severity of childhood asthma in Costa Rica. Am J Respir Crit Care Med. 2009;179(9):765-71. doi: 10.1164/rccm.200808-1361OC. PubMed PMID: 19179486; PubMed Central PMCID: PMCPMC2675563.

43. Granell R, Henderson AJ, Evans DM, Smith GD, Ness AR, Lewis S, et al. Effects of BMI, fat mass, and lean mass on asthma in childhood: a Mendelian randomization study. PLoS Med. 2014;11(7):e1001669. doi: 10.1371/journal.pmed.1001669. PubMed PMID: 24983943; PubMed Central PMCID: PMCPMC4077660.

44. Vimaleswaran KS, Cavadino A, Berry DJ, LifeLines Cohort Study i, Jorde R, Dieffenbach AK, et al. Association of vitamin D status with arterial blood pressure and hypertension risk: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2014;2(9):719-29. doi: 10.1016/S2213-8587(14)70113-5. PubMed PMID: 24974252.

45. Prentice AM, Jebb SA. Beyond body mass index. Obes Rev. 2001;2(3):141-7. PubMed PMID: 12120099.

46. Lochte L, Nielsen KG, Petersen PE, Platts-Mills TA. Childhood asthma and physical activity: a systematic review with meta-analysis and Graphic Appraisal Tool for Epidemiology assessment. BMC Pediatr. 2016;16(1):50. doi: 10.1186/s12887-016-0571-4. PubMed PMID: 27091126.

47. Searing DA, Zhang Y, Murphy JR, Hauk PJ, Goleva E, Leung DY. Decreased serum vitamin D levels in children with asthma are associated with increased corticosteroid use. J

Allergy Clin Immunol. 2010;125(5):995-1000. doi: 10.1016/j.jaci.2010.03.008. PubMed PMID: 20381849; PubMed Central PMCID: PMCPMC2866800.

48. Wang S. Epidemiology of vitamin D in health and disease. Nutr Res Rev. 2009;22(2):188-203. doi: 10.1017/S0954422409990151. PubMed PMID: 19860998.

49. Bosse Y, Lemire M, Poon AH, Daley D, He JQ, Sandford A, et al. Asthma and genes encoding components of the vitamin D pathway. Respir Res. 2009;10:98. doi: 10.1186/1465-9921-10-98. PubMed PMID: 19852851; PubMed Central PMCID: PMCPMC2779188.

50. Papadopoulou A, Kouis P, Middleton N, Kolokotroni O, Karpathios T, Nicolaidou P, et al. Association of vitamin D receptor gene polymorphisms and vitamin D levels with asthma and atopy in Cypriot adolescents: a case-control study. Multidiscip Respir Med. 2015;10(1):26. doi: 10.1186/s40248-015-0025-0. PubMed PMID: 26346690; PubMed Central PMCID: PMCPMC4559891.

51. Leung TF, Wang SS, Tang MF, Kong AP, Sy HY, Hon KL, et al. Childhood asthma and spirometric indices are associated with polymorphic markers of two vitamin D 25-hydroxylase genes. Pediatr Allergy Immunol. 2015;26(4):375-82. doi: 10.1111/pai.12392. PubMed PMID: 25845986.

52. Hallau J, Hamann L, Schumann RR, Worm M, Heine G. A Promoter Polymorphism of the Vitamin D Metabolism Gene Cyp24a1 is Associated with Severe Atopic Dermatitis in Adults. Acta Derm Venereol. 2016;96(2):169-72. doi: 10.2340/00015555-2226. PubMed PMID: 26315479.

53. Suzuki H, Makino Y, Nagata M, Furuta J, Enomoto H, Hirota T, et al. A rare variant in CYP27A1 and its association with atopic dermatitis with high serum total IgE. Allergy. 2016;71(10):1486-9. doi: 10.1111/all.12950. PubMed PMID: 27259383.

54. Shaheen SO, Rutterford C, Zuccolo L, Ring SM, Davey Smith G, Holloway JW, et al. Prenatal alcohol exposure and childhood atopic disease: a Mendelian randomization approach. J Allergy Clin Immunol. 2014;133(1):225-32 e1-5. doi: 10.1016/j.jaci.2013.04.051. PubMed PMID: 23806636; PubMed Central PMCID: PMCPMC3884122.

55. Hysinger EB, Roizen JD, Mentch FD, Vazquez L, Connolly JJ, Bradfield JP, et al. Mendelian randomization analysis demonstrates that low vitamin D is unlikely causative for pediatric asthma. J Allergy Clin Immunol. 2016. doi: 10.1016/j.jaci.2016.06.056. PubMed PMID: 27554823.

56. Zittermann A, Schleithoff SS, Frisch S, Gotting C, Kuhn J, Koertke H, et al. Circulating calcitriol concentrations and total mortality. Clin Chem. 2009;55(6):1163-70. doi: 10.1373/clinchem.2008.120006. PubMed PMID: 19359534.

4.9 Tables and Figures

4.9.1 Tables

Table 1: Characteristics of SNPs used as instrumental variables and their association with Asthma, Atopic Dermatitis and IgE levels

Vitamin D (25OHD) Results								Asthma Results			Childhood Asthma Results§			Atopic Dermatitis Results∞			IgE Results§		
Locus	25OHD Associated SNP	EA	EAF	Effect on 25OHD*	P#	F-Statistic♮	Variance in 25OHD explained by each SNP (%)	OR (95% CI)	P	N	OR (95% CI)	P	N	OR (95% CI)	P	N	Beta (95% CI)	P	N
CYP2R1	rs10741657	C	0.62	-0.052	$3.3x10^{-20}$	18.78	0.13	0.99 (0.97,1.01)	0.54	142,551	1.02 (0.96,1.07)	0.56	15,008	1.02 (0.99,1.05)	0.27	40,834	-0.02 (-.23,0.19)	0.86	12,853
DHCR7	rs12785878	G	0.27	-0.056	$2.1x10^{-27}$	18.29	0.12	1.01 (0.98,1.03)	0.64	142,551	0.95 (0.90,1.01)	0.11	15,008	1.02 (0.98,1.06)	0.32	40,834	-0.15 (-.36,0.06)	0.20	12,853
GC	rs2282679	C	0.3	-0.047	$1.9x10^{-109}$	13.38	0.09	1.01 (0.99,1.04)	0.31	144,243	1.00 (0.95,1.06)	0.96	15,008	0.98 (0.94,1.02)	0.32	40,531	0.06 (-.17,0.29)	0.60	12,853
CYP24A1	rs6013897	A	0.19	-0.027	$6.0x10^{-10}$	3.13	0.02	1.02 (0.99,1.05)	0.14	144,243	1.03 (0.97,1.10)	0.38	15,008	1.03 (0.99,1.07)	0.22	40,529	0.02 (-.25,0.29)	0.90	12,853

EA: Effect allele. EAF: Effect Allele frequency

*Effect on Natural Log-Transformed 25OHD levels in the CaMos Cohort, adjusted for age, age^2, sex, season of blood draw and BMI.

#P-values derived from the SUNLIGHT Consortium

♮F-Statistic derived from Multiply-Adjusted Natural Log-Transformed 25OHD levels in the CaMos Cohort

♭ Results are derived from the meta-analysis of UKBIOBANK study and GABRIEL asthma consortium

§ Results are derived from the GABRIEL asthma consortium

∞ Results are derived from the EAGLE eczema consortium

Table 2: MR Estimates of the Association of Decreased 25OHD on the Risk of Asthma, Atopic dermatitis and IgE Levels

Outcome	MR Estimate Odds Ratio or Beta (95% CI)	P	I^2 (95% CI)
Asthma	1.03 (0.90,1.19)*	0.63	0 (0-85)
Childhood Asthma	0.95 (0.69,1.31)*	0.76	0 (0-85)
Atopic Dermatitis	1.12 (0.92,1.37)*	0.27	15 (0-87)
IgE Levels	-0.40 (-1.65,0.85)**	0.54	0 (0-85)

*OR is expressed as the odds of asthma or atopic dermatitis per standard deviation decrease in natural log-transformed 25OHD levels.

**beta is the effect per standard deviation decrease in natural log-transformed 25OHD levels on natural log-transformed total IgE levels.

Table 3: Sensitivity Analyses Testing MR Assumptions.

	Asthma		Atopic Dermatitis		IgE levels	
Sensitivity Analysis Model	**OR (95% CI) ***	**P**	**OR (95% CI) ***	**P**	**Beta (95% CI) ****	**P**
Excluding the *DHCR7* locus	1.02 (0.86,1.22)	0.8	1.08 (0.85,1.39)	0.53	0.17 (-1.41,1.75)	0.83
Synthesis loci (*CYP2R1* and *DHCR7*)	0.99 (0.85,1.15)	0.89	1. 20 (0.94,1.52)	0.14	-0.80 (-2.28,0.68)	0.29
Metabolism loci (*GC* and *CYP24A1*)	1.20 (0.96,1.51)	0.12	0.95(0.65,1.37)	0.77	0.63 (-1.73,2.99)	0.60
Excluding the *CYP2R1* and *CYP24A1* loci	1.09 (0.92,1.30)	0.31	NA	NA	NA	NA
Excluding the *CYP2R1* locus	NA	NA	1.07 (0.83,1.38)	0.27	-0.50 (-2.04,1.04)	0.53

*OR is expressed as the odds of asthma or atopic dermatitis per standard deviation decrease in natural log-transformed 25OHD levels.

**beta is the effect per standard deviation decrease in natural log-transformed 25OHD levels on natural log-transformed total IgE levels.

"NA" is denoted when potential bias was not detected for the analysis

4.9.2 Figures

Figure 1: Direct acyclic graph (DAG) of the Mendelian Randomization analysis for Asthma

Blue arrows represent the effect of SNPs on the change in natural log-transformed 25OHD levels. Green arrows represent the causal effect of genetically decreased 25OHD levels on the risk of Asthma.

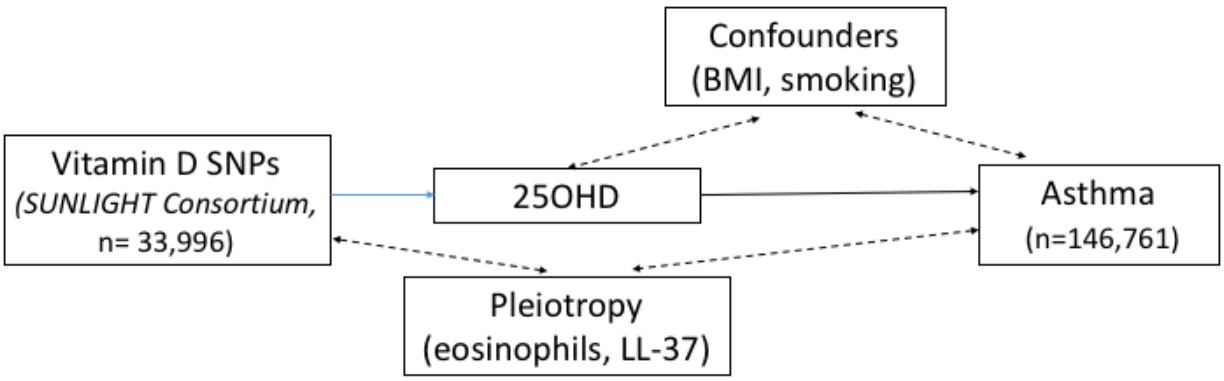


Figure 2: Direct acyclic graph (DAG) of the Mendelian Randomization analysis for Atopic Dermatitis

Blue arrows represent the effect of SNPs on the change in natural log-transformed 25OHD levels. Green arrows represent the causal effect of genetically decreased 25OHD levels on the risk of Atopic Dermatitis.

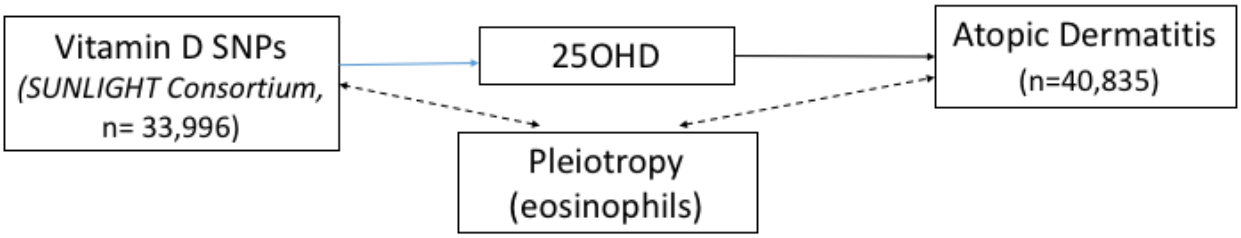


Figure 3: Direct acyclic graph (DAG) of the Mendelian Randomization analysis for IgE Levels

Blue arrows represent the effect of SNPs on the change in natural log-transformed 25OHD levels. Green arrows represent the causal effect of genetically decreased 25OHD levels on IgE levels.

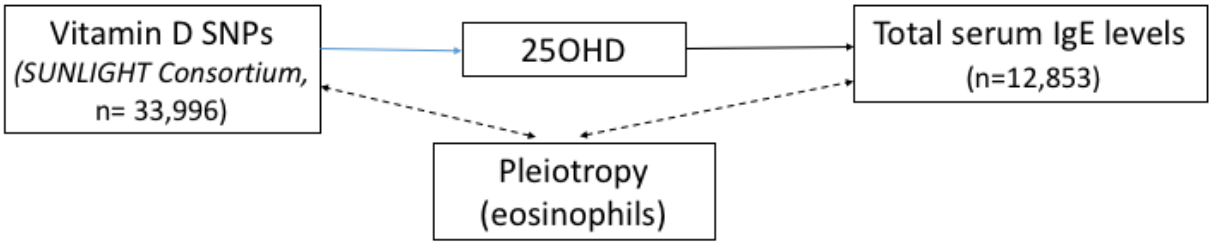
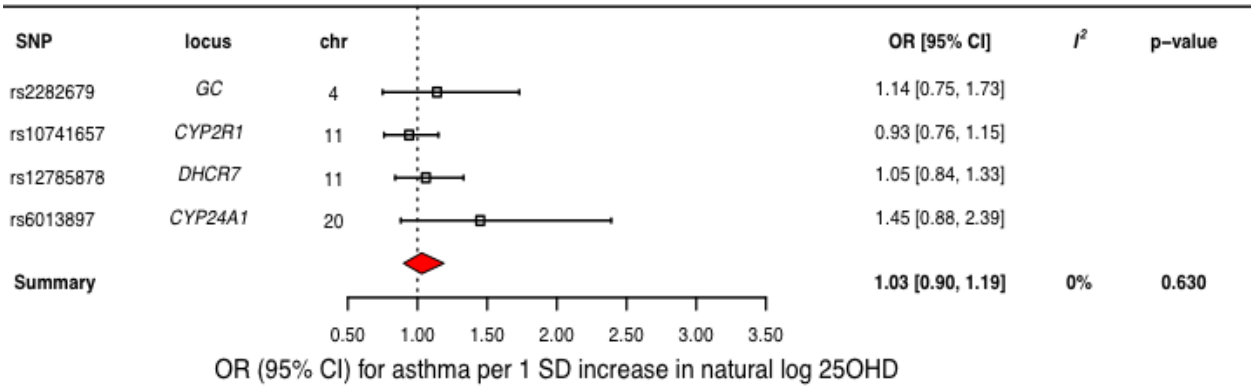
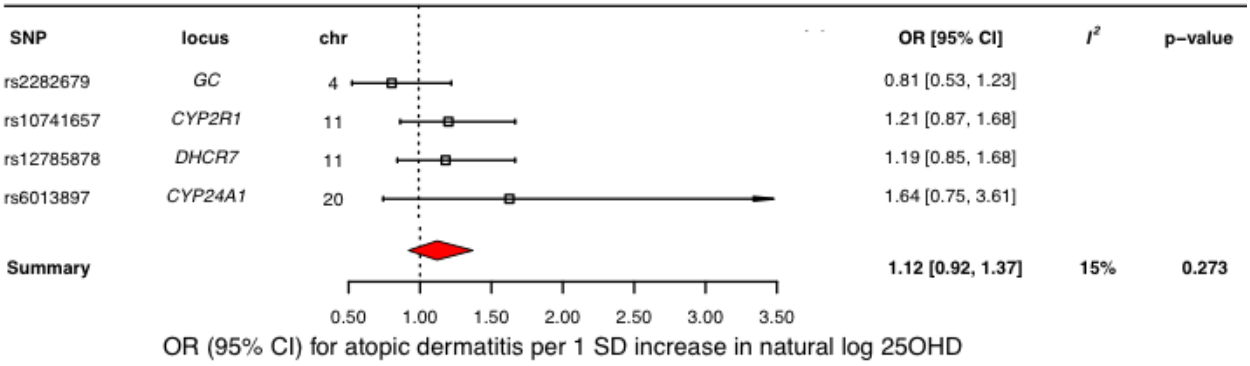


Figure 4: Mendelian Randomization Estimate of the Association of 25OHD Levels with Risk of Asthma



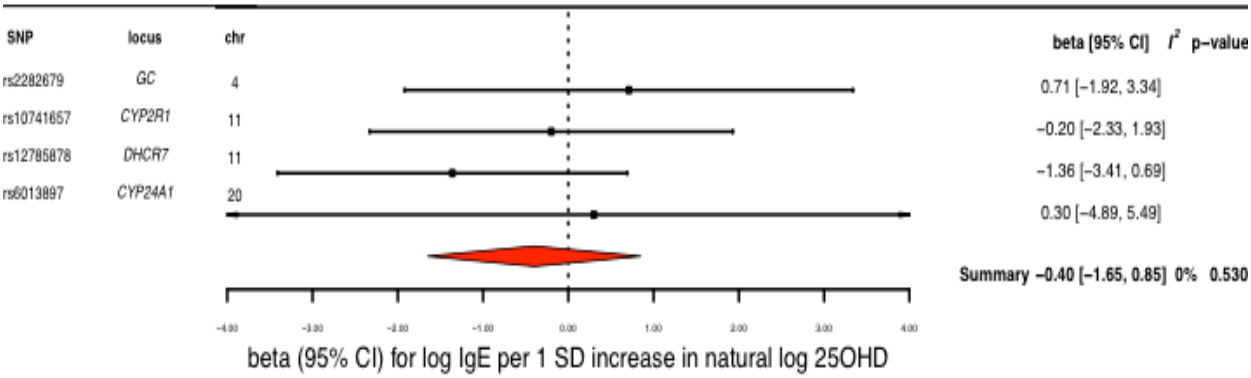
Estimates obtained from using a fixed-effects model

Figure 5: Mendelian Randomization Estimate of the Association of 25OHD Levels with Risk of Atopic Dermatitis



Estimates obtained from using a fixed-effects model

Figure 6: Mendelian Randomization Estimate of the Association of 25OHD Levels with IgE levels



Estimates obtained from using a fixed-effects model

4.10 Supplemental Material

4.10.1 S1 Text

Phenotype Definition in the Participating Studies

The UK Biobank Asthma GWAS was conducted using baseline phenotypic data collected in the UK Biobank study. UK Biobank[1] comprises ~500,000 people aged between 40-69 years, recruited between 2006 and 2010 across Great Britain[2]. Genotypic data were available in 120,286 UK Biobank participants with known asthma status (15,106 cases/105,180 controls). Asthma was defined as self-reported asthma, without further specification on the age of asthma onset.

The GABRIEL study is a meta-analysis of 37 GWAS studies of individuals of European descent (10,003 cases/13,954 controls), of which 20 studies included individuals with childhood-onset asthma (7,047 cases/7,961 controls), aiming to identify risk factors for asthma[3]. Asthma was considered present if diagnosed by a physician. Childhood-onset asthma was defined as onset of asthma before the age of 16 years. After quality control and filtering, genotypes were retained on 10,365 asthma cases and 16,110 controls from 37 different studies, including studies with both childhood onset and adult onset asthma. Only one of the four SUNLIGHT SNPs used as instruments for the MR was present in the published genotyped dataset of the GABRIEL consortium. Thus, we opted to use estimates on the effect of the four SUNLIGHT SNPs on asthma from a meta-analysis of Hapmap2-imputed GABRIEL data (unpublished), comprising all 37 studies of the original GABRIEL meta-analysis. We used a meta-analysis of data imputed in the same panel for the 20 childhood-onset asthma studies for our secondary analysis. The UK Biobank

and GABRIEL summary level results for the four SNPs were then pooled in a fixed-effects meta-analysis using the GWAMA software[4].

For atopic dermatitis, we obtained the effects of the four SUNLIGHT 25OHD-associated SNPs on this disease using data from The EAGLE Eczema Consortium[5]. EAGLE is a large meta-analysis of 26 studies, including 116,863 individuals (21,399 cases/95,464 controls), imputed to the 1000 Genomes Project Phase 1 reference panel. The data used in this MR study were obtained from the fixed-effect meta-analysis of 22 European studies (excluding 23andMe). This resulted in a sample of 10,788 cases and 30,047 controls. Phenotype definition of atopic dermatitis differed among studies and was either self-reported or based on dermatological exam.

For IgE levels, we obtained the effects for the four SUNLIGHT SNPs from 34 of the 37 studies of the GABRIEL consortium. The GWAS for IgE was conducted separately in studies with asthma cases only (total n=5,888) and in studies with controls only (n=6,965), and the results were combined by fixed-effects meta-analysis within cases (17 datasets) and within controls (17 datasets). The combination of these two pooled results, inversely weighted for their variance, is arithmetically equivalent to doing a single fixed-effects meta-analysis of all the study-stratum-specific results (from 34 datasets) in one calculation. Therefore the betas from this analysis represent the pooled estimates of the effect of the SNPs on naturally log transformed total IgE, derived from a fixed-effects (inverse-variance-weighted) meta-analysis of the results from the asthmatics and from the controls.

PubMed Search for Pleiotropy

The following terms were searched on the PubMed database to investigate possible pleiotropic mechanisms of our chosen SNPs corresponding to gene name, gene mutations, encoded protein, encoded protein with asthma, atopic dermatitis and IgE levels.

For rs2282679: *"GC", "GC* gene", "*GC* gene mutations", "vitamin D binding protein", "vitamin D binding protein asthma", "vitamin D binding protein atopic dermatitis", "vitamin D binding protein IgE levels"

The search term GC uncovered 69,152 results, most of which were not relevant to genetics, therefore the search term "GC gene" was used instead to refine search results.

For rs12785878: *"DHCR7", "DHCR7* mutations", "7-dehydrocholesterol reductase", "7-dehydrocholesterol reductase asthma", 7-dehydrocholesterol reductase atopic dermatitis", "7-dehydrocholesterol reductase IgE levels".

For rs6013897: "*CYP24A1", "CYP24A1* mutations", "1,25-dihydroxyvitamin D3 24-hydroxylase", "1,25-dihydroxyvitamin D3 24-hydroxylase asthma" "1,25-dihydroxyvitamin D3 24-hydroxylase atopic dermatitis","1,25-dihydroxyvitamin D3 24-hydroxylase IgE levels"

For rs10741657: "*CYP2R1", "CYP2R1* mutations", "vitamin-D hydroxylase", "vitamin-D hydroxylase asthma", "vitamin-D hydroxylase atopic dermatitis", "vitamin-D hydroxylase IgE levels".

Abstracts were selected for further review if they made reference to the search term and a pathway distinct from vitamin D or vitamin D insufficiency/ deficiency on the three outcomes. Only studies in mammals were considered. Findings are reported in the Results section.

References

1. Ollier W, Sprosen T, Peakman T. UK Biobank: from concept to reality. Pharmacogenomics 2005;6:639-46.

2. Sudlow C, Gallacher J, Allen N, 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 2015;12:e1001779.

3. Moffatt MF, Gut IG, Demenais F, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363:1211-21.

4. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 2010;11:288.

5. Paternoster L, Standl M, Chen CM, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet 2012;44:187-92.

4.10.2 S1 Table

S1 Table. P-values of the association of the SNPs used as instrumental variables with potential confounders

25OHD-SNP Look-up in GWAS Consortia				
Trait	**p for association with *GC* (rs2282679)**	**p for association with *CYP2R1* (rs10741657)**	**p for association with *DHCR7* (rs12785878)**	**p for association with *CYP24A1* (rs6013897)**
BMI[a]	0.91	0.29	0.78	0.61
Smoking Quantity[b]	0.80	0.90	0.18	0.73

[a] p-value for association between 25OHD SNP and BMI obtained from the GIANT Consortium[1]

[b]p-value for association between 25OHD SNP and smoking quantity obtained from the Tobacco and Genetics Consortium[2]

References

1. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012;44(6):659-69. doi: 10.1038/ng.2274. PubMed PMID: 22581228; PubMed Central PMCID: PMCPMC3613127.

2. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet. 2010;42(5):441-7. doi: 10.1038/ng.571. PubMed PMID: 20418890; PubMed Central PMCID: PMCPMC2914600.

Chapter 5: Genome-wide association study for vitamin D levels reveals 63 novel loci

Preface: Bridge between Chapters 2, 3, 4 and Chapter 5

In Chapter 5, we took advantage of the UK Biobank's 25OHD levels release, which came out two years after the GWAS meta-analysis of Chapter 2, to perform the largest to date GWAS on 25OHD levels. In collaboration with some of the authors of our previous GWAS, we analyzed data from the UK Biobank under a very competitive environment. We therefore identified 69 vitamin D loci, among which 63 are novel, harboring 138 independent common, low frequency and rare variants. This manuscript is currently under review at the American Journal of Human Genetics. Chapters 3 and 4 showed that identifying genetic determinants of vitamin D enables interrogation of a causal role of vitamin D in health outcomes through MR. In Chapter 5, we unveil an enhanced set of genocopies of vitamin D status, which can be used in the future to test causal associations between vitamin D levels and complex human diseases and traits. At the same time, we highlight that these novel genetic instruments for 25OHD should be used with caution in future MR analyses, given that they lie mostly in genes outside the canonical vitamin D metabolic pathway, and therefore they carry with them an increased risk for pleiotropy.

Title page

Title: Genome-wide association study for vitamin D levels reveals 63 novel loci.

Authors: Despoina Manousaki*[1,2], Ruth Mitchell*[3], Tom Dudding[3,4], Simon Haworth[3,4], Adil Harroud[5], Vince Forgetta[2], Rupal L. Shah[6], Jian'an Luan[6], Claudia Langenberg[6], Nicholas J. Timpson[3] and J. Brent Richards[1,2,7,8,9]

[1] Department of Human Genetics, McGill University, Montreal, Quebec, Canada

[2] Centre for Clinical Epidemiology, Department of Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

[3] MRC Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[4] Bristol Dental School, University of Bristol, Bristol, United Kingdom

[5] Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada

[6]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

[7] Department of Medicine, McGill University Montreal, Quebec, Canada

[8]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

[9] Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom

* denotes equal contribution

Correspondence:

J. Brent Richards, Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, 3755 Côte Ste-Catherine Road, Montreal, Quebec H3T 1E2

Tel: 514-340-8222, Fax: 514-340-7502

Email: brent.richards@mcgill.ca

Word count (main text): 3,896

Number of references: 70

Number of figures and tables: 4

Number of supplemental tables: 9

Abstract

We aimed to increase our understanding of the genetic determinants of vitamin D levels by undertaking a large-scale genome-wide association study (GWAS) of serum 25 hydroxyvitamin D (25OHD). To do so, we used imputed genotypes from 401,460 white British UK Biobank participants with available 25OHD levels, retaining single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) > 0.1%, and imputation quality score > 0.3. We performed a linear mixed model GWAS on standardized log-transformed 25OHD, adjusting for age, sex, season of measurement and vitamin D supplementation. These results were combined with those from a previous GWAS including 42,274 Europeans. *In silico* functional follow-up of the GWAS results was undertaken to identify enrichment in gene sets, pathways and expression in tissues, and to investigate the partitioned heritability of 25OHD, and its shared heritability with other traits. Using this approach, the SNP heritability of 25OHD was estimated to 16.1%. 138 conditionally independent SNPs were detected (p-value< 6.6 $x10^{-9}$) among which 53 had MAF<5%. Single variant association signals mapped to 69 distinct loci, among which 63 were novel. We identified enrichment in hepatic and lipid metabolism gene pathways, and enriched expression of the 25OHD genes in liver, skin and gastrointestinal tissues. We observed partially shared heritability between 25OHD and socio-economic traits, a feature which may be mediated through time spent outdoors. Therefore, through the largest 25OHD GWAS to date, we identified 63 novel loci, which underline the contribution of genes outside the vitamin D canonical metabolic pathway to the genetic architecture of 25OHD.

5.2 Introduction

Vitamin D status, as ascertained by 25-hydroxy-vitamin D level (25OHD), is associated with numerous health outcomes[1]. However, it is unclear if lowered 25OHD level plays a causal role in these outcomes and its exact biological mechanisms of action remains unknown[2; 3]. 25OHD is a steroid pro-hormone and a fat-soluble metabolite of cholecalciferol, which is predominately synthesized by exposure to ultra-violet light or obtained from dietary sources including fortified foods, supplements and oily fish. It plays an important role in regulating calcium and phosphorus concentrations and influences cell proliferation, differentiation, apoptosis and has immune modulating effects[4]. Understanding the etiology of low vitamin D levels could have important public health implications by prioritizing individuals who would benefit from supplementation. The body's vitamin D stores are best reflected by serum 25OHD, which is influenced not only by diet and exposure to ultra-violet light, but also by age, body mass index, skin color, and numerous factors regulating exposure to ultra-violet B radiation (including season, geographical latitude, skin coverage)[5; 6]. In addition to these environmental factors, classical twin studies show that 50-80% of the variability in the concentration of 25OHD is explained by genetic factors[7; 8] indicating that this is a highly heritable trait.

In recent years, several genome-wide association studies (GWAS) of serum 25OHD have been conducted on participants of Europeans ancestry, with the largest including 79,366 individuals[9]. These studies have identified six independent common genetic variants (minor allele frequency

[MAF] >5%) which are associated with 25OHD level.[9-12] These variants are in loci near genes having an established role in vitamin D synthesis (*DHCR7*/*NADSYN1* [MIM: 602858] (rs12785878) and *CYP2R1* [MIM: 608713] (rs10741657)), transportation (*GC* [MIM: 139200] (rs2282679)) and degradation (*CYP24A1* [MIM: 126065] (rs17216707)), as well as outside of known vitamin D metabolism pathways, such as *SEC23A* (Sec23 homolog A, coat protein complex II component [MIM: 610511], rs8018720), involved in endoplasmic reticulum (ER)-Golgi protein trafficking, and *AMDHD1* (amidohydrolase domain containing 1, rs10745742) an enzyme involved in the histidine, lysine, phenylalanine, tyrosine, proline and tryptophan catabolic pathway[9]. Additionally, a low frequency genetic variant (MAF <5%) at *CYP2R1* (rs117913124), with a four-fold larger effect than common variants at that locus was identified through whole-genome sequencing and deep imputation for low-frequency and rare variants[12].

An improved understanding of the genetic determinants of 25OHD has helped re-assess the role of vitamin D in the aetiology of complex diseases, such as musculoskeletal disorders[1], autoimmune disease, such as multiple sclerosis[13-23] and cancer[24] , through methods for causal inference, such as Mendelian randomization (MR) [25; 26]. For example, four separate MR studies have supported a protective effect of vitamin D against multiple sclerosis [12-14; 27], and these results have clinical implications, reflected in recent clinical care guidelines for the use of vitamin D in preventing multiple sclerosis in those at risk, published by the MS Society of Canada[28]. More than 60 MR studies have been published to date utilising genetic variants associated with 25OHD to aid causal effect estimation[29-46]. A deeper understanding of the genetic determinants contributing to variation

in circulating vitamin D levels could enable an improved instrumentation of vitamin D in MR studies, allow better genomic prediction of vitamin D levels and provide insights into biological mechanisms.

Although the most recent 25OHD GWAS study on 79,366 Europeans[9] had double the sample size of the previous GWASs, it yielded only two new 25OHD loci (the *SEC23A* and *AMDHD1*), indicating that 25OHD may be a metabolite with a simple oligogenic architecture. In the same study, little of the 25OHD heritability estimated using all common SNPs was explained (SNP heritability of 7.5%), suggesting that much of its heritability remains to be identified. Against this backdrop, we sought to further understand the phenotypic variance explained by genetic variants and investigate the genetic architecture of 25OHD by increasing substantially the GWAS sample size.

We hypothesized that we could identify new enzymes, or carrier proteins affecting the levels of this metabolite, unveiling a classical polygenic architecture, or confirm its oligogenic architecture. We therefore undertook a GWAS of serum 25OHD levels in 401,460 White British individuals from UK Biobank and combined results of this GWAS in a meta-analysis with results from a previous GWAS study including up to 42,274 Europeans. Using this approach, we validated previously described 25OHD loci and identified novel genetic determinants of vitamin D. To gain further insight into the genetic control of the vitamin D metabolic pathway, we looked for overlap

of our findings with those of the an unpublished GWAS on 1,25-dihydroxyvitamin D, the active form of vitamin D, which is downstream from 25OHD in the vitamin D metabolic pathway (**Figure 1**). Finally, we undertook an *in silico* functional follow-up of our GWAS findings, to identify enrichments in gene sets, pathways, and expression in tissues, and explore the partitioned heritability of 25OHD and its shared genetic architecture with other GWAS traits.

5.3 Methods

5.3.1 Phenotype

Between 2006 and 2010 approximately half a million British adults were recruited by UK Biobank[47]. Participants provided biological samples, physical measurements, and answered questionnaires relating to general health and lifestyle. Ethical approval was granted by the Northwest Multi-Centre Research Ethics Committee, and informed consent was obtained from all participants prior to participation.

Data on 25OHD level (in nmol/L) measured using the Diasorin assay were available from 465,415 samples, representing 449,978 UK Biobank participants. Measurements were performed at baseline (2006-2010), and/or the first follow-up visit (2012-2013). In the present study, we used baseline 25OHD measurements from 401,460 individuals from the White British subset of UK Biobank, as defined below. To account for vitamin D supplement use, we adjusted 25OHD levels by subtracting 21.2 nmol/L from the 25OHD measurement for vitamin D supplement users (see

Supplemental Material and Methods for definition of vitamin D supplementation). We used 21.2 nmol/L because it is the mean increase in 25OHD levels conferred by taking daily 400IU of cholecalciferol, the amount of vitamin D most often found in vitamin D supplements[48]. Whenever 25OHD levels were lower than 10nmol/L (the detection threshold for Diasorin assay) after subtraction, they were set to 10nmol/L. 25OHD levels were then log transformed and standardized to a mean of 0 and standard deviation of 1 (because of skewness in the distribution of 25OHD levels, and to allow comparison with previous 25OHD GWAS). Distribution of the 25OHD levels appears in **Figure S1.**

5.3.2 GWAS

After stringent quality control, the UK Biobank genotypes, imputed to the combined Haplotype Reference Consortium (HRC)[49] and UK10K haplotype resource panel, provided 20,370,874 genetic variants from the autosomes and the X chromosome to test for their association with 25OHD levels. This quality control removed low quality genetic variants, by retaining only SNPs with a minor allele frequency (MAF) > 0.1%, imputation quality score of >0.3 and Hardy–Weinberg $P > 1 \times 10^{-6}$. For details on genotyping and imputation in UK Biobank see the **Supplemental Material and Methods.**

To minimize bias from population stratification, an issue which is particularly relevant in the search for rare genetic variants associated with traits and disease[50], analysis was restricted to

individuals of White British ancestry, which comprises the largest single ancestral group represented in the UK Biobank. It is important to distinguish between the self-identified "White British" in UK Biobank, and the White British subset used in our analysis, where the latter was defined using a principal component analysis. Specifically, we previously defined this White British subset using high-quality genotypes, employing FlashPCA [51] and linkage-disequilibrium-pruned HapMap3 SNPs (MAF > 1%, minor allele count > 5, Hardy-Weinberg Equilibrium P > $1x\ 10^{-6}$), which were projected onto previously computed principal components using the same SNPs set from 1000 Genomes Phase 3 dataset (N=2,504)[52]. Henceforth, whenever the term "White British" appears in this paper, it refers to the White British subset defined as above. Details on this analysis are provided in the **Supplemental Material and Methods.** Descriptive statistics of this White British subset of UK Biobank are detailed in **Table S1.**

We then tested the additive allelic effects of SNPs on 25OHD levels, using a linear mixed-model in the BOLT-LMM software[53]. The model-fitting was performed on hard-called genotypes from 488,377 participants consisting of 803,113 SNPs. Age, sex, season of 25OHD measurement (as a categorical variable; 1 for Winter [January to March];2 for Spring [April to June];3 for Summer [July to September], and 4 for Fall [Oct to Dec]), genotype batch, genotype array, and assessment center (as a proxy for latitude) were included as covariates in the BOLT-LMM. We have previously estimated that 6.6×10^{-9} is an appropriate p-value threshold for genome wide significance for analyzing data from the UK Biobank using the above criteria, accounting for multiple testing[52].

5.3.3 Meta-analysis

We compared the results of the GWAS on UK Biobank to those of a previous 25OHD GWAS published by our group (n=42,274 samples of European ancestry)[12], by performing Pearson correlation of the betas of all variants with p-values $<1 \times 10^{-6}$ in both GWAS using the 'cor.test' function in R. We then combined the summary level results of the two GWAS in an inverse variance weighted fixed effects meta-analysis, using the GWAMA[54] software.

Approximate conditional association analysis

To identify conditionally independent SNPs from this meta-analysis, we used GCTA-COJO version 1.91.1[55; 56], which conditions upon the lead SNP per locus by approximating the genotype-phenotype covariance with correlation matrices and summary statistics (**Supplemental Material and Methods**). Variants with high collinearity (multiple regression $R^2 > 0.9$) were excluded, and those situated more than 20,000 pairs away were assumed to be independent. A reference sample of 50,000 unrelated white British individuals randomly selected from the UK Biobank was created for a previous GWAS[52], and was used to model patterns of linkage disequilibrium (LD) between variants. We retained as conditionally independent variants those reaching a genome-wide significant p-value pre- and post-conditioning, and with at least one genome-wide significant satellite SNP within 250,000 pairs. These variants were then positionally and functionally annotated to the physically closest gene using the hg19 gene range list, and the Variant Effect Predictor[57] as implemented in PhenoScanner v2.[58]

5.3.4 Estimation of variance explained by significant variants and SNP heritability

We estimated the proportion of 25OHD phenotypic variance tagged by all SNPs on the genotyping array (that is, the SNP heritability) using BOLT-REML function [53] in the UK Biobank GWAS. To estimate the variance explained by independent genome-wide significant SNPs (that is, all the genome-wide significant conditionally independent lead SNPs), we summed the variance explained per independent SNP using the formula: variance explained $\approx 2\beta^2 f(1-f)$, where β and f denote the effect estimate and the effect allele frequency of the allele on a standardized phenotype, respectively[59].

5.3.5 Assessment of inflationary bias in GWAS results

By estimating the lambda GC and the LD score regression (LDSR) intercept, BOLT-LMM software estimated the amount of genomic inflation present in the data that was due to residual population stratification, cryptic relatedness, and other latent sources of bias in the UK Biobank GWAS. We used the lambda GC from GWAMA to estimate the genomic inflation in the meta-analysis of the UK Biobank GWAS and compared this with the previous GWAS meta-analysis[12].

5.3.6 *In silico* functional follow-up

Functional follow-up of the meta-analysis summary statistics was performed using Complex Trait Genomic-Virtual Lab[60] web application (https://genoma.io), which implements a variety of follow-up methods for GWAS summary statistics output from the COJO analysis (**Supplemental Material and Methods**). In brief, association between predicted gene transcription and 25OHD was estimated using S-MultiXcan[61] in the MetaXcan package with the default options implemented. Association statistics for the 48 tissues were combined accounting for correlation between tissues to give transcript-level results, and a Bonferroni correction was applied to account for the number of gene transcripts tested. Gene prioritisation, gene set and tissue enrichment analysis were performed using DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits)[62; 63] to identify likely causal genes at associated loci, highlight gene pathways which are over-represented by associated loci in the single variant results and test whether expression of these genes is enriched in specific tissue types. Genetic correlation between 25OHD and a range of other traits available as publicly available GWAS summary statistics was examined using bivariate LDSR[64] implemented in the LD Hub platform[65]. Finally, partitioned heritability by functional annotation with 53 overlapping categories was performed using stratified LDSR using the baseline model from 1000 Genomes phase 3 data (baselineLD_v2.2, February 2019)[64; 66]. Cell specific heritability was examined using the --h2-cts flag in LDSR and the multi-tissue gene expression file ("Multi_tissue_gene_expr" containing both GTEx data and Franke lab dataset of microarray gene expression)[65]. These final two analyses were restricted to common variants

present in HapMap3 (approximately 1,500,000 SNPs), excluding those within the HLA region defined as Chr6: 25000000 to 34000000 bases inclusive.

5.3.7 GWAS on 1,25-dihydroxyvitamin D

Study participants, genotyping and imputation

The Ely Study, established in 1990, is a prospective study of the aetiology of type 2 diabetes and has been described in detail elsewhere. We studied Ely participants with measures of 1,25-dihydroxyvitamin D to estimate genetic effects the active form of vitamin D[67; 68]. Briefly, Ely comprises individuals of European ancestry aged 40-69 years, registered at a single medical practice in Ely, Cambridgeshire, UK and evaluated in 3 phases. All participants of the Ely Study gave their written informed consent and the study was approved by the local ethics committee. Participants at Phase 3 were genotyped using the HumanCoreExome-24 and InfiniumCoreExome arrays. Details of the genotype quality control appear in **Supplemental Material and Methods.** A total of 1,591 samples and 546,486 variants met the quality control criteria. Imputation was performed using the Sanger Imputation Server (pre-phase with EAGLE2 and impute with PBWT pipeline), and the HRC 1.1 reference panel[49]. Additional variants not captured by the HRC reference panel were imputed using a combined UK10K and 1000 Genomes Phase 3 reference panel resulting in data available for >14 million variants.

1,25-dihydroxyvitamin D phenotype and look-up for the 25OHD conditionally independent SNPs

Phase 1 1,25-dihydroxyvitamin D levels and genetic data were available for 748 Ely participants. Levels of 1,25-dihydroxyvitamin D were natural log transformed before regressing with the inclusion of age, sex, body mass index and season as covariates. Residuals from the regression were standardised and used as the final 1,25-dihydroxyvitamin D phenotype. Genetic association analysis was performed for the conditionally independent variants from the 25OHD GWAS meta-analysis using SNPTEST v2.5.4-beta3[69]. Bonferroni adjustment was applied to association test p-values such that variants with GWAS p-values $<4.10 \times 10^{-4}$ (0.05/122) were considered to meet the corrected significance threshold.

5.4 Results

5.4.1 GWAS for 25OHD levels

The GWAS in UK Biobank included 401,460 participants and 20,370,874 variants. The genomic control lambda in BOLT-LMM was 1.23, and the LDSR intercept was 1.06. We found a strong correlation between the effect sizes of the UK Biobank GWAS with our previous GWAS meta-analysis[12]. Specifically, we compared the betas of 20,787 SNPs achieving p-values $< 1 \times 10^{-6}$ in both GWAS (minimum MAF 0.3%) and found a coefficient of correlation (r) of 0.88 (**Figure S3**). We then performed a meta-analysis of the two GWAS on a total of 16,668,957 SNPs (**Figure 2**). The lambda GC of the meta-analysis was 1.23. Using approximate conditional analysis as

implemented by GCTA–COJO, we observed 138 conditionally independent signals (pre- and post-conditioning p-value< 6.6 x10^{-9}), mapping to 69 loci (a locus was defined as 1 Mb region around the SNP reaching the lowest p-value), 63 of which were not reported in previous 25OHD GWAS (**Table S2**). Of these conditionally independent SNPs, 53 (38%) had MAF<5%, and 85 (62%) were common (MAF≥5%). The 53 SNPs with MAF <5% conferred an average absolute effect of 0.23 standard deviations on standardized log transformed 25OHD levels per effect allele, compared to 0.03 standard deviations for the 85 SNPs with MAF≥5%.

The total variance explained by the 138 conditionally independent genome-wide significant vitamin D SNPs was 4.9%. When partitioning the variance explained by these lead SNPs into two MAF categories, we found that low-frequency and rare variants explained 1.8% of the variance in 25OHD levels, whereas common variants explained 3.1% of the variance, respectively. The SNP heritability from all SNPs, independent of GWAS p-value, as estimated by BOLT-LMM on 805,426 hard called variants in UK Biobank was 16.1%, indicating that genome-wide significant independent variants capture less than a third of the variance explained in 25OHD levels by all directly genotyped markers.

5.4.2 Look-up of the 25OHD GWAS variants in the 1,25-dihydroxyvitamin D GWAS

We tested 122 out of the 138 conditionally independent variants from the 25OHD GWAS for genetic association with 1,25-dihydroxyvitamin D. The 16 variants that were not tested were not

available in the Ely dataset, either because they were not reliably captured through imputation, or had low MAF (<0.001), and no suitable proxy variant could be identified. Among the 122 conditionally independent variants tested in Ely for association with 1,25-duhydroxyvitamin D, only one rs6127099 in the *CYP24A1* locus on chromosome 20 reached the multiple testing corrected threshold for significance (20:52731402:T_A; β=0.231; p=2.5 x 10^{-4})(**Table S2).**

5.4.3 *In silico* functional follow-up

Gene prioritisation and enrichment analyses

Gene prioritisation analysis suggested 70 genes with FDR<5% which might plausibly underlie the distribution of association statistics seen in the single variant results. At many loci, genes within the vitamin D metabolism pathway were suggested as plausible candidates. For example, DEPICT prioritized *DHCR7* at the lead associated chr11:70313961-71239227 locus and *GC* at chr4:72607410-72669758 locus. Interestingly, *ADH6* [MIM:103735] was a plausible candidate at locus chr4:99916771-100274184 suggesting this locus may have pleiotropic effects on vitamin D and alcohol metabolism (**Table S3**).

Gene set enrichment analysis identified enrichment in 418 pre-defined gene sets with a false discovery rate (FDR) < 5%. The strongest statistical evidence for enrichment was in the following gene sets: the alpha-2-HS Glycoprotein (AHSG), a negatively-charged serum glycoprotein that is

synthesized by hepatocytes involved in several processes, including endocytosis, brain development, and the formation of bone tissue ($p=4.18x10^{-7}$); the reactome gene set for "metabolism of lipids and lipoprotein" ($p=7.91x10^{-7}$); several genes involved in immune pathways and therefore expressed in the blood such as 'Elastase, Neutrophil Expressed (ELANE)' ($p=8.43x10^{-7}$); the 'Serum albumin (ALB)' ($p=1.19x10^{-6}$), 'Acidic form of complement factor 4 (C4A)' ($p=1.51x10^{-6}$) and 'ENSG00000211949' gene sets, belonging to the immunoglobulin (Ig) heavy chain locus ($p=1.51x10^{-6}$); biosynethic pathways such as "GO:0044283, small molecule biosynthetic process, $p=1.89x10^{-6}$", "GO:0016053, organic acid biosynthetic process, $p=2.29x10^{-6}$; GO:0046394" and "carboxylic acid biosynthetic process, $p=2.29x10^{-6}$"; and finally liver associated pathways including "MP:0000599, enlarged liver, $p=1.33x10^{-6}$", "GO:0001889, liver development, p=3.35x $x10^{-6}$" and "GO:0061008, hepatobiliary system development, $p=4.15x10^{-6}$" (**Table S4**). Collectively these results suggest that detectable serum levels of 25OHD are influenced by a range of metabolic processes extending within and beyond the canonical vitamin D metabolic pathway.

Finally, expression of 25OHD genes was enriched in 17 cell types with an FDR < 5%, including cell lines representing the liver (hepatocytes, $p=1.63x10^{-6}$) and skin (keratinocytes, $p=7.73x10^{-3}$). The tissue-specific analysis found greatest evidence for enrichment in the liver ($p=1.34x10^{-6}$) and the gastrointestinal tract ($p=2.22x10^{-3}$) (**Table S5**).

Predicted gene transcription levels

After applying a Bonferroni-corrected multiple testing threshold ($p<1.94x10^{-6}$), varying expression levels at 377 gene transcripts were predicted to influence 25OHD, out of a total of 25,816 that were tested. Results for all gene transcripts are shown in **Figure 3**. This indicates that although there are 69 loci associated with vitamin D phenotype, there are potentially 377 gene transcripts across multiple tissues whose expression may influence vitamin D. The lead associated genetic transcripts using S-MulTiXcan[61] were consistent with the lead association signals in the single variant results, for example identifying association at *NADSYN1* [MIM:608285](Z-test $p<1.81x10^{-309}$); *DHCR7* (Z-test $p<1.15x10^{-245}$); *GC* (Z-test $p<1.81x10^{-309}$); *CYP2R1* (Z-test $p=2.85x10^{-277}$); *UGT1A4* [MIM:606429] (Z-test $p=3.25x10^{-34}$); *PADI1* [MIM: 607934] (Z-test $p=3.64x10^{-23}$). The S-MulTiXcan[61] method integrates information from multiple tissue-specific predictions improving the statistical power over the single variant method and highlights additional transcripts associated with 25OHD, with the strongest evidence in various forms of *Keratin Associated Protein 5* (*KRTAP5* [MIM:608822]) (Z-test $p<1.81x10^{-309}$), a protein coding gene involved in keratinization and has been identified as a potential read through for *NADSYN1*. This adds further evidence that 25OHD is affected through processes beyond the established vitamin D metabolic pathway. Results are shown in **Table S6.**

Genetic correlation

Genetic correlation results for 25OHD were available for 774 traits from the LD hub catalogue[65], including 517 raw traits from UK Biobank and 257 from other GWAS studies and consortia

(**Figure 4**). A total of 101 traits passed a multiple testing corrected Bonferroni p-value threshold of $p<6.46x10^{-5}$. The strongest evidence of negative genetic correlation with 25OHD were 'Time spent using a computer' (r_g=-0.22) and 'Qualifications: College or University degree' (r_g=-0.17); 'Intelligence' (r_g=-0.24). Traits pertaining to exercise ('Duration of vigorous activity' (r_g=0.22) and 'Number of days/week walked 10+ minutes' (r_g=0.18)) had positive genetic correlations with vitamin D. Traits related to body mass index (BMI) including lipids and diabetes, had a negative correlation: 'BMI' (r_g=-0.14); 'Triglycerides' (r_g=-0.25); 'Type 2 Diabetes' (r_g=-0.19). A full list of results can be found in **Table S7**. This highlights that the genetic determinants of 25OHD levels capture biological influences that extend beyond the canonical synthesis and metabolic pathways of this metabolite.

Tests for enrichment in functional annotations

Using information from all the SNPs in the 25OHD GWAS summary statistics and modelling LD with the 53 functional categories not specific to any cell type in the baseline model, there was evidence for enrichment in 3 out of the 95 functional annotations tested. These were annotations providing evidence for evolutionary conservation with 2% of variants annotated as highly conserved accounting for 20% of the heritability of vitamin D (9-fold enrichment over baseline, $p=1.48x10^{-5}$) (**Table S8**). There was little evidence from stratified LDSR[66] that vitamin D heritability is enriched in gene sets expressed specifically in given cells or tissue types. However, it is worth noting that the highest LDSR coefficients were seen for genomic regions specifically

expressed in hepatocytes (coefficient = $1.17x10^{-8}$), liver (coefficient = $1.73\ x10^{-8}$) and whole blood (coefficient= $1.16\ x10^{-8}$), corroborating the cell and tissue predicted gene enrichment (**Table S9**).

5.5 Discussion

This large-scale GWAS meta-analysis identified 63 novel genetic loci which were associated with 25OHD levels in people of European ancestry and at least doubled the estimate of SNP heritability of 25OHD levels. Our study also replicated the 6 known vitamin D loci (in or near *CYP2R1, DHCR7, GC, CYP24A1, AMDHD1, SEC23A)*. *In silico* follow-up identified enrichment in gene sets and pathways mostly independent from canonical vitamin D synthesis and metabolism pathways. Taken together, these results identify new biological pathways that influence 25OHD levels and demonstrate that this metabolite is moderately polygenic.

The hypothesis-free approach of GWAS has served to highlight the role of lipid biology in 25OHD levels—a fat-soluble hormone. Specifically, among the 69 identified 25OHD loci, 22 loci are related to serum lipid phenotypes. Examples of these loci are the lipase C gene *(LIPC* [MIM:151670]*)* on chromosome 15, the low density lipoprotein receptor gene (*LDLR* [MIM:606945]) and the apolipoprotein C1 gene (*APOC1* [MIM107710]) on chromosome 19, and the cholesteryl ester transfer protein gene (*CETP* [MIM:118470]) on chromosome 16. Additionally, our gene enrichment analysis prioritized the metabolism of lipids and lipoprotein gene set, and lipid traits were strongly genetically correlated with 25OHD using LDSR. These

findings suggest that 25OHD levels share several of the same biological pathways influencing circulating lipids.

We also found enrichment in loci harboring genes associated with skin keratinization. Among these, an interesting finding was the *FLG* gene [MIM:135940] on the chromosome 1, which encodes fillagrin, a protein which plays an important role in the skin barrier's function, and deregulation of this function might affect vitamin D in the skin, which is also synthetized in the skin. Another locus related to skin keratinization was the *KRTAP5* gene, which was prioritized by our *in silico* analyses. However, functional follow-up of these novel loci is required, to characterize the causal genes and/or mechanisms underlying the associations with 25OHD levels. Also, we observed enrichment in loci associated with traits outside the vitamin D pathway, which are not directly linked to 25OHD synthesis and metabolism. We can speculate on the exact mechanism of action of these genes on 25OHD—for instance through their effect on time spend outdoors and consequently exposure to sunlight—but follow-up experiments are necessary to validate these hypotheses. Collectively these results suggest that serum levels of 25OHD are in crosstalk with a range of metabolic processes extending within and beyond the canonical vitamin D metabolic pathway. This highlights that the genetic instrument for vitamin D is instrumenting more than the vitamin D pathway, and specifically also captures variance in traits that relate to environmental confounders that could influence 25OHD levels. Taken together, our findings present a cautionary tale for future MR studies, since there is a risk of pleiotropic effects for a substantial number of novel 25OHD-related SNPs mapping to genes not directly involved in 25OHD biology.

The findings of the look-up of the significant 25OHD SNPs in the 1,25-dihyxroxyvitamin D GWAS provide evidence that the two biomarkers of vitamin D in humans have, to a certain extent, a shared genetic component. This may be expected as both biomarkers share the same vitamin D catabolic pathway. However, the small sample size of the 1,25-dihydroxyvitamin D GWAS, the only available GWAS on this trait to date, limits the power for characterization of 1,25-dyhydroxyvitamin D loci. We can therefore speculate that there might be a larger overlap of the genetic architecture of the two biomarkers. 1,25-dihydroxyvitamin D is the active metabolite of vitamin D, and although its levels directly regulate the effects of vitamin D on a cellular level, it remains understudied because of its short half-life, low concentration in blood[70] and the body's ability to buffer 1,25-dihydroxyvitamin D in deficient individuals by increasing parathyroid hormone. In that aspect, any additional evidence, from larger 1,25-dihydroxyvitamin D GWAS, linking 25OHD levels to those of 1,25-dihydroxyvitamin D in the genetic level will be important, as it will add to our understanding of the vitamin D physiology.

In summary, we described 63 novel loci which are associated with 25OHD levels in Europeans. Further research is warranted to better characterize the novel genetic variants, validate these findings in other ethnic groups, and to better understand the biological pathways influencing 25OHD levels. The novel genetic instruments for 25OHD identified here should be used with

caution in future MR analyses assessing the association between vitamin D and other complex traits and diseases.

5.6 Declaration of Interests: The authors declare no competing interests.

5.7 Author Contributions: JBR and NJT conceived the experiment, and DM, RM, TD, SH undertook the analyses. All authors revised the manuscript. DM and RM were the main analysts, performed the GWAS in UK Biobank, the meta-analysis with the previous GWAS and *in silico* follow-up, as well as the first draft of the manuscript. CL, RS and JL undertook the 1,25 dihydroxyvitamin D GWAS and the look-up for the cojo-independent 25OHD SNPs. JBR is the guarantor of the manuscript.

5.8 Ethical approval: All data sources used in this study (UK Biobank, Ely Study) received approval from respective national ethical committees for medical research and obtained informed consent from all participants. Additional ethical approval was not required for this study.

5.9 Funding: JBR is supported by the Canadian Institutes of Health Research (CIHR), the Canadian Foundation for Innovation and the Fonds de Recherche du Québec – Santé (FRQS). JBR is funded by a FRQS Clinical Research Scholarship and received research support from the National MS Society and the MS Society of Canada. DM is funded by the JDRF. REM works in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol (grant number MC_UU_00011/1). The funding agencies had no role in the study design, analysis, or

interpretation of data; the writing of the manuscript; or in the decision to submit the article for publication.

5.10 Web resources

OMIM, http://www.omim.org/

The GWAS summary-level results will become available through GRASP

https://grasp.nhlbi.nih.gov/

5.11 References

1. Bouillon, R., Marcocci, C., Carmeliet, G., Bikle, D., White, J.H., Dawson-Hughes, B., Lips, P., Munns, C.F., Lazaretti-Castro, M., Giustina, A., et al. (2018). Skeletal and extra-skeletal actions of vitamin D: Current evidence and outstanding questions. Endocr Rev.

2. Theodoratou, E., Tzoulaki, I., Zgaga, L., and Ioannidis, J.P. (2014). Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ 348, g2035.

3. Autier, P., Mullie, P., Macacu, A., Dragomir, M., Boniol, M., Coppens, K., Pizot, C., and Boniol, M. (2017). Effect of vitamin D supplementation on non-skeletal disorders: a systematic review of meta-analyses and randomised trials. Lancet Diabetes Endocrinol 5, 986-1004.

4. Haroon, M., and Fitzgerald, O. (2012). Vitamin D and its emerging role in immunopathology. Clin Rheumatol 31, 199-202.

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

6. Shea, M.K., Benjamin, E.J., Dupuis, J., Massaro, J.M., Jacques, P.F., D'Agostino, R.B., Sr., Ordovas, J.M., O'Donnell, C.J., Dawson-Hughes, B., Vasan, R.S., et al. (2009). Genetic and non-genetic correlates of vitamins K and D. Eur J Clin Nutr 63, 458-464.

7. Karohl, C., Su, S., Kumari, M., Tangpricha, V., Veledar, E., Vaccarino, V., and Raggi, P. (2010). Heritability and seasonal variability of vitamin D concentrations in male twins. Am J Clin Nutr 92, 1393-1398.

8. Hunter, D., De Lange, M., Snieder, H., MacGregor, A.J., Swaminathan, R., Thakker, R.V., and Spector, T.D. (2001). Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. J Bone Miner Res 16, 371-378.

9. Jiang, X., O'Reilly, P.F., Aschard, H., Hsu, Y.H., Richards, J.B., Dupuis, J., Ingelsson, E., Karasik, D., Pilz, S., Berry, D., et al. (2018). Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. Nat Commun 9, 260.

10. Wang, T.J., Zhang, F., Richards, J.B., Kestenbaum, B., van Meurs, J.B., Berry, D., Kiel, D.P., Streeten, E.A., Ohlsson, C., Koller, D.L., et al. (2010). Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet 376, 180-188.

11. Ahn, J., Yu, K., Stolzenberg-Solomon, R., Simon, K.C., McCullough, M.L., Gallicchio, L., Jacobs, E.J., Ascherio, A., Helzlsouer, K., Jacobs, K.B., et al. (2010). Genome-wide association study of circulating vitamin D levels. Hum Mol Genet 19, 2739-2745.

12. Manousaki, D., Dudding, T., Haworth, S., Hsu, Y.H., Liu, C.T., Medina-Gomez, C., Voortman, T., van der Velde, N., Melhus, H., Robinson-Cohen, C., et al. (2017). Low-

Frequency Synonymous Coding Variation in CYP2R1 Has Large Effects on Vitamin D Levels and Risk of Multiple Sclerosis. Am J Hum Genet 101, 227-238.

13. Mokry, L.E., Ross, S., Ahmad, O.S., Forgetta, V., Smith, G.D., Goltzman, D., Leong, A., Greenwood, C.M., Thanassoulis, G., and Richards, J.B. (2015). Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLoS Med 12, e1001866.

14. Rhead, B., Baarnhielm, M., Gianfrancesco, M., Mok, A., Shao, X., Quach, H., Shen, L., Schaefer, C., Link, J., Gyllenberg, A., et al. (2016). Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. Neurol Genet 2, e97.

15. Giulietti, A., Gysemans, C., Stoffels, K., van Etten, E., Decallonne, B., Overbergh, L., Bouillon, R., and Mathieu, C. (2004). Vitamin D deficiency in early life accelerates Type 1 diabetes in non-obese diabetic mice. Diabetologia 47, 451-462.

16. Riachy, R., Vandewalle, B., Moerman, E., Belaich, S., Lukowiak, B., Gmyr, V., Muharram, G., Kerr Conte, J., and Pattou, F. (2006). 1,25-Dihydroxyvitamin D3 protects human pancreatic islets against cytokine-induced apoptosis via down-regulation of the Fas receptor. Apoptosis 11, 151-159.

17. (1999). Vitamin D supplement in early childhood and risk for Type I (insulin-dependent) diabetes mellitus. The EURODIAB Substudy 2 Study Group. Diabetologia 42, 51-54.

18. Hypponen, E., Laara, E., Reunanen, A., Jarvelin, M.R., and Virtanen, S.M. (2001). Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. Lancet 358, 1500-1503.

19. Mayer-Davis, E.J., Dabelea, D., Crandell, J.L., Crume, T., D'Agostino, R.B., Jr., Dolan, L., King, I.B., Lawrence, J.M., Norris, J.M., Pihoker, C., et al. (2013). Nutritional factors and preservation of C-peptide in youth with recently diagnosed type 1 diabetes: SEARCH Nutrition Ancillary Study. Diabetes Care 36, 1842-1850.

20. Littorin, B., Blom, P., Scholin, A., Arnqvist, H.J., Blohme, G., Bolinder, J., Ekbom-Schnell, A., Eriksson, J.W., Gudbjornsdottir, S., Nystrom, L., et al. (2006). Lower levels of plasma 25-hydroxyvitamin D among young adults at diagnosis of autoimmune type 1 diabetes compared with control subjects: results from the nationwide Diabetes Incidence Study in Sweden (DISS). Diabetologia 49, 2847-2852.

21. Baumgartl, H.J., Standl, E., Schmidt-Gayk, H., Kolb, H.J., Janka, H.U., and Ziegler, A.G. (1991). Changes of vitamin D3 serum concentrations at the onset of immune-mediated type 1 (insulin-dependent) diabetes mellitus. Diabetes Res 16, 145-148.

22. Bierschenk, L., Alexander, J., Wasserfall, C., Haller, M., Schatz, D., and Atkinson, M. (2009). Vitamin D levels in subjects with and without type 1 diabetes residing in a solar rich environment. Diabetes Care 32, 1977-1979.

23. Pozzilli, P., Manfrini, S., Crino, A., Picardi, A., Leomanni, C., Cherubini, V., Valente, L., Khazrai, M., Visalli, N., and group, I. (2005). Low levels of 25-hydroxyvitamin D3 and 1,25-dihydroxyvitamin D3 in patients with newly diagnosed type 1 diabetes. Horm Metab Res 37, 680-683.

24. Feldman, D., Krishnan, A.V., Swami, S., Giovannucci, E., and Feldman, B.J. (2014). The role of vitamin D in reducing cancer risk and progression. Nat Rev Cancer 14, 342-357.

25. Smith, G.D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32, 1-22.

26. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 37, 658-665.

27. Gianfrancesco, M.A., Stridh, P., Rhead, B., Shao, X., Xu, E., Graves, J.S., Chitnis, T., Waldman, A., Lotze, T., Schreiner, T., et al. (2017). Evidence for a causal relationship between low vitamin D, high BMI, and pediatric-onset MS. Neurology 88, 1623-1629.

28. Canada., M.S.S.o. (2019). Vitamin D and Multiple Sclerosis Recommendations 2018. In. (

29. Larsson, S.C., Traylor, M., Markus, H.S., and Michaelsson, K. (2018). Serum Parathyroid Hormone, 25-Hydroxyvitamin D, and Risk of Alzheimer's Disease: A Mendelian Randomization Study. Nutrients 10.

30. He, Y., Timofeeva, M., Farrington, S.M., Vaughan-Shaw, P., Svinti, V., Walker, M., Zgaga, L., Meng, X., Li, X., Spiliopoulou, A., et al. (2018). Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: a large Mendelian randomisation study. BMC Med 16, 142.

31. Aspelund, T., Grubler, M.R., Smith, A.V., Gudmundsson, E.F., Keppel, M., Cotch, M.F., Harris, T.B., Jorde, R., Grimnes, G., Joakimsen, R., et al. (2019). Effect of Genetically

Low 25-Hydroxyvitamin D on Mortality Risk: Mendelian Randomization Analysis in 3 Large European Cohorts. Nutrients 11.

32. Michaelsson, K., Melhus, H., and Larsson, S.C. (2018). Serum 25-Hydroxyvitamin D Concentrations and Major Depression: A Mendelian Randomization Study. Nutrients 10.

33. Bowman, K., Jones, L., Pilling, L.C., Delgado, J., Kuchel, G.A., Ferrucci, L., Fortinsky, R.H., and Melzer, D. (2019). Vitamin D levels and risk of delirium: A mendelian randomization study in the UK Biobank. Neurology 92, e1387-e1394.

34. Lund-Nielsen, J., Vedel-Krogh, S., Kobylecki, C.J., Brynskov, J., Afzal, S., and Nordestgaard, B.G. (2018). Vitamin D and Inflammatory Bowel Disease: Mendelian Randomization Analyses in the Copenhagen Studies and UK Biobank. J Clin Endocrinol Metab 103, 3267-3277.

35. Sun, J.Y., Zhao, M., Hou, Y., Zhang, C., Oh, J., Sun, Z., and Sun, B.L. (2019). Circulating serum vitamin D levels and total body bone mineral density: A Mendelian randomization study. J Cell Mol Med 23, 2268-2271.

36. Jiang, X., Dimou, N.L., Al-Dabhani, K., Lewis, S.J., Martin, R.M., Haycock, P.C., Gunter, M.J., Key, T.J., Eeles, R.A., Muir, K., et al. (2018). Circulating vitamin D concentrations and risk of breast and prostate cancer: a Mendelian randomization study. Int J Epidemiol.

37. Larsson, S.C., Traylor, M., Mishra, A., Howson, J.M.M., Michaelsson, K., Markus, H.S., and Consortium, M.P.o.t.I.S.G. (2018). Serum 25-Hydroxyvitamin D Concentrations and Ischemic Stroke and Its Subtypes. Stroke 49, 2508-2511.

38. Yarmolinsky J, R.C., Lophatananon A, et al. . (November 2018:472696.). Evaluating causal associations between previously reported risk factors and epithelial ovarian cancer: a Mendelian randomization analysis. . BioRxiv.

39. Mai, X.M., Videm, V., Sheehan, N.A., Chen, Y., Langhammer, A., and Sun, Y.Q. (2019). Potential causal associations of serum 25-hydroxyvitamin D with lipids: a Mendelian randomization approach of the HUNT study. Eur J Epidemiol 34, 57-66.

40. Dong, J., Gharahkhani, P., Chow, W.H., Gammon, M.D., Liu, G., Caldas, C., Wu, A.H., Ye, W., Onstad, L., Anderson, L.A., et al. (2019). No Association Between Vitamin D Status and Risk of Barrett's Esophagus or Esophageal Adenocarcinoma: A Mendelian Randomization Study. Clin Gastroenterol Hepatol.

41. Tan, V.Y., Biernacka, K.M., Dudding, T., Bonilla, C., Gilbert, R., Kaplan, R.C., Qibin, Q., Teumer, A., Martin, R.M., Perks, C.M., et al. (2018). Reassessing the Association between Circulating Vitamin D and IGFBP-3: Observational and Mendelian Randomization Estimates from Independent Sources. Cancer Epidemiol Biomarkers Prev 27, 1462-1471.

42. Havdahl, A., Mitchell, R., Paternoster, L., and Davey Smith, G. (2019). Investigating causality in the association between vitamin D status and self-reported tiredness. Sci Rep 9, 2880.

43. Yuri Milaneschi, W.J.P., Michel G Nivard, Hamdi Mbarek, Dorret I Boomsma, Brenda WJH Penninx. (2019). A role for vitamin D and omega-3 fatty acids in major depression? An exploration using genomics. BioRxiv.

44. Libuda, L., Laabs, B.H., Ludwig, C., Buhlmeier, J., Antel, J., Hinney, A., Naaresh, R., Focker, M., Hebebrand, J., Konig, I.R., et al. (2019). Vitamin D and the Risk of Depression: A Causal Relationship? Findings from a Mendelian Randomization Study. Nutrients 11.

45. Liyanage, U.E., Law, M.H., Melanoma Meta-analysis, C., Barrett, J.H., Iles, M.M., and MacGregor, S. (2019). Is there a causal relationship between vitamin D and melanoma risk? : A Mendelian randomization study. Br J Dermatol.

46. Xiangrui Meng, X.L., Maria Timofeeva, Yazhou He, Athina, and Spiliopoulou, W.-Q.W., Aliya Gifford, Hongjiang Wu, Timothy Varley, Peter Joshi,Joshua C. Denny, Susan Farrington, Lina Zgaga, Malcolm G. Dunlop, Paul McKeigue, Harry Campbell, Evropi Theodoratou. (2019). Phenome Wide Mendelian Randomization Study of genetically determined Vitamin D on multiple health outcomes using the UK Biobank Study. Int J Epidemiol In press.

47. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.

48. McKenna, M.J., and Murray, B.F. (2013). Vitamin D dose response is underestimated by Endocrine Society's Clinical Practice Guideline. Endocr Connect 2, 87-95.

49. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., Y, A.R., H, K.F., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 48, 1443-1448.

50. Persyn, E., Redon, R., Bellanger, L., and Dina, C. (2018). The impact of a fine-scale population stratification on rare variant association test results. PLoS One 13, e0207677.

51. Galinsky, K.J., Bhatia, G., Loh, P.R., Georgiev, S., Mukherjee, S., Patterson, N.J., and Price, A.L. (2016). Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. Am J Hum Genet 98, 456-472.

52. Morris, J.A., Kemp, J.P., Youlten, S.E., Laurent, L., Logan, J.G., Chai, R.C., Vulpescu, N.A., Forgetta, V., Kleinman, A., Mohanty, S.T., et al. (2019). An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet 51, 258-266.

53. Loh, P.R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., Schizophrenia Working Group of Psychiatric Genomics, C., de Candia, T.R., Lee, S.H., Wray, N.R., et al. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat Genet 47, 1385-1392.

54. Magi, R., and Morris, A.P. (2010). GWAMA: software for genome-wide association meta-analysis. Bmc Bioinformatics 11.

55. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190-2191.

56. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88, 76-82.

57. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol 17, 122.

58. Kamat, M.A., Blackshaw, J.A., Young, R., Surendran, P., Burgess, S., Danesh, J., Butterworth, A.S., and Staley, J.R. (2019). PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics.

59. Park, J.H., Wacholder, S., Gail, M.H., Peters, U., Jacobs, K.B., Chanock, S.J., and Chatterjee, N. (2010). Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 42, 570-575.

60. Cuellar-Partida G, L.M., Kho PF, D'Urso S, Gutierrez-Mondragon LF, Hwang L-D. (2019). Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses. bioRxiv.

61. Barbeira, A.N., Pividori, M., Zheng, J., Wheeler, H.E., Nicolae, D.L., and Im, H.K. (2019). Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet 15, e1007889.

62. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun 6, 5890.

63. Ellinghaus, D., Jostins, L., Spain, S.L., Cortes, A., Bethune, J., Han, B., Park, Y.R., Raychaudhuri, S., Pouget, J.G., Hubenthal, M., et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 48, 510-518.

64. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen, C., Psychiatric Genomics, C., Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control, C., Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236-1241.

65. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.R., Lareau, C., Shoresh, N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet 50, 621-629.

66. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47, 1228-1235.

67. Forouhi, N.G., Luan, J., Hennings, S., and Wareham, N.J. (2007). Incidence of Type 2 diabetes in England and its association with baseline impaired fasting glucose: the Ely study 1990-2000. Diabet Med 24, 200-207.

68. Williams, D.R., Wareham, N.J., Brown, D.C., Byrne, C.D., Clark, P.M., Cox, B.D., Cox, L.J., Day, N.E., Hales, C.N., Palmer, C.R., et al. (1995). Undiagnosed glucose intolerance in the community: the Isle of Ely Diabetes Project. Diabet Med 12, 30-35.

69. Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. Nat Rev Genet 11, 499-511.

70. Zittermann, A., Schleithoff, S.S., Frisch, S., Gotting, C., Kuhn, J., Koertke, H., Kleesiek, K., Tenderich, G., and Koerfer, R. (2009). Circulating calcitriol concentrations and total mortality. Clin Chem 55, 1163-1170.

5.12 Figures

Figure 1. The vitamin D metabolic pathway

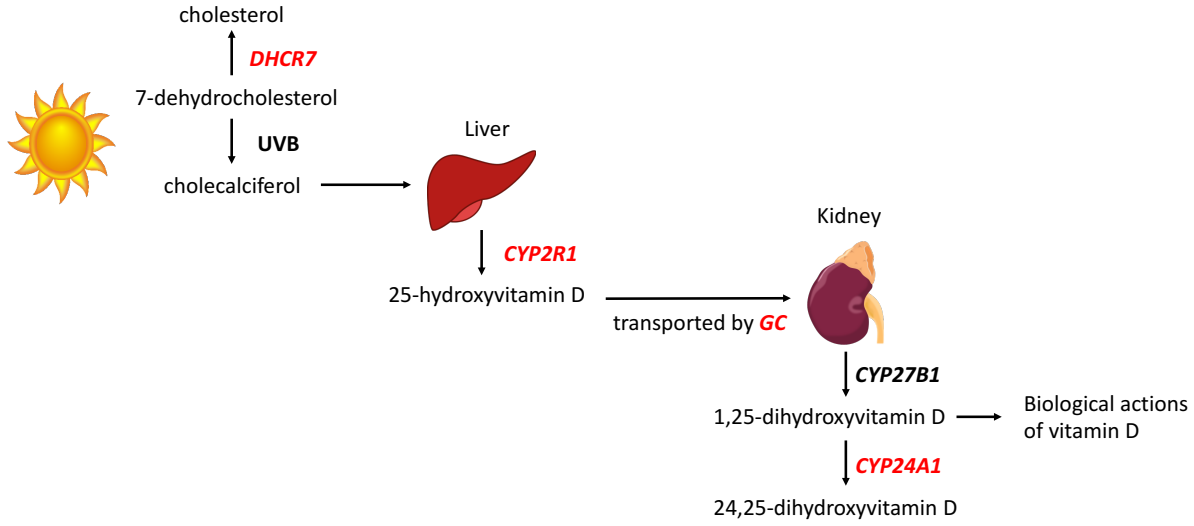


Figure 2. Genome-wide association of 25OHD graphed by chromosome positions and −log10 P-value (Manhattan plot), and quantile-quantile plot of all 20,370,874 SNPs from the GWAS meta-analysis (QQ-plot) on 443,374 European individuals. a Manhattan plot: The P-values were obtained from the fixed-effects inverse variance weighted meta-analysis. The Y axis shows −log10 P-values, and the X axis shows chromosome positions. Horizontal red dash line represents the thresholds of $P = 6.6\times10^{-9}$ for genome-wide significance. Known loci were colored coded as blue diamonds, novel rare loci were color coded as red diamonds, and novel common loci were color coded as white diamonds. b QQ-plot: The Y axis shows observed −log10 P-values (it is truncated at 310), and the X axis shows the expected −log10 P-values. Each SNP is plotted as a blue dot, and the dash red line indicates null hypothesis of no true association. Deviation from the expected P-value distribution is evident only in the tail area, with a lambda of 1.23.

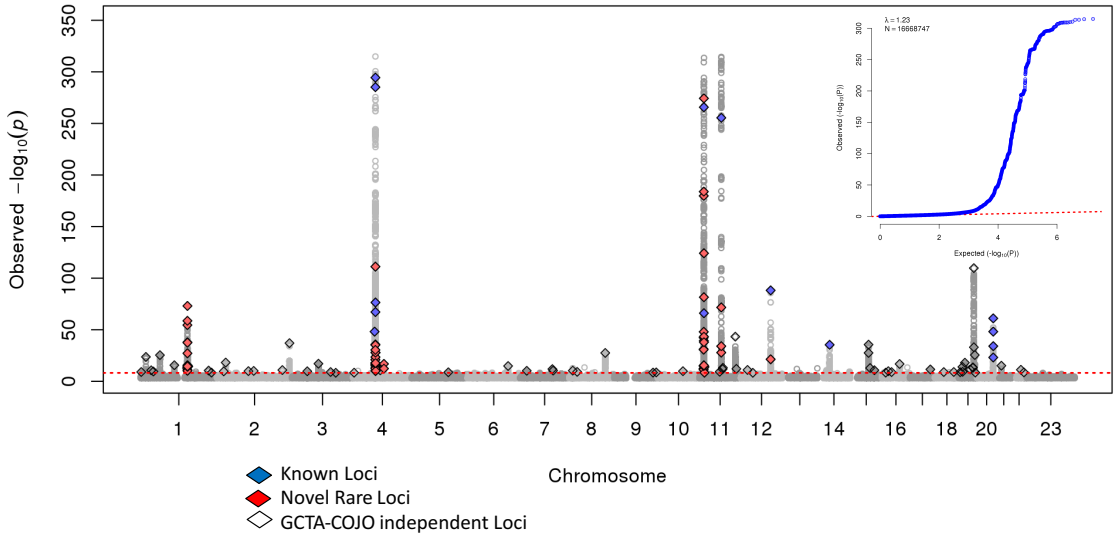


Figure 3. Effect of predicted increased transcription of all genes on circulating vitamin D.

Each dot represents the effect of increased transcription (averaged across all tissue-specific predictions using S-MultiXcan) on 25OHD.

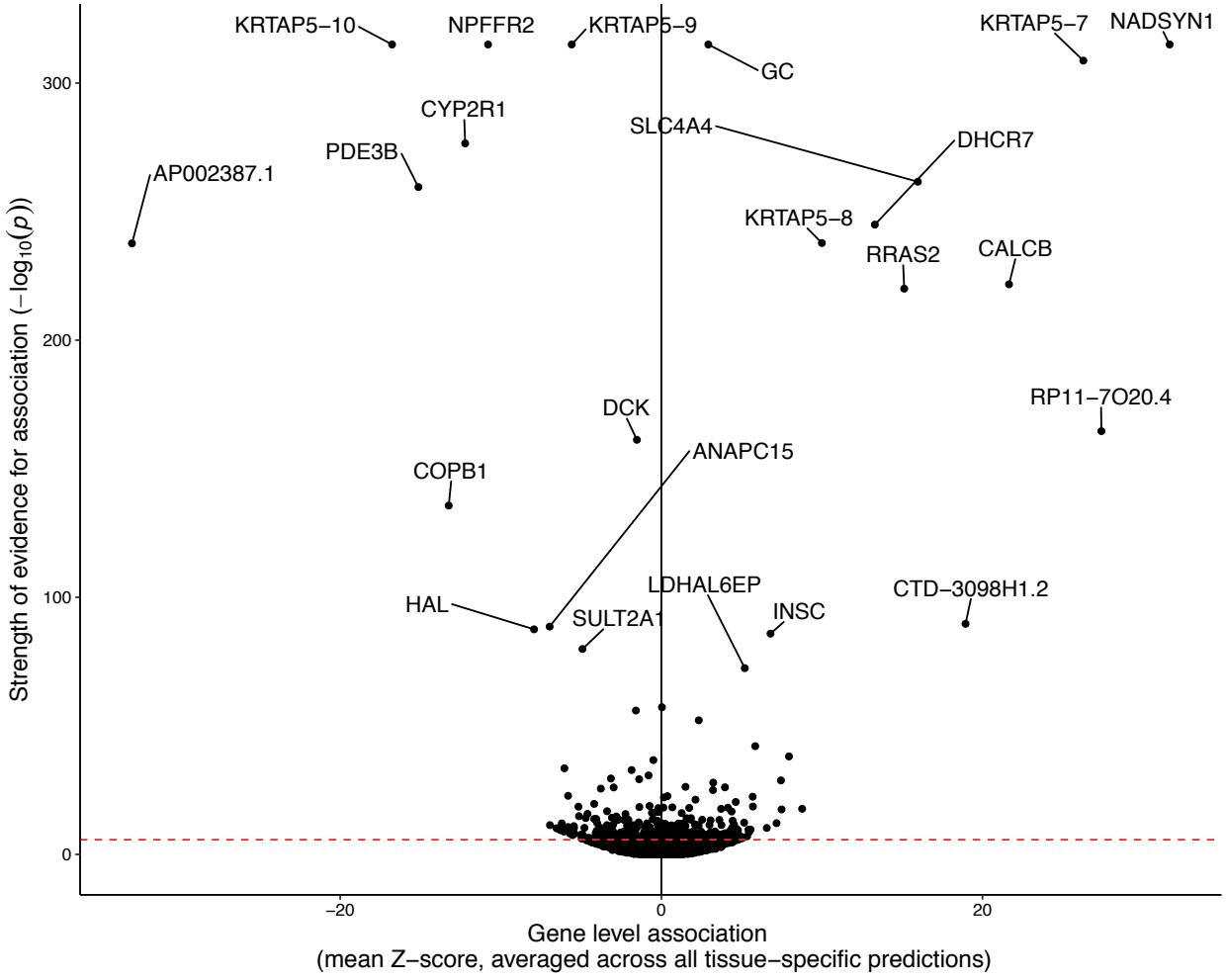
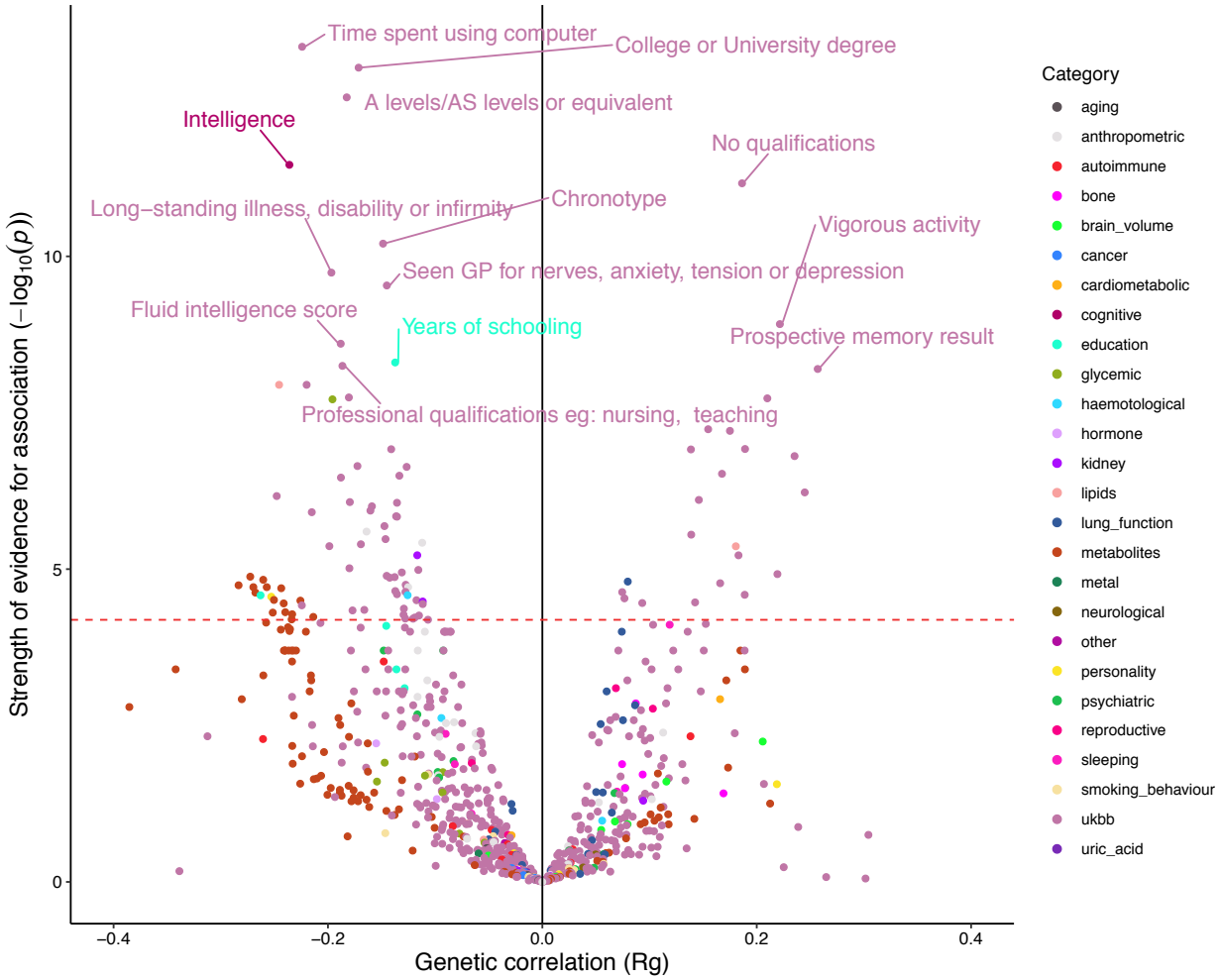


Figure 4. Genetic correlation between 25OHD levels and GWAS traits available within LD hub. Each dot represented the Rg between 25OHD and an individual trait. The red dashed line represents the Bonferroni-corrected multiple testing threshold at the 5% level.



5.13 Supplemental Material and Methods

5.13.1 Supplemental Text and Figures

UK Biobank data fields

Information on vitamin D supplementation in UK Biobank participants was retrieved from data fields 20084 (Vitamin and/or mineral supplement use), using the code 479 (Vitamin D) and from data field 6155 (Vitamin and mineral supplements), using the code 4 (Vitamin D).

The specific description for data field 20084 is: Category: Vitamin / mineral supplements yesterday -Diet by 24-hour recall –online follow-up

For data field 6155, the description is as follows: Category: Medication-Health and medical history-Touchscreen

Information on 25OHD levels was retrieved from data field 30890.

25OHD measurements

Blood was drawn from consenting participants and transported to the UK Biobank central laboratory in Stockport, UK. The whole blood was aliquoted and frozen within an average time of 24 hours (+/- 2.4 hours). 25(OH)D was measured by immunoassay on a DiaSorin LIASON® XL chemiluminescence analyzer with a detection range of 10 nmol/L < 25(OH)D < 375 nmol/L. Where samples were above 375 nmol/L and sufficient sample remained, a dilution was

performed and the sample reanalysed. During the automated aliquoting process, an unintended dilution of the aliquots occurred however, approximately 92% of sample assay concentrations are affected by less than 1% dilution, approximately 8% of assay concentrations are affected by less than 10% dilution and very few are affected by more than this. Correction for this unintentional dilution has been made centrally by UK Biobank and more details can be found here: http://biobank.ndph.ox.ac.uk/showcase/docs/biomarker_issues.pdf.

Genotype QC and Imputation in UK Biobank

In this study, we used the Version 3 of the imputed genotype data. Description of category can be found through the following link: https://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=263)

Specifically, genome-wide genetic data is available for 488,000 UK Biobank participants.

Genotype calling was performed by Affymetrix (now part of ThermoFisher Scientific) on two closely related purpose-designed arrays. ~50,000 participants were run on the UK BiLEVE Axiom array (Resource 149600) and the remaining ~450,000 were run on the UK Biobank Axiom array (Resource 149601). The dataset combines results from both arrays (data field 22000) and there are 805,426 markers in the released genotype data. The positions of markers in the data are in GRCh37 coordinates. It was not possible to assay genotypes for some participants (~3%) as sufficient DNA could not be extracted from their blood samples.

The genotype data were quality controlled (QC). In addition, the dataset was phased and ~96M genotypes were imputed using computationally efficient methods combined with the Haplotype Reference Consortium and UK10K haplotype resources. Classical allelic variations at eleven HLA genes were imputed. Information from the QC pipeline, such as array, and important genetic properties of the data such as population structure and relatedness are available.

Details of these analyses, and the methods used to derive other data such as imputation and haplotypes, is given in the paper by Bycroft et al[1].

Ancestry assignment

A sample of 409,728 White British individuals was identified centrally by the UK Biobank, using a combination of self-reported ethnicity and genetic information. However, the reliance on self-reported information was deemed too conservative and we chose to redefine a White British sample (n=440,414) using genetic information only. The approach is described in detail in a previous publication by our group[2]. Specifically, we projected the UK Biobank sample onto the first 20 principal components estimated from the 1000 Genomes Phase 3 (1000G) project data[3] (where ancestry was known) using FastPCA version 2[4]. Projections used a curated set of 38,551 LD-pruned HapMap 3 Release 3 (HM3)[5] bi-allelic SNPs that were shared between the 1000G and UK Biobank datasets (MAF>1%, minor allele count >5, genotyping call rate >95%, Hardy-Weinberg p>1x10-6 and regions of extensive LD removed). Expectation

Maximization (EM) clustering (implemented in R using EMCluster[6]) was used to compute probabilities of cluster membership based on a finite mixture of multivariate Gaussian distributions with unstructured dispersion. Eigenvectors 1, 2 and 5 were used for clustering as they represented the smallest number of eigenvectors that were able to resolve the British 1000G subpopulation (GBR) from other ethnicities. Twelve predefined clusters were chosen for EM clustering as sensitivity analyses suggested that this number provided a good compromise between model fit (as quantified by log likelihood, Bayesian information criterion, and Akaike information criterion) and computational burden. UK Biobank participants (n=440,414) that clustered together with the 1000G GBR subpopulation were termed White British and used for downstream genetic analyses.

GWAS

Meta-analysis

Summary level files from the two GWAS were filtered using the following SNP-level exclusion criteria: i) Info score <0.3, ii) HWE P-value $<10^{-6}$ iii) MAF <0.1% and iv) being a bi-allelic variant (indels were excluded). Alignment of the SNPs across studies was done using the chromosome and position information for each variant according to genome build hg19. Fixed – effects beta-based meta-analysis was performed using the software package GWAMA32[7] without adjusting for genomic control. Meta-analysis results were combined across the two

studies for consistency using inverse variance weighted fixed effects. QQ and Manhattan plots for single variant associations with 25OHD are shown in **Figure 2a and 2b**, respectively.

Approximate conditional association analysis

Conditional analyses for significant SNPs was performed using GCTAversion1.91.1[8; 9].This method uses an approximate conditional analysis approach from summary-level statistics from the meta-analysis and LD corrections between SNPs estimated from a reference sample. 50,000 unrelated white British individuals randomly selected from the UK Biobank was created for a previous GWAS[2], and was used to model patterns of LD between variants. We used the following parameters: --cojo-wind 20000 --cojo-slct --cojo-collinear 0.9 --cojo-p 6.6e-9 in order to extract associated regions flanking within 20kb of the top SNPs to then conduct conditional analyses within these regions.

1,25-dihydroxyvitamin D GWAS

Study participants

The Ely Study, established in 1990, is a prospective study of the aetiology of Type 2 diabetes and has been described in detail elsewhere [10; 11]. Briefly, the study comprises of three phases: Phase 1 (Baseline examination: 1990-1992), Phase 2 (1994-1996) and Phase 3 (2000-2003). At baseline, European ancestry individuals aged 40-69 years registered at a single medical practice in Ely, Cambridgeshire, UK were invited to participate. Baseline blood samples were collected after an

overnight fast. Genotyping was performed for participants attending examination at Phase 3. All participants of the Ely Study gave their written informed consent and the study was approved by the local ethics committee.

Sample genotyping, genotype quality control and imputation

Samples were genotyped using the HumanCoreExome-24 and InfiniumCoreExome arrays. Samples were excluded for: call rate <98% (11 samples); common variant heterozygosity outliers (MAF >1%, heterozygosity <0.341 or >0.363 and not ethnic outlier; 7 samples), rare variant heterozygosity outlier (MAF <1%, heterozygosity >0.005 and not ethnicity outlier; 19 samples); rare allele count outlier and not ethnic outlier (6 samples); duplicate samples, lower call rate (5 samples) and sex mismatch (6 samples). Monoalleleic variants and variants with a call rate <95% were excluded (705 variants). A total of 1,591 samples and 546,486 variants met the above quality control criteria. Imputation was performed using the Sanger Imputation Server (pre-phase with EAGLE2 and impute with PBWT pipeline), and the HRC 1.1 reference panel. Additional variants not captured by the HRC reference panel were imputed using a combined UK10K and 1000 Genomes Phase 3 reference panel resulting in data available for >14 million variants.

***In-silico* functional analysis**

Functional analysis of the meta-analysis summary statistics was performed using Complex Trait Genomic-Virtual Lab[12] web application (https://genoma.io) which implements a variety of follow-up methods for GWAS summary statistics.

Association between predicted gene transcription and 25OHD was estimated using S-MultiXcan[13] in the MetaXcan package with the default options implemented. This analysis uses genetically predicted gene expression variation across multiple tissues as exposure and 25OHD as outcome. Elastic net prediction models for 48 tissues were trained using GTEx (version 7)[14] data and obtained from the PredictDB repository (http://predictdb.org). Association statistics for the 48 tissues were combined accounting for correlation between tissues to give transcript-level results, and a Bonferroni correction was applied to account for the number of gene transcripts tested.

Gene prioritisation, gene set and tissue enrichment analysis were performed using DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits)[15; 16] to identify likely causal genes at associated loci, highlight gene pathways which are over-represented by associated loci in the single variant results and test whether expression of these genes is enriched in specific tissue types. First, DEPICT defines associated loci by clumping genome-wide single variant results. The default criteria ($p<5x10^{-8}$, LDwindow< 1Mb, LD $r^2>0.1$) were used for this step. Next, DEPICT looks for co-expressed gene pairs present within these associated loci, using precomputed co-regulated gene-networks defined using publicly available gene expression data. DEPICT anticipates that

association signals for a biologically causal gene will co-localise with genes encoding other members of a co-regulated gene network elsewhere in the genome. Finally, DEPICT tests for enrichment in tissue/cell types by assessing whether genes in associated loci are highly expressed in any of the 209 Medical Subject Heading (MESH) tissue and cell type annotations from human gene expression microarrays. DEPICT constructs a tissue/cell type expression matrix by averaging gene expression levels of microarray samples with the same MeSH annotation and uses this to assess enrichment in an identical method to the gene set enrichment. To take sources of confounding into account, DEPICT calibrates results against a null expectation from precomputed GWAS based randomly distributed dummy phenotypes.

Genetic correlation between vitamin D and a range of other traits available as publicly available GWAS summary statistics was examined using bivariate LD score regression[17] implemented in the LD Hub platform[18].

Partitioned heritability by functional annotation with 53 overlapping categories was performed using stratified linkage disequilibrium (LD) score regression (LDSR) using the baseline model from 1000 Genomes phase 3 data (baselineLD_v2.2, February 2019)[17; 19]. Cell specific heritability was examined using the --h2-cts flag in LDSR and the multi-tissue gene expression file ("Multi_tissue_gene_expr" containing both GTEx data and Franke lab dataset of microarray gene

expression)[20]. Both of these analyses are restricted to common variants in present in HapMap3 (approximately 1,500,000 SNPs) excluding those within the HLA.

Figure S1. Density plot of the distribution of baseline standardized log-transformed 25OHD measurements from 401,460 White-British UK Biobank participants. The y-axis is the probability density function for the kernel density estimation, whereas the x-axis depicts values of the standardized log-transformed 25OHD.

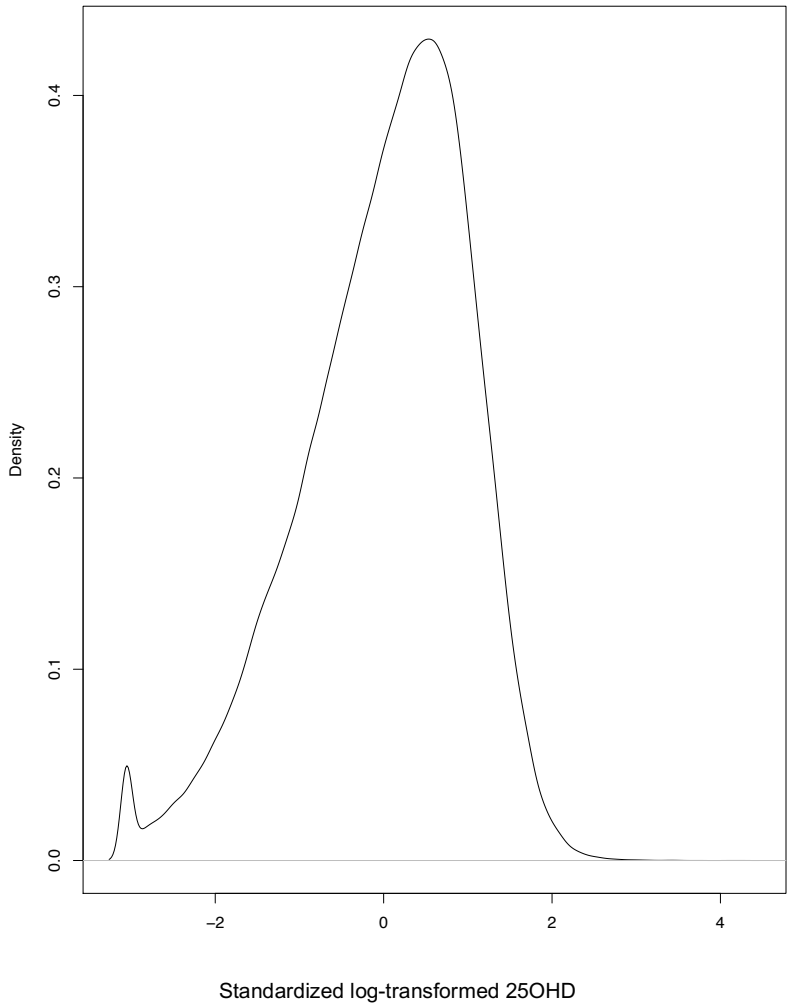
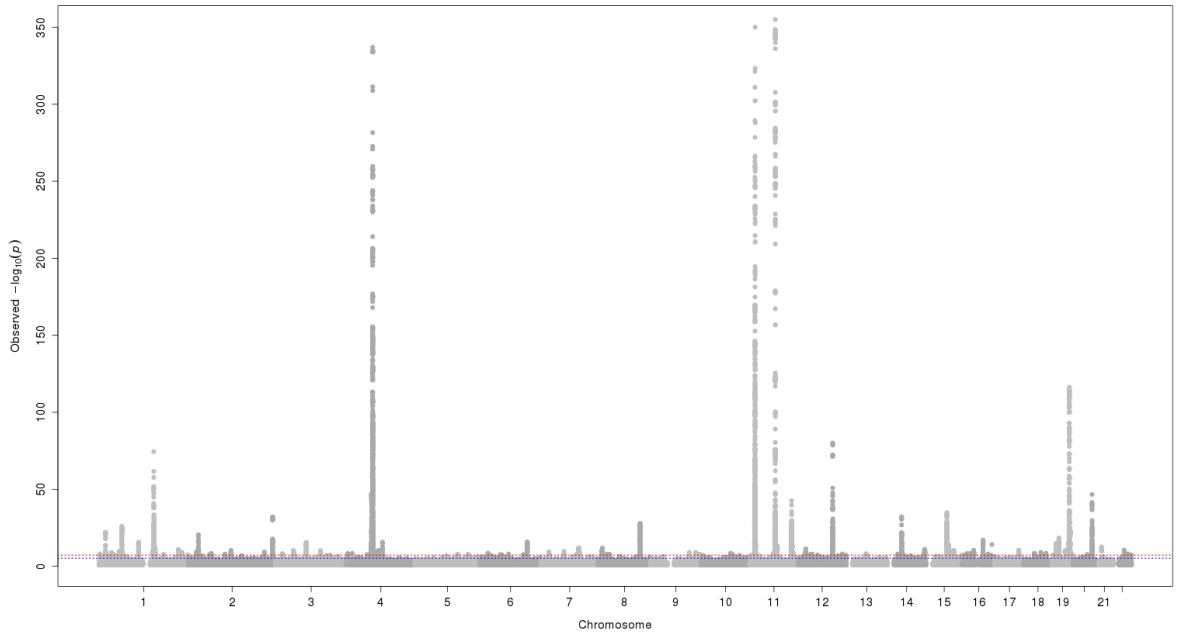


Figure S2. Genome-wide association of 25OHD graphed by chromosome positions and −log10 P-value (Manhattan plot), and quantile-quantile plot of 20,370,875 SNPs from the UK Biobank GWAS (QQ-plot) in 401,460 White-British individuals. a Manhattan plot: The P-values were obtained from the BOLT-LMM. The Y axis shows −log10 P-values, and the X axis shows chromosome positions. Horizontal red and blue dash line represents the thresholds of $P = 6.6x10^{-9}$ and $5x10^{-6}$ for genome-wide significant and genome-wide suggestive P-values. b QQ-plot: The Y axis shows observed −log10 P-values (and is truncated at 350), and the X axis shows the expected −log10 P-values. Each SNP is plotted as a blue dot, and the dash red line indicates null hypothesis of no true association. Deviation from the expected P-value distribution is evident only in the tail area, with a lambda of 1.2, suggesting that population stratification was adequately controlled.

a


Observed $-\log_{10}(p)$
0
50
100
150
200
250
300
350
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
21
Chromosome

b

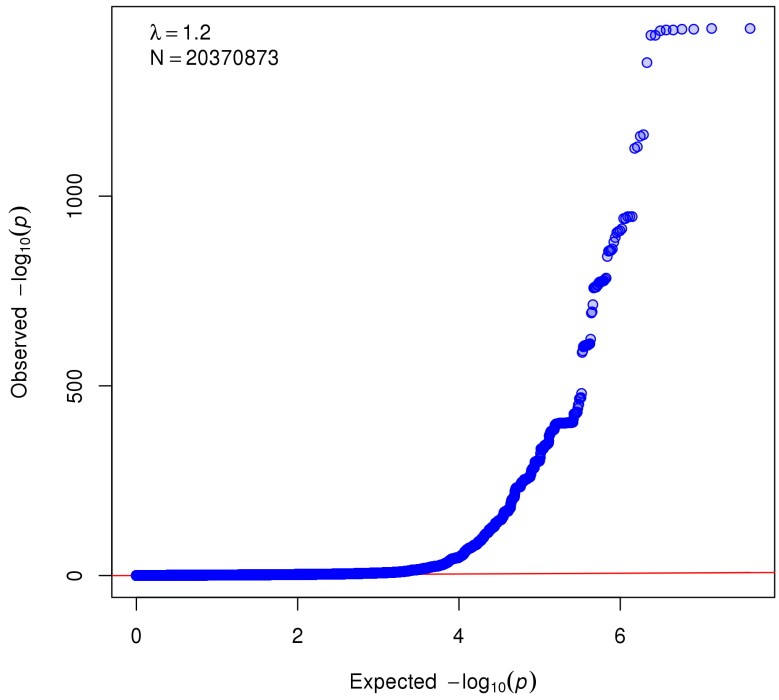
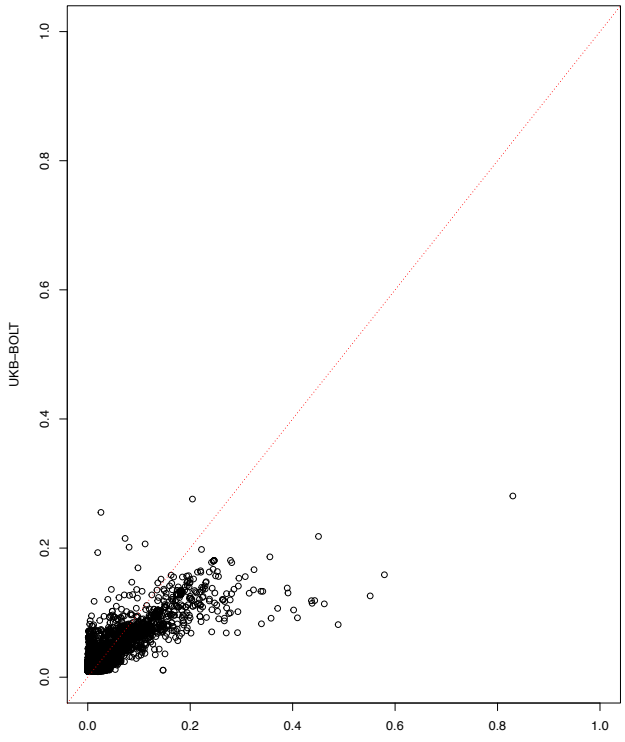

λ = 1.2
N = 20370873
Observed $-\log_{10}(p)$
0
500
1000
0
2
4
6
Expected $-\log_{10}(p)$

Figure S3. Correlation between the betas of the UK Biobank GWAS and of a previous 25OHD GWAS on 42,274 Europeans *(Manousaki et al, AJHG 2017)*[21]. 20,787 SNPs achieving p-values < 1x 10^{-6} in both GWAS are included in this analysis The y-axis depicts the betas from the BOLT-LMM in UK Biobank, whereas the x-axis depicts the betas from the aforementioned GWAS meta-analysis.



Manousaki et al, AJHG 2017

Supplemental References

1. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203-209.

2. Morris, J.A., Kemp, J.P., Youlten, S.E., Laurent, L., Logan, J.G., Chai, R.C., Vulpescu, N.A., Forgetta, V., Kleinman, A., Mohanty, S.T., et al. (2019). An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet 51, 258-266.

3. Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56-65.

4. Galinsky, K.J., Bhatia, G., Loh, P.R., Georgiev, S., Mukherjee, S., Patterson, N.J., and Price, A.L. (2016). Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. Am J Hum Genet 98, 456-472.

5. International HapMap, C., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., et al. (2010). Integrating common and rare genetic variation in diverse human populations. Nature 467, 52-58.

6. Chen, W.-C.M., R. . (2015). EM Algorithm for Model-Based Clustering of Finite Mixture Gaussian Distribution.

7. Magi, R., and Morris, A.P. (2010). GWAMA: software for genome-wide association meta-analysis. Bmc Bioinformatics 11.

8. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190-2191.

9. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88, 76-82.

10. Forouhi, N.G., Luan, J., Hennings, S., and Wareham, N.J. (2007). Incidence of Type 2 diabetes in England and its association with baseline impaired fasting glucose: the Ely study 1990-2000. Diabet Med 24, 200-207.

11. Williams, D.R., Wareham, N.J., Brown, D.C., Byrne, C.D., Clark, P.M., Cox, B.D., Cox, L.J., Day, N.E., Hales, C.N., Palmer, C.R., et al. (1995). Undiagnosed glucose intolerance in the community: the Isle of Ely Diabetes Project. Diabetic medicine : a journal of the British Diabetic Association 12, 30-35.

12. Cuellar-Partida G, L.M., Kho PF, D'Urso S, Gutierrez-Mondragon LF, Hwang L-D. (2019). Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses. bioRxiv.

13. Barbeira, A.N., Pividori, M., Zheng, J., Wheeler, H.E., Nicolae, D.L., and Im, H.K. (2019). Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet 15, e1007889.

14. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun 6, 5890.

15. Munger, K.L., Levin, L.I., Hollis, B.W., Howard, N.S., and Ascherio, A. (2006). Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. JAMA 296, 2832-2838.

16. Ellinghaus, D., Jostins, L., Spain, S.L., Cortes, A., Bethune, J., Han, B., Park, Y.R., Raychaudhuri, S., Pouget, J.G., Hubenthal, M., et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 48, 510-518.

17. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen, C., Psychiatric Genomics, C., Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control, C., Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236-1241.

18. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.R., Lareau, C., Shoresh, N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet 50, 621-629.

19. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Early, G., et al. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential

of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33, 272-279.

20. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47, 1228-1235.

21. Manousaki, D., Dudding, T., Haworth, S., Hsu, Y.H., Liu, C.T., Medina-Gomez, C., Voortman, T., van der Velde, N., Melhus, H., Robinson-Cohen, C., et al. (2017). Low-Frequency Synonymous Coding Variation in CYP2R1 Has Large Effects on Vitamin D Levels and Risk of Multiple Sclerosis. Am J Hum Genet 101, 227-238.

5.13.2 Supplemental Tables

Note: Due to size restrictions, Tables S3 to S9 are not included in the present e-Thesis file. They can be found in a separate excel file submitted with the present e-Thesis.

Table S1. UK Biobank Study descriptives.

		UK BIOBANK COHORT					
TRAIT	**UNIT**	**MIN**	**MAX**	**RANGE**	**MEDIAN**	**MEAN**	**SD**
AGE	years	39.0	73.0	34.0	58.0	56.8	8.0
MALE_SEX	%	45.9					
25OHD	nmol/L	10.0	130.0	120.0	69.9	70.0	34.7
VITD SUPPLEMEI	%	6.2					
SEASON_FALL	%	22.26					
SEASON_WINTEI	%	25.28					
SEASON_SPRINC	%	28.92					
SEASON_SUMME	%	23.43					

MIN	Minimum recorded value of trait
MAX	Maximum recorded value of trait
RANGE	Difference between maximum trait and minimum trait value
MEDIAN	Median value of trait
MEAN	Mean value of trait
SD	Standard deviation

Table S2. Association results for 138 conditionally independent SNPs that reach genome-wide significance in the UKBB GWAS, the GWAS meta-analysis for 25OHD, and the 1,25 dihydroxyvitamin D GWAS.

CONDITIONALLY INDEPENDENT 25OHD ASSOCIATED VARIANTS					BOLT-LMM UK BIOBANLK (N=401,460)							META-ANALYSIS (N=443,734)		
RSID	CHR	BP	EA	NEA	EAF	MAF	INFO	β	SE	P	P.I	EAF	MAF	β
rs6698680	1	2329661	G	A	0.465	0.465	0.995	-0.011	0.002	8.80E-09	1.90E-08	0.464	0.464	-0.012
rs3750296	1	17559656	C	G	0.342	0.342	0.998	-0.021	0.002	1.00E-22	3.60E-22	0.341	0.341	-0.021
rs7519574	1	34726552	A	G	0.183	0.183	0.989	0.016	0.003	2.20E-09	9.40E-10	0.182	0.182	0.017
rs56044892	1	41830086	T	C	0.211	0.211	0.948	0.013	0.003	5.80E-08	2.20E-07	0.211	0.211	0.015
rs2934744	1	63048045	A	C	0.643	0.357	0.998	-0.022	0.002	1.20E-25	4.50E-26	0.644	0.356	-0.022
rs7528419	1	109817192	G	A	0.225	0.225	1.000	0.019	0.002	2.00E-16	2.10E-15	0.225	0.225	0.019
rs3768013	1	150815411	A	G	0.370	0.370	0.996	-0.014	0.002	4.60E-11	1.00E-10	0.370	0.370	-0.015
rs115045402	1	152029548	A	G	0.026	0.026	0.793	0.105	0.007	2.30E-51	6.50E-50	0.026	0.026	0.107
rs12123821	1	152179152	T	C	0.048	0.048	1.000	0.075	0.005	2.00E-58	1.50E-56	0.048	0.048	0.074
rs201561609	1	152187902	T	C	0.987	0.013	0.569	-0.129	0.012	8.00E-29	1.20E-27	0.987	0.013	-0.129
rs185433896	1	152249021	A	C	0.993	0.007	0.422	-0.246	0.019	1.20E-40	2.00E-38	0.993	0.007	-0.246
rs189918701	1	152254152	G	A	0.997	0.003	0.458	-0.238	0.029	1.20E-16	2.50E-16	0.997	0.003	-0.238
rs375984409	1	152255772	G	A	0.992	0.008	0.407	-0.229	0.018	3.80E-40	3.20E-38	0.992	0.008	-0.229
rs144613541	1	152270875	G	A	0.304	0.304	0.963	0.015	0.002	3.10E-11	9.80E-12	0.291	0.291	0.015
rs150597413	1	152277622	T	G	0.004	0.004	1.000	0.105	0.016	9.70E-11	9.00E-11	0.004	0.004	0.105
rs138726443	1	152280023	A	G	0.005	0.005	1.000	0.117	0.015	9.90E-17	1.80E-15	0.005	0.005	0.112
rs61816761	1	152285861	A	G	0.023	0.977	0.889	0.128	0.007	3.10E-75	4.80E-72	0.023	0.023	0.125
rs576242124	1	152390763	A	G	0.011	0.011	0.645	0.113	0.015	3.00E-15	8.40E-15	0.011	0.011	0.114
rs184958517	1	153111312	T	A	0.993	0.007	0.501	-0.135	0.017	1.30E-15	5.60E-15	0.993	0.007	-0.135
rs558560635	1	153147997	G	A	0.998	0.002	0.409	-0.272	0.034	1.20E-16	5.80E-16	0.998	0.002	-0.272
rs11264360	1	155284586	A	T	0.244	0.244	0.978	0.017	0.002	6.90E-13	3.10E-12	0.243	0.243	0.018
rs867772	1	220972343	G	A	0.684	0.316	0.994	-0.014	0.002	1.60E-11	5.80E-11	0.682	0.318	-0.014
rs10127775	1	230295789	T	A	0.607	0.393	1.000	0.012	0.002	1.60E-09	3.00E-09	0.605	0.395	0.012
rs12997242	2	21381177	A	G	0.438	0.438	0.999	-0.012	0.002	1.00E-09	1.70E-09	0.438	0.438	-0.013
rs11127048	2	27752463	A	G	0.618	0.382	0.968	0.020	0.002	3.00E-21	1.60E-20	0.617	0.383	0.018
rs6724965	2	101440151	G	A	0.171	0.171	1.000	-0.015	0.003	1.20E-08	1.80E-08	0.172	0.172	-0.017
rs7569755	2	118648261	A	G	0.292	0.292	0.988	0.014	0.002	7.20E-11	1.30E-10	0.292	0.292	0.014
rs1047891	2	211540507	A	C	0.316	0.316	1.000	-0.013	0.002	1.00E-09	1.20E-09	0.316	0.316	-0.014
rs2011425	2	234627608	G	T	0.079	0.079	0.986	-0.045	0.004	3.80E-32	3.50E-32	0.079	0.079	-0.046
rs7650253	3	49431160	A	T	0.691	0.309	0.914	0.015	0.002	5.20E-11	1.60E-10	0.690	0.310	0.015
rs1972994	3	85631142	T	A	0.649	0.351	0.996	-0.017	0.002	3.00E-16	2.80E-16	0.647	0.353	-0.018
rs6438900	3	125148287	G	C	0.260	0.260	0.987	0.014	0.002	3.60E-10	1.90E-09	0.261	0.261	0.014
rs6773343	3	141825598	T	C	0.724	0.276	0.999	0.012	0.002	1.80E-07	9.00E-08	0.720	0.280	0.013
rs78649910	4	3482213	A	T	0.110	0.110	0.978	-0.019	0.003	8.40E-09	3.30E-09	0.110	0.110	-0.018
rs7699711	4	69947596	T	G	0.454	0.454	0.998	-0.029	0.002	5.70E-47	1.30E-45	0.455	0.455	-0.029
rs529640451	4	72177044	C	G	0.993	0.007	0.637	0.176	0.015	5.50E-31	1.00E-30	0.997	0.003	0.233
rs528776789	4	72486140	A	G	0.993	0.007	0.637	0.176	0.015	5.50E-31	1.00E-30	0.993	0.007	0.178
rs113938679	4	72488025	A	G	0.006	0.006	0.771	-0.177	0.015	1.90E-32	1.20E-31	0.006	0.006	-0.184
rs564377207	4	72488525	G	C	0.996	0.004	0.527	-0.196	0.021	2.80E-21	1.30E-20	0.996	0.004	-0.199
rs186897112	4	72528565	G	A	0.998	0.002	0.450	0.247	0.034	2.60E-12	3.80E-13	0.998	0.002	0.247
rs557657187	4	72539857	G	A	0.999	0.001	0.468	0.365	0.045	3.20E-16	6.20E-16	0.999	0.001	0.365
rs145432346	4	72575017	C	T	0.826	0.174	0.775	0.108	0.003	2.80E-282	8.40E-282	0.826	0.174	0.109
rs705117	4	72608115	T	C	0.852	0.148	0.994	-0.035	0.003	8.20E-35	2.50E-34	0.849	0.151	-0.034
rs11723621	4	72615362	G	A	0.292	0.292	0.999	-0.181	0.002	3.0E-1443	7.1E-1434	0.291	0.291	-0.187
rs560384646	4	72616618	C	A	0.021	0.021	0.664	-0.188	0.009	1.30E-105	2.40E-104	0.024	0.024	-0.193
rs200641845	4	72620895	T	A	0.545	0.455	0.733	0.018	0.002	1.90E-14	6.50E-14	0.545	0.455	0.018
rs565277381	4	72625772	T	G	0.999	0.001	0.572	0.308	0.047	3.30E-11	6.60E-11	0.999	0.001	0.308
rs3775150	4	72640750	C	T	0.264	0.264	0.815	-0.090	0.003	2.30E-273	2.90E-273	0.262	0.262	-0.091
rs222026	4	72643760	T	A	0.873	0.127	0.945	-0.051	0.003	6.40E-62	8.10E-61	0.871	0.129	-0.052
rs190688847	4	72705716	C	T	0.998	0.002	0.438	0.291	0.033	4.80E-19	1.00E-18	0.998	0.002	0.291
rs184291421	4	72752846	C	T	0.994	0.006	0.749	0.170	0.015	3.20E-28	1.30E-28	0.994	0.006	0.170
rs188838036	4	72783385	A	G	0.995	0.005	0.721	0.179	0.018	1.20E-24	3.10E-24	0.995	0.005	0.179
rs186881826	4	72785743	A	T	0.220	0.220	0.904	0.044	0.003	1.60E-65	9.20E-66	0.223	0.223	0.046
rs186441690	4	72820969	G	A	0.997	0.003	0.391	-0.267	0.030	2.00E-18	6.40E-18	0.997	0.003	-0.267
rs546541682	4	72864566	T	G	0.994	0.006	0.556	-0.155	0.018	4.70E-18	5.30E-18	0.994	0.006	-0.157
rs143106299	4	72920085	T	A	0.006	0.006	0.745	-0.167	0.016	1.80E-26	1.20E-26	0.006	0.006	-0.169
rs192785674	4	73505826	A	G	0.997	0.003	0.449	0.169	0.026	1.60E-10	8.10E-11	0.997	0.003	0.169
rs58073039	4	88287363	G	A	0.299	0.299	1.000	-0.015	0.002	6.70E-11	2.20E-11	0.298	0.298	-0.014
rs28364331	4	100201295	G	A	0.019	0.019	0.977	0.061	0.007	2.90E-16	5.40E-16	0.019	0.019	0.061
rs1229984	4	100239319	C	T	0.977	0.023	1.000	-0.047	0.007	2.50E-13	4.30E-12	0.973	0.027	-0.047
rs7718395	5	118652574	G	C	0.318	0.318	0.986	0.012	0.002	1.20E-08	1.40E-08	0.320	0.320	0.013
rs3822868	6	131934986	G	A	0.835	0.165	0.993	0.022	0.003	8.70E-16	7.30E-16	0.835	0.165	0.022
rs111529171	7	21571932	C	G	0.217	0.217	0.984	-0.015	0.002	3.10E-10	2.70E-09	0.216	0.216	-0.015
rs1011468	7	104613791	A	G	0.476	0.476	0.996	-0.014	0.002	1.90E-12	1.50E-11	0.476	0.476	-0.014
rs1858889	7	107117447	C	A	0.501	0.499	0.999	0.012	0.002	1.60E-09	1.80E-09	0.501	0.499	0.013
rs804280	8	11612698	A	C	0.582	0.418	1.000	0.014	0.002	1.40E-12	7.00E-12	0.582	0.418	0.013
rs34726834	8	25889606	T	C	0.254	0.254	0.990	0.013	0.002	1.60E-08	1.00E-08	0.254	0.254	0.014
rs7828742	8	116960729	G	A	0.598	0.402	0.984	-0.022	0.002	1.20E-28	5.00E-27	0.597	0.403	-0.022
rs10818769	9	125719923	G	C	0.860	0.140	0.998	-0.017	0.003	1.60E-09	3.10E-09	0.857	0.143	-0.017
rs532436	9	136149830	A	G	0.183	0.183	1.000	-0.015	0.003	1.70E-08	6.90E-09	0.184	0.184	-0.015
rs10887718	10	82042624	T	C	0.528	0.472	0.997	-0.012	0.002	3.20E-09	3.90E-09	0.527	0.473	-0.012
rs538325438	11	13414030	C	A	0.999	0.001	0.723	0.227	0.032	8.00E-13	6.10E-13	0.999	0.001	0.227
rs373514022	11	13955649	C	T	0.998	0.002	0.673	0.203	0.029	1.60E-12	4.80E-12	0.998	0.002	0.203
rs571618690	11	13996822	A	C	0.999	0.001	0.746	0.366	0.031	3.60E-32	1.90E-31	0.999	0.001	0.366
rs191379475	11	14075712	G	A	0.989	0.011	0.560	-0.103	0.013	2.20E-15	1.60E-15	0.989	0.011	-0.103
rs561089663	11	14100539	G	C	0.998	0.002	0.773	0.409	0.030	3.40E-43	4.80E-43	0.998	0.002	0.409
rs10832218	11	14181174	C	T	0.180	0.180	0.672	-0.036	0.003	1.50E-29	7.80E-29	0.198	0.198	-0.034
rs117206369	11	14335876	T	C	0.998	0.002	0.611	0.468	0.032	5.70E-51	1.10E-48	0.998	0.002	0.468
rs567876843	11	14414139	G	T	0.995	0.005	0.607	0.542	0.019	4.90E-182	1.80E-180	0.995	0.005	0.542
rs148514005	11	14464878	T	C	0.005	0.005	0.756	-0.446	0.016	1.50E-175	2.80E-173	0.006	0.006	-0.447
rs571484036	11	14512559	A	G	0.998	0.002	0.856	-0.217	0.027	6.90E-17	4.10E-16	0.998	0.002	-0.217
rs577185477	11	14612563	C	T	0.012	0.012	0.892	-0.377	0.010	4.3E-311	5.6E-322	0.015	0.015	-0.379
rs554808052	11	14636390	C	A	0.998	0.002	0.877	0.349	0.026	3.80E-39	5.40E-40	0.998	0.002	0.349
rs10832289	11	14669496	T	A	0.409	0.409	0.999	-0.066	0.002	4.20E-233	2.70E-227	0.410	0.410	-0.069
rs187443664	11	14768892	T	G	0.989	0.011	0.537	-0.106	0.013	3.20E-16	4.00E-16	0.989	0.011	-0.106
rs188480917	11	14785870	G	C	0.011	0.011	0.943	-0.335	0.010	3.10E-248	6.30E-241	0.011	0.011	-0.343
rs534042887	11	14818258	G	T	0.997	0.003	0.797	0.392	0.020	5.00E-84	2.80E-82	0.997	0.003	0.392
rs532836473	11	14822853	G	A	0.998	0.002	0.663	0.436	0.031	2.30E-45	4.90E-44	0.998	0.002	0.436
rs201501563	11	14882470	T	C	0.123	0.123	0.660	-0.066	0.004	2.20E-68	3.20E-66	0.122	0.122	-0.066
rs117913124	11	14900931	A	G	0.028	0.028	0.985	-0.348	0.006	2.4E-695	3.0E-687	0.028	0.028	-0.354
rs117576073	11	14912573	T	G	0.012	0.012	1.000	-0.117	0.009	7.20E-38	1.80E-37	0.012	0.012	-0.115
rs150585703	11	14951216	G	C	0.996	0.004	0.632	0.482	0.020	4.40E-128	7.30E-125	0.996	0.004	0.482
rs574992951	11	16580958	C	T	0.991	0.009	0.558	0.087	0.015	8.10E-09	4.00E-09	0.991	0.009	0.087
rs567415847	11	16854631	G	A	0.998	0.002	0.379	0.283	0.037	3.40E-15	1.00E-14	0.998	0.002	0.283
rs523583	11	66070146	C	A	0.470	0.470	0.980	0.012	0.002	2.30E-09	1.10E-08	0.469	0.469	0.012
rs12803256	11	71132868	G	A	0.778	0.222	0.993	0.101	0.002	1.3E-378	5.6E-376	0.771	0.229	0.100
rs536006581	11	71135151	G	G	0.006	0.006	0.881	-0.168	0.014	9.30E-32	1.90E-31	0.009	0.009	-0.174
rs574615332	11	71144427	A	C	0.997	0.003	0.572	-0.287	0.026	3.10E-29	1.40E-28	0.997	0.003	-0.287
rs549940584	11	71222408	T	C	0.010	0.010	0.932	0.183	0.010	4.60E-70	1.00E-69	0.013	0.013	0.183
rs200454003	11	71228990	T	C	0.264	0.264	0.828	-0.087	0.003	3.60E-256	2.20E-253	0.265	0.265	-0.087
rs10793129	11	75459865	A	G	0.090	0.090	0.982	0.025	0.004	8.80E-12	2.90E-12	0.090	0.090	0.024
rs1149605	11	76485216	C	T	0.173	0.173	0.987	0.020	0.003	1.10E-13	8.20E-14	0.171	0.171	0.019
rs964184	11	116648917	C	G	0.869	0.131	1.000	0.041	0.003	2.50E-43	2.80E-43	0.864	0.136	0.040
rs2847500	11	120114421	A	G	0.123	0.123	1.000	-0.022	0.003	9.00E-14	7.00E-13	0.124	0.124	-0.021
rs12317268	12	21352541	G	A	0.151	0.151	0.995	-0.019	0.003	8.40E-12	8.20E-12	0.152	0.152	-0.019
rs9668081	12	38602911	T	C	0.471	0.471	0.992	0.012	0.002	1.30E-08	4.50E-09	0.471	0.471	0.012
rs61937878	12	96371731	T	C	0.006	0.006	1.000	0.118	0.013	4.40E-21	2.00E-20	0.006	0.006	0.119
rs10859995	12	96375682	C	T	0.584	0.416	1.000	-0.039	0.002	8.90E-81	5.40E-80	0.581	0.419	-0.039
rs8018720	14	39556185	C	G	0.823	0.177	1.000	-0.030	0.003	6.50E-33	3.30E-30	0.820	0.180	-0.032
rs261291	15	58680178	C	T	0.356	0.356	0.997	-0.024	0.002	3.60E-30	4.50E-29	0.356	0.356	-0.022
rs1800588	15	58723675	T	C	0.215	0.215	1.000	-0.030	0.002	1.40E-35	4.50E-34	0.215	0.215	-0.030
rs17765311	15	63789952	C	A	0.344	0.344	0.994	-0.016	0.002	1.60E-14	1.90E-14	0.345	0.345	-0.015
rs62007299	15	77711719	A	G	0.713	0.287	0.999	-0.014	0.002	6.30E-11	3.90E-10	0.709	0.291	-0.014
rs8063706	16	11909552	T	A	0.272	0.272	0.989	0.014	0.002	7.50E-09	1.80E-09	0.273	0.273	0.013
rs77924615	16	20392332	A	G	0.198	0.198	0.980	-0.015	0.003	3.20E-09	2.10E-09	0.198	0.198	-0.016
rs71383766	16	30930233	T	C	0.425	0.425	0.939	0.014	0.002	2.50E-10	1.90E-10	0.420	0.420	0.013
rs1800775	16	56995236	A	C	0.486	0.486	1.000	-0.017	0.002	1.70E-17	3.30E-17	0.486	0.486	-0.017
rs2909218	17	66464546	T	C	0.796	0.204	0.990	0.017	0.003	3.50E-11	3.40E-11	0.793	0.207	0.017
rs8091117	18	28919794	A	C	0.065	0.065	1.000	-0.024	0.004	3.40E-09	3.30E-09	0.065	0.065	-0.024
rs2037511	18	61366207	A	G	0.165	0.165	0.999	0.017	0.003	2.90E-09	5.40E-10	0.165	0.165	0.016
rs57631352	19	4338173	G	A	0.297	0.297	0.998	-0.013	0.002	4.40E-08	2.00E-08	0.297	0.297	-0.013
rs73015021	19	11192915	G	A	0.121	0.121	0.996	0.024	0.003	2.20E-15	1.10E-14	0.121	0.121	0.023
rs10500209	19	11979164	C	T	0.282	0.282	1.000	-0.014	0.002	2.00E-09	2.50E-09	0.282	0.282	-0.013
rs58542926	19	19379549	T	C	0.076	0.076	1.000	0.033	0.004	5.20E-19	6.50E-18	0.076	0.076	0.032
rs3814995	19	36342212	T	C	0.312	0.312	1.000	-0.014	0.002	7.00E-11	1.20E-10	0.312	0.312	-0.015
rs1065853	19	45413233	T	T	0.082	0.082	0.998	0.028	0.004	2.30E-15	1.90E-14	0.082	0.082	0.027
rs157595	19	45425460	G	A	0.615	0.385	0.951	-0.015	0.002	2.30E-12	5.20E-12	0.614	0.386	-0.016
rs112285002	19	48374320	T	C	0.165	0.165	0.995	0.062	0.003	2.90E-115	5.30E-113	0.160	0.160	0.060
rs62130059	19	48461240	C	A	0.332	0.332	0.874	-0.028	0.002	5.00E-35	9.10E-35	0.336	0.336	-0.027
rs10426	19	51517798	A	G	0.213	0.213	1.000	0.025	0.002	1.00E-23	4.90E-24	0.213	0.213	0.025
rs8103262	19	53065814	C	T	0.305	0.305	0.992	0.012	0.002	4.00E-07	1.80E-07	0.305	0.305	0.013
rs6123359	20	52714706	G	A	0.106	0.106	0.972	0.031	0.003	1.80E-21	7.60E-21	0.105	0.105	0.032
rs6127099	20	52731402	T	A	0.278	0.278	0.949	-0.034	0.002	2.20E-47	1.40E-48	0.279	0.279	-0.037
rs2585442	20	52737123	G	C	0.247	0.247	0.967	0.033	0.002	1.10E-40	4.80E-43	0.246	0.246	0.034
rs2762942	20	52788925	A	G	0.946	0.054	1.000	0.052	0.004	1.80E-30	3.30E-31	0.942	0.058	0.053
rs2229742	21	16339172	C	G	0.103	0.103	1.000	-0.025	0.003	2.00E-13	2.10E-13	0.104	0.104	-0.026
rs2074735	22	31535872	C	G	0.064	0.064	1.000	0.028	0.004	2.60E-11	4.00E-11	0.064	0.064	0.027
rs960596	22	41393520	T	C	0.339	0.339	0.986	0.012	0.002	7.50E-09	1.60E-08	0.340	0.340	0.012

	META-ANALYSIS (N=443,734)		GCTA-COJO							Ely 1,25-dihydroxyvitamin D GWAS				CLOSEST KNOWN 25OHD VARIANT				LOCUS ANNOTATION		
RSID	SE	P	EAF.J	MAF.J	β.J	SE.J	P.J	LD.R	VAR.J	EAF	β	SE	P	T.VAR	T.DIST	T.GENE	STATUS	A.GENE	FUNCTION	FUNCTION
rs6698680	0.002	8.99E-10	0.465	0.465	-0.012	0.002	7.47E-10	-0.003	0.0001	0.4624	-0.0578	0.0523	0.2689				NOVEL	RER1	intron	intron
rs3750296	0.002	2.09E-24	0.340	0.340	-0.021	0.002	3.04E-24	-0.008	0.0002	0.3396	0.0041	0.0553	0.9416				NOVEL	PADI1	intron	intron
rs7519574	0.003	2.09E-11	0.180	0.180	0.017	0.003	4.03E-11	0.002	0.0001	0.1669	-0.0763	0.0707	0.2811				NOVEL	RP4-657M3.	intergenic	intergeni
rs56044892	0.002	2.85E-10	0.204	0.204	0.015	0.002	3.13E-10	0.000	0.0001	0.2126	-0.0068	0.0661	0.9186				NOVEL	FOXO6	intron	intron
rs2934744	0.002	3.96E-26	0.643	0.643	-0.022	0.002	4.13E-26	0.000	0.0002	0.6331	-0.1131	0.0557	0.0426				NOVEL	DOCK7	intron	intron
rs7528419	0.002	2.41E-16	0.224	0.224	0.019	0.002	2.43E-16	0.000	0.0001	0.2253	-0.1585	0.0596	0.0080				NOVEL	CELSR2	3_prime_UT	3_prime_
rs3768013	0.002	1.37E-13	0.366	0.366	-0.012	0.002	3.86E-09	-0.039	0.0001	0.3778	-0.0566	0.0544	0.2989				NOVEL	ARNT	intron	intron
rs115045402	0.007	3.05E-55	0.013	0.013	0.070	0.008	1.58E-19	0.472	0.0003	0.0233	0.1985	0.2186	0.3643				NOVEL	FLG	intergenic	intergeni
rs12123821	0.005	2.25E-59	0.049	0.049	0.053	0.005	1.28E-24	0.007	0.0003	0.0431	0.0411	0.1504	0.7845				NOVEL	FLG	intron	intron
rs201561609	0.012	6.99E-28	0.996	0.004	-0.097	0.012	6.63E-16	-0.004	0.0002	0.9844	0.0236	0.2742	0.9314				NOVEL	FLG	missence	missence
rs185433896	0.019	1.50E-38	0.999	0.001	-0.209	0.019	7.24E-28	-0.002	0.0006	0.9963	-0.2418	0.6025	0.6882				NOVEL	FLG	intron	intron
rs189918701	0.029	2.47E-16	0.999	0.001	-0.183	0.029	3.29E-10	-0.002	0.0002	Not available in Ely datasets due to low MAF of varia							NOVEL	FLG	intron	intron
rs375984409	0.018	3.22E-38	0.999	0.001	-0.186	0.018	1.53E-25	-0.058	0.0006	Not available in Ely datasets due to low MAF of varia							NOVEL	FLG	intron	intron
rs144613541	0.002	6.49E-12	0.291	0.291	0.016	0.002	1.52E-12	-0.036	0.0001	0.2828	0.1190	0.0605	0.0497				NOVEL	FLG	downstream	downstre
rs150597413	0.016	6.18E-11	0.004	0.004	0.113	0.016	1.56E-12	-0.002	0.0001	0.0029	0.1257	0.5599	0.8224				NOVEL	FLG		
rs138726443	0.014	8.81E-15	0.005	0.005	0.124	0.014	1.36E-17	0.004	0.0001	0.0057	0.4800	0.4157	0.2486				NOVEL	FLG		
rs61816761	0.007	8.57E-74	0.016	0.016	0.110	0.007	5.39E-54	-0.003	0.0005	0.0172	-0.0118	0.2361	0.9602				NOVEL	FLG	stop_lost	stop_lost
rs576242124	0.015	3.08E-15	0.003	0.003	0.092	0.015	2.59E-10	-0.003	0.0002	0.0124	0.2931	0.3632	0.4200				NOVEL	FLG	upstream	upstream
rs184958517	0.017	5.55E-15	0.999	0.001	-0.105	0.017	1.21E-09	-0.001	0.0002	0.9960	0.1670	0.6270	0.7900				NOVEL	FLG	downstream	downstre
rs558560635	0.034	5.83E-16	1.000	0.000	-0.243	0.034	4.45E-13	0.004	0.0003	Not available in Ely datasets due to low MAF of varia							NOVEL	FLG	intron	intron
rs11264360	0.002	3.34E-15	0.241	0.241	0.018	0.002	1.12E-15	0.000	0.0001	0.2333	-0.0857	0.0607	0.1586				NOVEL	FDPS	indels	indels
rs867772	0.002	3.64E-11	0.684	0.316	-0.014	0.002	3.31E-11	0.002	0.0001	0.6880	-0.0022	0.0555	0.9686				NOVEL	MARC_1	intron	intron
rs10127775	0.002	3.43E-09	0.609	0.391	0.012	0.002	3.11E-09	0.000	0.0001	0.5956	0.0085	0.0532	0.8735				NOVEL			
rs12997242	0.002	2.23E-10	0.441	0.441	-0.012	0.002	2.32E-10	-0.001	0.0001	0.4304	-0.0073	0.0530	0.8903				NOVEL	TDRD15	intergenic	intergeni
rs11127048	0.002	6.41E-19	0.620	0.380	0.018	0.002	6.72E-19	0.000	0.0002	0.6313	0.0266	0.0546	0.6267				NOVEL	GCKR	intergenic	intergeni
rs6724965	0.003	1.29E-10	0.173	0.173	-0.017	0.003	1.34E-10	-0.001	0.0001	0.1810	0.0701	0.0692	0.3120				NOVEL	NPAS2	intron	intron
rs7569755	0.002	8.03E-11	0.291	0.291	0.014	0.002	8.35E-11	0.000	0.0001	0.2813	-0.0274	0.0585	0.6396				NOVEL	HTR5BP	intron	intron
rs1047891	0.002	1.16E-11	0.315	0.315	-0.014	0.002	1.16E-11	0.000	0.0001	0.3225	-0.0137	0.0570	0.8094				NOVEL	CPS1	missence	missence
rs2011425	0.004	9.66E-38	0.076	0.076	-0.046	0.004	9.93E-38	0.000	0.0003	0.0644	0.0810	0.1076	0.4518				NOVEL	UGT1A4	missence	missence
rs7650253	0.002	1.76E-10	0.695	0.305	0.015	0.002	1.76E-10	0.000	0.0001	0.6884	0.0004	0.0595	0.9947				NOVEL	RHOA	intron	intron
rs1972994	0.002	7.99E-18	0.650	0.350	-0.018	0.002	8.04E-18	0.000	0.0001	0.6659	-0.1074	0.0554	0.0530				NOVEL	CADM2	intron	intron
rs6438900	0.002	9.59E-10	0.255	0.255	0.014	0.002	1.16E-09	0.005	0.0001	0.2515	-0.0052	0.0617	0.9334				NOVEL	MRPL3	intergenic	intergeni
rs6773343	0.002	5.20E-09	0.722	0.278	0.013	0.002	6.28E-09	0.000	0.0001	0.7241	0.0175	0.0577	0.7621				NOVEL	TFDP2	intron	intron
rs78649910	0.003	4.32E-09	0.106	0.106	-0.018	0.003	3.41E-09	0.006	0.0001	0.1193	-0.0216	0.0818	0.7921				NOVEL	DOK7	intron	intron
rs7699711	0.002	6.97E-49	0.453	0.453	-0.030	0.002	4.85E-50	-0.004	0.0004	0.4339	0.0078	0.0517	0.8804				NOVEL	UGT2B7	intron	intron
rs529640451	0.027	2.25E-17	0.999	0.001	0.174	0.027	2.20E-10	-0.002	0.0002	Not available in Ely datasets due to low MAF of varia				rs2282679	431339	GC	KNOWN	GC	intergenic	intergeni
rs528776789	0.015	3.67E-31	0.998	0.002	0.122	0.015	2.45E-15	0.005	0.0002	0.9941	0.0579	0.4681	0.9015	rs2282679	122243	GC	KNOWN	GC	intergenic	intergeni
rs113938679	0.015	5.88E-36	0.003	0.003	-0.099	0.015	2.21E-11	0.003	0.0001	0.0077	0.1986	0.4416	0.6530	rs2282679	120358	GC	KNOWN	GC	intergenic	intergeni
rs564377207	0.021	1.05E-21	0.999	0.001	-0.159	0.021	2.23E-14	-0.001	0.0002	0.9960	-0.6351	0.5139	0.2169	rs2282679	119858	GC	KNOWN	GC	intergenic	intergeni
rs186897112	0.034	3.79E-13	1.000	0.000	0.201	0.034	3.81E-09	-0.001	0.0002	Not available in Ely datasets due to low MAF of varia				rs2282679	79818	GC	KNOWN	GC	intergenic	intergeni
rs557657187	0.045	6.18E-16	1.000	0.000	0.287	0.045	2.19E-10	0.005	0.0002	Not available in Ely datasets due to low MAF of varia				rs2282679	68526	GC	KNOWN	GC	intergenic	intergeni
rs145432346	0.003	6.78E-286	0.873	0.127	0.035	0.003	2.26E-27	0.040	0.0003	0.8220	0.1976	0.0799	0.0136	rs2282679	33366	GC	KNOWN	GC	intergenic	intergeni
rs705117	0.003	1.71E-36	0.852	0.148	0.031	0.003	1.12E-27	0.270	0.0003	0.8667	0.0551	0.0756	0.4658	rs2282679	268	GC	KNOWN	GC	intron	intron
rs11723621	0.002	2.903E-1689	0.294	0.294	-0.156	0.003	0	0.153	0.0101	0.2898	-0.0766	0.0579	0.1862	rs2282679	6979	GC	KNOWN	GC	intron	intron
rs560384646	0.009	6.91E-112	0.010	0.010	-0.089	0.009	3.23E-24	0.051	0.0004	0.0190	-0.5370	0.2940	0.0700	rs2282679	8235	GC	KNOWN	GC	indel	indel
rs200641845	0.002	6.92E-14	0.550	0.450	0.017	0.002	5.23E-12	0.016	0.0001	0.5567	-0.1588	0.0692	0.0221	rs2282679	12512	GC	KNOWN	GC	intron	intron
rs565277381	0.047	6.62E-11	1.000	0.000	0.279	0.047	3.55E-09	-0.004	0.0002	Not available in Ely datasets due to low MAF of varia				rs2282679	17389	GC	KNOWN	GC	intron	intron
rs3775150	0.002	3.90E-295	0.218	0.218	-0.071	0.003	3.46E-109	-0.569	0.0019	0.2746	0.0288	0.0699	0.6804	rs2282679	32367	GC	KNOWN	GC	indel	indel
rs222026	0.003	6.98E-68	0.882	0.118	-0.051	0.004	1.09E-40	-0.003	0.0006	0.8630	-0.0528	0.0774	0.4957	rs2282679	35377	GC	KNOWN	GC	intron	intron
rs190688847	0.033	1.02E-18	1.000	0.000	0.254	0.033	1.26E-14	-0.002	0.0003	Not available in Ely datasets due to low MAF of varia				rs2282679	97333	GC	KNOWN	GC	intergenic	intergeni
rs184291421	0.015	1.25E-28	0.997	0.003	0.090	0.015	5.03E-09	-0.004	0.0001	0.9954	-0.3770	0.4362	0.3876	rs2282679	144463	GC	KNOWN	GC	intergenic	intergeni
rs188838036	0.018	3.07E-24	0.998	0.002	0.118	0.018	3.14E-11	0.024	0.0001	0.9950	0.5427	0.4213	0.1980	rs2282679	175002	GC	KNOWN	GC	intergenic	intergeni
rs186881826	0.002	3.64E-77	0.190	0.190	0.020	0.003	1.43E-15	-0.038	0.0001	0.2256	-0.0512	0.0659	0.4377	rs2282679	177360	GC	KNOWN	GC	intergenic	intergeni
rs186441690	0.030	1.96E-18	1.000	0.000	-0.234	0.031	1.79E-14	-0.001	0.0003	Not available in Ely datasets due to low MAF of varia				rs2282679	212586	GC	KNOWN	GC	intergenic	intergeni
rs546541682	0.018	2.06E-18	0.999	0.001	-0.113	0.018	3.45E-10	-0.005	0.0001	0.9945	0.3745	0.5041	0.4578	rs2282679	256183	GC	KNOWN	GC	intergenic	intergeni
rs143106299	0.015	1.50E-28	0.002	0.002	-0.090	0.015	4.62E-09	-0.013	0.0001	0.0034	-0.0212	0.6069	0.9722	rs2282679	311702	GC	KNOWN	GC	intron	intron
rs192785674	0.026	8.14E-11	1.000	0.000	0.181	0.026	3.48E-12	-0.001	0.0002	Not available in Ely datasets due to low MAF of varia				rs2282679	897443	GC	KNOWN	GC	intergenic	intergeni
rs58073039	0.002	2.16E-11	0.300	0.300	-0.013	0.002	2.84E-10	0.006	0.0001	0.2753	-0.0439	0.0581	0.4509				NOVEL	HSD17B11	intron	intron
rs28364331	0.007	1.31E-17	0.018	0.018	0.063	0.007	3.06E-18	0.019	0.0001	0.0170	0.0808	0.2093	0.6996				NOVEL	ADH1A	splice_region	splice_re
rs1229984	0.006	4.85E-13	0.977	0.023	-0.047	0.006	2.43E-13	0.000	0.0001	0.9718	-0.0350	0.1741	0.8406				NOVEL	ADH1A	missence	missence
rs7718395	0.002	1.67E-09	0.319	0.319	0.013	0.002	1.68E-09	0.000	0.0001	0.3105	0.0359	0.0557	0.5195				NOVEL	TNFAIP8	intron	intron
rs3822868	0.003	1.41E-15	0.837	0.163	0.022	0.003	1.41E-15	0.000	0.0001	0.8403	0.1539	0.0684	0.0247				NOVEL	MED23	intron	intron
rs111529171	0.002	6.24E-11	0.216	0.216	-0.015	0.002	6.26E-11	0.000	0.0001	0.2156	0.0423	0.0631	0.5031				NOVEL	DNAH11	intergenic	intergeni
rs1011468	0.002	1.35E-12	0.474	0.474	-0.014	0.002	1.39E-12	0.005	0.0001	0.4446	-0.1249	0.0540	0.0210				NOVEL	LINC01004	intron	intron
rs1858889	0.002	3.85E-11	0.500	0.500	0.013	0.002	3.03E-11	0.000	0.0001	0.4976	-0.0231	0.0527	0.6619				NOVEL	COG5	intron	intron
rs804280	0.002	4.43E-11	0.584	0.416	0.016	0.002	9.90E-16	-0.011	0.0001	0.5747	-0.0647	0.0523	0.2162				NOVEL	GATA4	intron	intron
rs34726834	0.002	6.65E-10	0.251	0.251	0.014	0.002	3.39E-10	0.000	0.0001	0.2677	-0.0429	0.0604	0.4781				NOVEL	EBF2	intron	intron
rs7828742	0.002	3.06E-28	0.598	0.402	-0.024	0.002	2.85E-33	0.000	0.0003	0.5953	-0.0340	0.0522	0.5146				NOVEL	LINC00536	downstream	downstre
rs10818769	0.003	3.35E-09	0.862	0.138	-0.017	0.003	2.99E-09	-0.003	0.0001	0.8450	0.0375	0.0699	0.5920				NOVEL	DNAH11	intergenic	intergeni
rs532436	0.003	2.17E-09	0.186	0.186	-0.015	0.003	1.94E-09	0.000	0.0001	0.2105	0.0375	0.0630	0.5519				NOVEL	ABO	intron	intron
rs10887718	0.002	1.44E-10	0.528	0.472	-0.013	0.002	1.18E-10	0.005	0.0001	0.5118	0.0309	0.0507	0.5419				NOVEL	MAT1A	intron	intron
rs538325438	0.032	6.07E-13	0.999	0.001	-0.451	0.038	4.61E-32	-0.002	0.0006	Not available in Ely datasets due to low MAF of varia				rs10741657	1500848	CYP2R1	KNOWN	CYP2R1	intron	intron
rs373514022	0.029	4.77E-12	0.999	0.001	0.213	0.029	4.15E-13	-0.001	0.0002	0.9980	-0.8925	0.6801	0.1898	rs10741657	959229	CYP2R1	KNOWN	CYP2R1	intergenic	intergeni
rs571618690	0.031	1.90E-31	0.999	0.001	0.228	0.032	1.40E-12	-0.002	0.0001	Not available in Ely datasets due to low MAF of varia				rs10741657	918056	CYP2R1	KNOWN	CYP2R1	intron	intron
rs191379475	0.013	1.70E-15	0.998	0.002	-0.088	0.013	1.22E-11	-0.002	0.0002	0.9924	-0.3289	0.4137	0.4269	rs10741657	839166	CYP2R1	KNOWN	CYP2R1	intron	intron
rs561089663	0.030	4.79E-43	0.999	0.001	0.204	0.031	4.31E-11	0.000	0.0001	Not available in Ely datasets due to low MAF of varia				rs10741657	814339	CYP2R1	KNOWN	CYP2R1	intron	intron
rs10832218	0.003	7.09E-32	0.109	0.109	-0.018	0.003	3.06E-10	-0.018	0.0001	0.1688	-0.1167	0.0913	0.2016	rs10741657	733704	CYP2R1	KNOWN	CYP2R1	intron	intron
rs117206369	0.032	1.07E-48	0.999	0.001	0.233	0.033	1.10E-12	0.044	0.0002	0.9981	-0.3892	0.8084	0.6304	rs10741657	579002	CYP2R1	KNOWN	CYP2R1	intron	intron
rs567876843	0.019	1.83E-180	0.998	0.002	0.538	0.023	3.35E-116	0.002	0.0027	0.9929	0.3396	0.3611	0.3473	rs10741657	500739	CYP2R1	KNOWN	CYP2R1	intergenic	intergeni
rs148514005	0.015	1.37E-184	0.003	0.003	-0.144	0.018	4.99E-15	0.004	0.0002	0.0090	-0.0118	0.3060	0.9693	rs10741657	450000	CYP2R1	KNOWN	CYP2R1	downstream	downstre
rs571484036	0.027	4.13E-16	0.999	0.001	-0.246	0.027	3.43E-20	0.003	0.0002	Not available in Ely datasets due to low MAF of varia				rs10741657	402319	CYP2R1	KNOWN	CYP2R1	intron	intron
rs577185477	0.010	1.624E-342	0.007	0.007	-0.153	0.012	7.55E-37	0.005	0.0007	0.0144	-0.2018	0.2472	0.4146	rs10741657	302315	CYP2R1	KNOWN	CYP2R1	intron	intron
rs554808052	0.026	5.41E-40	0.999	0.001	0.198	0.028	7.88E-13	0.026	0.0001	0.9971	0.9106	0.5379	0.0909	rs10741657	278488	CYP2R1	KNOWN	CYP2R1	intron	intron
rs10832289	0.002	2.03E-266	0.409	0.409	-0.086	0.002	0	-0.098	0.0036	0.4229	-0.0795	0.0537	0.1386	rs10741657	245382	CYP2R1	KNOWN	CYP2R1	intron	intron
rs187443664	0.013	3.49E-16	0.998	0.002	-0.079	0.013	1.52E-09	0.004	0.0001	0.9911	-0.4051	0.3859	0.2942	rs10741657	145986	CYP2R1	KNOWN	CYP2R1	intron	intron
rs188480917	0.010	5.00E-275	0.009	0.009	-0.167	0.013	3.21E-37	-0.294	0.0006	0.0148	-0.3677	0.2372	0.1216	rs10741657	129008	CYP2R1	KNOWN	CYP2R1	intron	intron
rs534042887	0.020	2.82E-82	0.998	0.002	0.192	0.021	2.21E-19	-0.002	0.0002	0.9966	0.6935	0.5670	0.2217	rs10741657	96620	CYP2R1	KNOWN	CYP2R1	intron	intron
rs532836473	0.031	4.90E-44	0.999	0.001	0.267	0.032	4.77E-17	0.014	0.0002	0.9973	-0.1641	0.5965	0.7833	rs10741657	92025	CYP2R1	KNOWN	CYP2R1	intron	intron
rs201501563	0.004	9.17E-67	0.065	0.065	-0.035	0.004	1.96E-18	-0.045	0.0003	0.1261	-0.0653	0.1055	0.5365	rs10741657	32408	CYP2R1	KNOWN	CYP2R1		
rs117913124	0.006	1.653E-775	0.026	0.026	-0.206	0.009	2.94E-107	-0.023	0.0023	0.0408	-0.2699	0.1333	0.0433	rs10741657	13947	CYP2R1	KNOWN	CYP2R1	synonymous	synonym
rs117576073	0.009	1.22E-38	0.013	0.013	-0.166	0.009	1.40E-78	0.004	0.0007	0.0082	-0.0562	0.3778	0.8818	rs10741657	2305	CYP2R1	KNOWN	CYP2R1	5_prime_UT	5_prime_
rs150585703	0.020	7.16E-125	0.998	0.002	0.243	0.022	1.56E-27	0.008	0.0005	0.9947	-0.1599	0.4394	0.7160	rs10741657	36338	CYP2R1	KNOWN	CYP2R1	intron	intron
rs574992951	0.015	4.04E-09	0.998	0.002	0.089	0.015	1.69E-09	-0.002	0.0001	0.9905	-0.4286	0.3502	0.2213				NOVEL	PLEKHA7	intron	intron
rs567415847	0.037	1.03E-14	1.000	0.000	0.302	0.037	1.88E-16	0.000	0.0004	Not available in Ely datasets due to low MAF of variant and proxy variants							NOVEL	PLEKHA7	intron	intron
rs523583	0.002	5.58E-10	0.467	0.467	0.013	0.002	6.60E-12	-0.009	0.0001	0.4618	0.0231	0.0532	0.6635				NOVEL	TMEM151A	intergenic	intergeni
rs12803256	0.002	8.599E-407	0.778	0.222	0.087	0.003	1.64E-195	-0.124	0.0027	0.7576	0.0607	0.0586	0.3006	rs12785878	34581	NADSYN/DH	KNOWN	FLJ42102	non_coding_	non_codi
rs536006581	0.014	8.87E-35	0.004	0.004	-0.107	0.014	5.64E-14	-0.002	0.0002	0.0050	0.6150	0.4180	0.1400	rs12785878	32298	NADSYN/DH	KNOWN	FLJ42102	downstream	downstre
rs574615332	0.026	1.38E-28	0.999	0.999	-0.206	0.026	5.87E-15	0.005	0.0002	Not available in Ely datasets due to low MAF of varia				rs12785878	23022	NADSYN/DH	KNOWN	FLJ42102	intron	intron
rs549940584	0.010	2.31E-72	0.008	0.008	0.147	0.010	1.93E-45	-0.058	0.0006	0.0044	-0.5143	0.5753	0.3717	rs12785878	54959	NADSYN/DH	KNOWN	FLJ42102	intron	intron
rs200454003	0.003	3.68E-256	0.249	0.249	-0.029	0.003	3.49E-21	-0.028	0.0003	0.2909	-0.0055	0.0651	0.9323	rs12785878	61541	NADSYN/DH	KNOWN	FLJ42102	intron	intron
rs10793129	0.003	1.64E-12	0.086	0.086	0.025	0.003	4.11E-13	-0.016	0.0001	0.0796	-0.0015	0.0991	0.9876				NOVEL	RP11-21L23.	intergenic	intergeni
rs1149605	0.003	7.34E-14	0.174	0.174	0.020	0.003	3.36E-15	0.000	0.0001	0.1834	0.0150	0.0654	0.8193				NOVEL	RP11-21L23.	intergenic	intergeni
rs964184	0.003	5.11E-44	0.868	0.132	0.040	0.003	1.30E-43	-0.009	0.0004	0.8489	0.0453	0.0719	0.5289				NOVEL	ZPR1	3_prime_UT	3_prime_
rs2847500	0.003	7.79E-13	0.122	0.122	-0.021	0.003	1.93E-12	0.000	0.0001	0.1184	-0.0793	0.0937	0.3975				NOVEL	ZPR1	intron	intron
rs12317268	0.003	9.15E-12	0.149	0.149	-0.019	0.003	9.20E-12	0.000	0.0001	0.1445	0.1254	0.0732	0.0872				NOVEL	SLCO1B1	intron	intron
rs9668081	0.002	5.38E-09	0.470	0.470	0.012	0.002	5.40E-09	0.000	0.0001	0.4919	0.0403	0.0516	0.4355				NOVEL	FAM166AP9	upstream	upstream
rs61937878	0.012	4.43E-22	0.007	0.007	0.104	0.012	5.63E-17	-0.095	0.0001	0.0094	0.1517	0.2699	0.5744	rs10745742	13202	AMDHD1	KNOWN	HAL	missence	missence
rs10859995	0.002	7.03E-89	0.588	0.412	-0.041	0.002	3.03E-91	0.235	0.0008	0.5776	-0.0709	0.0524	0.1761	rs10745742	17153	AMDHD1	KNOWN	HAL	intron	intron
rs8018720	0.003	4.04E-36	0.824	0.176	-0.032	0.003	4.10E-36	0.000	0.0003	0.8384	-0.0894	0.0694	0.1985	rs8018720	0	SEC23A	KNOWN	SEC23A	missence	missence
rs261291	0.002	2.89E-28	0.356	0.356	-0.023	0.002	2.46E-29	-0.017	0.0002	0.3707	-0.0135	0.0535	0.8015				NOVEL	LIPC	intron	intron
rs1800588	0.002	2.65E-36	0.214	0.214	-0.030	0.002	3.17E-37	0.002	0.0003	0.2146	-0.0976	0.0625	0.1189				NOVEL	LIPC	intron	intron
rs17765311	0.002	1.35E-13	0.344	0.344	-0.015	0.002	1.18E-13	0.004	0.0001	0.3616	0.0391	0.0544	0.4725				NOVEL	AC007950.2	downstream	downstre
rs62007299	0.002	1.69E-11	0.712	0.288	-0.014	0.002	3.33E-11	0.000	0.0001	0.6934	0.0000	0.0563	0.9998				NOVEL	PEAK1	intron	intron
rs8063706	0.002	3.64E-09	0.269	0.269	0.013	0.002	4.27E-09	-0.002	0.0001	0.2923	0.0307	0.0592	0.6042				NOVEL	BCAR4	downstream	downstre
rs77924615	0.002	1.46E-10	0.194	0.194	-0.016	0.002	2.28E-10	-0.010	0.0001	0.1984	0.2055	0.0682	0.0027				NOVEL	PDILT	intron	intron
rs71383766	0.002	1.15E-09	0.431	0.431	0.012	0.002	1.86E-09	0.000	0.0001	0.4521	0.0299	0.0547	0.5847				NOVEL	FBXL19	upstream	upstream
rs1800775	0.002	1.56E-17	0.488	0.488	-0.017	0.002	1.57E-17	0.000	0.0001	0.4606	-0.0311	0.0513	0.5451				NOVEL	CETP	upstream	upstream
rs2909218	0.002	2.81E-12	0.799	0.201	0.017	0.002	2.82E-12	0.000	0.0001	0.7986	0.1088	0.0654	0.0967				NOVEL	RP11-120M1	intron	intron
rs8091117	0.004	1.03E-09	0.064	0.064	-0.024	0.004	9.48E-10	-0.002	0.0001	0.0689	0.0407	0.1007	0.6860				NOVEL	DSG1	missence	missence
rs2037511	0.003	9.29E-10	0.168	0.168	0.016	0.003	8.35E-10	0.000	0.0001	0.1638	0.0109	0.0677	0.8723				NOVEL	SERPINB11	intron	intron
rs57631352	0.002	1.48E-09	0.297	0.297	-0.013	0.002	1.50E-09	-0.007	0.0001	0.3052	0.0043	0.0550	0.9372				NOVEL	STAP2	intron	intron
rs73015021	0.003	1.15E-14	0.118	0.118	0.022	0.003	6.29E-14	-0.029	0.0001	0.1291	0.0413	0.0763	0.5885				NOVEL	LDLR	intergenic	intergeni
rs10500209	0.002	6.18E-10	0.284	0.284	-0.013	0.002	2.73E-09	0.000	0.0001	0.2814	-0.0770	0.0569	0.1767				NOVEL	LDLR	missence	missence
rs58542926	0.004	8.57E-19	0.077	0.077	0.033	0.004	2.63E-19	-0.032	0.0002	0.0722	0.0921	0.0979	0.3471				NOVEL	TM6SF2	missence	missence
rs3814995	0.002	2.83E-12	0.313	0.313	-0.015	0.002	1.08E-12	0.001	0.0001	0.3243	-0.0557	0.0657	0.3963				NOVEL	NPHS1	missence	missence
rs1065853	0.004	8.32E-14	0.081	0.081	0.028	0.004	2.24E-14	0.032	0.0001	0.0880	0.0140	0.0900	0.8700				NOVEL	APOC1	upstream	upstream
rs157595	0.002	2.95E-14	0.622	0.378	-0.016	0.002	4.25E-15	-0.002	0.0001	0.6154	-0.1436	0.0546	0.0087				NOVEL	APOC1	downstream	downstre
rs112285002	0.003	1.77E-110	0.162	0.162	0.056	0.003	1.49E-90	-0.256	0.0008	0.1548	0.0638	0.0702	0.3641				NOVEL	SULT2A1	3_prime_UT	3_prime_
rs62130059	0.002	9.25E-34	0.322	0.322	-0.016	0.002	2.64E-12	0.000	0.0001	0.3238	-0.0172	0.0623	0.7827				NOVEL	SULT2A1	intergenic	intergeni
rs10426	0.002	3.31E-26	0.213	0.213	0.025	0.002	1.59E-26	-0.009	0.0002	0.2002	0.0311	0.0641	0.6274				NOVEL	KLK10	3_prime_UT	3_prime_
rs8103262	0.002	3.18E-09	0.304	0.304	0.013	0.002	6.80E-10	-0.005	0.0001	0.3023	-0.0208	0.0569	0.7147				NOVEL	ZNF808	intron	intron
rs6123359	0.003	7.74E-24	0.102	0.102	0.024	0.003	7.48E-14	-0.102	0.0001	0.1115	0.0204	0.0886	0.8180	rs17216707	17656	CYP24A1	KNOWN	RP13-379L1	intergenic	intergeni
rs6127099	0.002	9.30E-62	0.270	0.270	-0.027	0.002	2.22E-32	-0.258	0.0003	0.2921	0.2307	0.0627	0.0003	rs17216707	960	CYP24A1	KNOWN	RP13-379L1	intergenic	intergeni
rs2585442	0.002	6.87E-49	0.241	0.241	0.023	0.002	3.96E-23	0.076	0.0002	0.2257	-0.0527	0.0659	0.4245	rs17216707	4761	CYP24A1	KNOWN	RP13-379L1	intergenic	intergeni
rs2762942	0.004	7.99E-35	0.945	0.055	0.043	0.004	1.69E-23	0.000	0.0002	0.9392	-0.0825	0.1286	0.5213	rs17216707	56563	CYP24A1	KNOWN	RP13-379L1	intron	intron
rs2229742	0.003	7.13E-16	0.105	0.105	-0.026	0.003	7.16E-16	0.000	0.0001	0.1001	0.0510	0.0941	0.5878				NOVEL	NRIP1	missence	missence
rs2074735	0.004	6.55E-12	0.065	0.065	0.027	0.004	7.12E-12	0.002	0.0001	0.0680	-0.0522	0.1093	0.6332				NOVEL	PLA2G3	missence	missence
rs960596	0.002	2.23E-09	0.340	0.340	0.012	0.002	2.43E-09	0.000	0.0001	0.3392	-0.0461	0.0550	0.4015				NOVEL	SCUBE1	intergenic	intergeni

N Sample size
RSID Reference SNP cluster ID
CHR Chromosome
BP Base pair position of the variant according to human reference sequence (GRCh37), Hg19
EA Effect allele
NEA Non-effect allele
EAF Effect allele frequency
MAF Minor allele frequency
INFO Imputation information score
β Per allele effect in standard deviations of standardized log-transformed 25OHD or 1,25 dihydroxvitamin D
SE Standard error of the β
P Strength of evidence against the null hypothesis of no association between variant and 25OHD (i.e. P-value) from standard linear regression
P.I p-value from infinitesimal mixed model assocoation test
EAF.J Frequency of the effect allele in the reference sample used for conditional and joint genome-wide association analysis
MAF.J Minor allele frequency of variant in reference databases used for conditional and joint genome-wide association analysis
β.J Per allele effect estimated from joint analysis of conditionally associated snps
SE.J Standard error of β.J
P.J Strength of evidence against the null hypothesis of no association between the variant and 25OHD as estimated by conditional and joint genome-wide association analysis (i.e. P-value)
LD.R LD correlation between the SNP i and SNP i + 1
VAR.J Proportion of variance explained by the conditionally associated variant
T.VAR Closest published 25OHD associated variant
T.DIST Distance between conditionally independent 25OHD associated variant and closest published 25OHD associated variant
T.GENE Closest gene to published 25OHD associated variant
STATUS Status of locus: novel - published 25OHD variant is not within 1mb of any conditionally associated variant
A.GENE The name of the gene situated closest to variant that has smallest P-value of all conditionally independent variants present in the same locus
FUNCTION Functional annotation of the conditionally independent variant

Chapter 6: General Discussion

The purpose of this thesis was to advance the knowledge on the genetic control of vitamin D, using methods from genetic epidemiology, bioinformatics, and statistical genetics. This thesis represents a dramatic leap forward in identifying novel vitamin D genes and in characterizing variants of large effects in known vitamin D genes. In Chapter 2, we identified a SNP in a known vitamin D locus, the 25-hydroxylase gene, with effects on vitamin D levels comparable to those of oral vitamin D supplementation. In Chapter 5, we identified tens of novel vitamin D genes. In Chapters 3 and 4, we used vitamin D genetic variants to interrogate the causal role of vitamin D in human diseases. Below, we discuss the strengths or shortcomings of each chapter.

Chapters 2 and 5 were both genome-wide association studies (GWAS) of 25OHD levels. They differ in their sample sizes and the described follow-up analyses. In Chapter 2, we meta-analyzed individuals from 19 European cohorts for a total of 42,326 participants. At the time, this was a sample size greater than any other GWAS of 25OHD, with the previous largest study reaching 33,996 participants[7]. The first strength of this study was the combination of our increased sample size with deeper imputation, which provided insight into variants with MAF below 5% and as low as 0.1%. The second strength, which will have important implications for the future aim described in the Appendix, was that we showed the presence of a variant in 5% of the general European population, with the largest effect on 25OHD levels ever described (-0.43 standard deviations on standardized log-transformed 25OHD per effect allele). Carrying one copy of this variant doubles the risk of vitamin D insufficiency and confers a 40% increase in the odds of multiple sclerosis,

while it might as well reduce the response to conventional vitamin D supplementation. If this hypothesis is confirmed by the pharmacogenetics trial described in Chapter 7, this might have important public health consequences, given that vitamin D insufficiency in carriers of this low-frequency variant can be easily treated using the active form of vitamin D (calcitriol).

In Chapter 5, by undertaking a GWAS on 401,460 UK Biobank participants, and combining these results to our previous GWAS described in Chapter 2, we maximized our yield in common, low-frequency and rare variants, and identified in total 69 25OHD loci, among which 63 are novel. These loci harbor 138 conditionally independent variants, among which 53 alleles with MAF <5% (average absolute effect of 0.23 standard deviations on standardized log-transformed 25OHD), and 85 alleles with MAF≥5% (average absolute effect 0.03 standard deviations on standardized log-transformed 25OHD). Of note, the *CYP2R1* variant described in Chapter 2 exerts the largest effect (-0.35 standard deviations on standardized log-transformed 25OHD per effect allele) upon 25OHD among all low-frequency variants in this larger and more recent GWAS. This study increased substantially our knowledge on the genetic architecture of vitamin D, and confirmed that 25OHD is a moderately polygenic trait, affected by genes inside and outside the canonical vitamin D metabolic pathway. Specifically, we described novel genes potentially involved in vitamin D synthesis in the skin (such as the *FLG* and *KRATP5*), and in vitamin D catabolism (*UGT1A4, SULT2A1*), as well as genes not directly related to vitamin D metabolism, notably genes involved in the lipid metabolism (such as the *LIPC, CETP, PCSK9, APOC1, LDLR*). These results are further supported by our *in silico* follow-up analyses using summary-level results of our GWAS,

which showed increased expression of vitamin D genes in skin, liver and gut, and enrichments in gene pathways involved in the lipid metabolism. However, *in vitro* and *in vivo* functional follow-up of these novel loci is required, to characterize the causal genes and/or mechanisms underlying the associations with 25OHD levels. Also, using LD score regression, we observed genetic correlation between 25OHD and lifestyle traits, which are controlled by genes not directly linked to 25OHD metabolism. We can speculate on the exact mechanism of action of these genes on 25OHD - for instance through their effect on time spend outdoors and consequently exposure to sunlight- but follow-up experiments are necessary to validate these hypotheses. Collectively these results suggest that serum levels of 25OHD are in crosstalk with a range of metabolic processes extending within and beyond the canonical vitamin D metabolic pathway. The findings of this GWAS highlight that the genetic instrument for vitamin D is instrumenting more than the vitamin D pathway, and specifically also captures variance in traits that relate to environmental confounders that could influence 25OHD levels. Taken together, these results present a cautionary tale for future MR studies, since there is a risk of pleiotropic effects for a substantial number of novel 25OHD-related SNPs mapping to genes not directly involved in 25OHD biology.

Both GWAS studies described in Chapters 2 and 5 have some limitations. We did not undertake any statistical fine-mapping of the newly identified loci, or *in vitro/ in vivo* functional follow-up experiments to identify causal variants. Although this should consist an aim for future studies, and will allow a better characterization of the role of the newly identified genes in vitamin D status, this exceeded the objectives of the present Thesis. Moreover, in regards to the *CYP2R1*

synonymous low-frequency variant described in Chapter 2, since this variant maps in the hepatic vitamin D hydroxylase, a gene with a well-established role in vitamin D synthesis, we judged that functional follow-up experiments were not necessary for its validation. Although we cannot unequivocally claim that this SNP is the causal variant in this locus, the fact that it replicated as the lead conditionally independent low-frequency variant in this locus in the larger GWAS in UK described in Chapter 5, supports this hypothesis.

In Chapters 3 and 4, we applied genomic discoveries from previous 25OHD GWAS, and data from the largest GWAS on 4 human diseases and traits, to test causal associations of low 25OHD with these outcomes in MR studies. This well-established study design typically overcomes the bias from confounding and reverse causation of the observational studies. Our MR experiments failed to provide evidence supporting a causal role of 25OHD in all 4 studied outcomes (coronary artery disease, asthma, atopic dermatitis, IgE levels), and exclude large effects (odds ratios > 1.3) of 25OHD on these outcomes. The interpretation of these results is that the observational associations of vitamin D status with these diseases are likely driven by environmental confounders, and vitamin D supplementation will likely not affect the risk of developing these chronic conditions. Contrarily, prevention efforts should rather target the environmental confounders that drive these associations. It is important to underline that these two Mendelian randomization studies were published before the GWASes of Chapter 2 and 5, and thus they do not include the novel variants described in these chapters. Although the results of these MR studies might have changed by including an enhanced set of genetic instruments, it is important to underline that the 4 common

variants from SUNLIGHT, which are used as instruments in these two MR papers, still explain a large part of the variance in 25OHD levels (2.4% vs 4.9% of all 138 conditionally independent variants identified in Chapter 5). Also, since these SNPs map directly in or near genes involved in vitamin D synthesis or metabolism, they minimize the risk of pleiotropy, a considerable limitation of the MR study design. Thus, we believe that these studies continue to provide valid evidence refuting causal effects of low vitamin D levels in the studied outcomes. It is also important to note that the results of the MR study on coronary artery disease presented in Chapter 3 were replicated in a large randomized controlled trial published two years later (Vitamin D and Omega-3 Trial-VITAL, N=25,871), which reported a hazard ratio very comparable to the MR estimate provided in our Circulation: Cardiovascular Genetics paper (hazard ratio 0.97, 95% CI 0.85-1.12 compared to an MR odds ratio of 0.99, 95% CI 0.94-1.17).

The MR studies described in Chapters 3 and 4, similar to other MR studies using 25OHD levels as an exposure, have some important limitations. First, they do not exclude effects of the active form of vitamin D ($1,25[OH]_2D$) in the serum and of its intracellular concentrations on the outcomes. Indeed, there is a poor correlation between 25OHD and $1,25[OH]_2D$ levels[95]. Nonetheless, measurement of 25OHD remains the gold standard for tracking response to vitamin D supplementation, and we showed in Chapter 2 that genetic effects on 25OHD levels are comparable in magnitude to the effects on 25OHD of oral vitamin D supplementation. Another limitation of the MR studies is that this design does not allow to interrogate non-linear effects, meaning that having 25OHD levels in the extremes of the normal distribution might affect the

studied outcomes. MR does, though, provide insight in effects of 25OHD within the normal distribution, which represents the majority of the population. Finally, the results of our MR studies are restricted to European populations, and cannot be generalized in other ancestries. A trans-ethnic GWAS for 25OHD, described in Chapter 7, might provide valid instruments to test causal associations between vitamin D and outcomes, for which available GWAS studies in other ancestries exist.

Chapter 7: Conclusion and Future Directions

This thesis was an exploration of the genetic determinants of vitamin D through the study of 25OHD levels. The obtained findings demonstrate the clinical utility of characterization of novel genetic associations, offer an improved understanding of the control of vitamin D, enable the identification of groups at increased risk for vitamin D insufficiency, and help to interrogate the causal role of low 25OHD levels in vitamin D related clinical outcomes.

As we demonstrated in Chapter 5, 25OHD level is a moderately polygenic trait with 69 associated genetic loci, among which loci including low-frequency and rare variants of large effects. Several future aims can be suggested to continue this work.

Chapter 5 can be continued with trans-ethnic analyses. We restricted our GWAS to either Europeans (Chapter 2) or White British participants in the UK Biobank (Chapter 5). GWAS of 25OHD can and should be expanded to other ethnicities, however previous efforts had limited statistical power in their discovery GWAS[96]. These studies signal that large trans-ethnic GWAS of 25OHD may be on the horizon, therefore as these data become available, they should be analyzed along with data from participants of various European ancestries from UK Biobank to identify even more genetic determinants of vitamin D.

The discovery of a low-frequency variant on *CYP2R1* with large effects on 25OHD levels described in Chapter 2 challenges the investigation of the impact of this variant on serum 25OHD in response to vitamin D oral supplementation. This is described in the proposal featuring in Appendix 3, which was approved as an ancillary study to the VITAL RCT[97] .

Hence, we have identified 63 novel loci for 25OHD, and demonstrating the moderate polygenicity of 25OHD suggests that larger GWAS (e.g. sample sizes exceeding 500,000) will continue to identify novel loci. Following our recent GWAS and future larger GWAS, we will need methods to identify target genes and to discern their function, since many of the novel genes are not directly involved in vitamin D synthesis and metabolism. Statistical fine-mapping of the candidate loci using FINEMAP, *in vitro* and *in vivo* functional exploration, and genome-editing through CRISPR-Cas9 are a few steps in this direction. This is a necessary follow-up of the results of our study, with potentially major clinical implications. Also, since information on vitamin D supplementation is available in UK Biobank (with ~25,000 individuals taking vitamin D supplements in the White-British subset of 440,345 individuals), gene-diet interaction studies could be well-powered to generate clinically relevant results.

Genetic prediction of vitamin D levels is another future aim, which can be enabled by the large sample size of the 25OHD GWAS of Chapter 5, and future larger 25OHD GWASes. Developing a polygenic risk score from 25OHD levels is important, since it will allow a more efficient identification and management of people at risk of vitamin D insufficiency and its complications. We have generated preliminary results on a 25OHD polygenic risk score, by using a machine learning approach (the LASSO regression), which we had developed from UK Biobank for ultrasound measured BMD (Forgetta et al, manuscript under review). To do so, we first used data from UK Biobank for training of an algorithm. Specifically, we selected to include various numbers of SNPs using 3 different p-value cut-offs from a GWAS performed on 80% of the White British participants included in the GWAS of Chapter 5. We trained seven algorithms in total,

using the various sets of SNPs, with and without covariates (age and sex). We then tested the performance of these algorithms in a separate validation subset from UK Biobank (a 12.5% among the remaining 20% of the White British participants). The performance of our polygenic risk score was consistently low, with a maximum r^2 of 4.2% (obtained using a p-value cut-off of 5 x10^{-6} and the including the covariates), which was significantly lower than the 25OHD SNP heritability of 16% described in Chapter 5. We therefore decided not to proceed with testing this genetic predictor in an independent cohort. Although we can speculate on various reasons behind the low performance of our polygenic risk score, we believe that the high representation of low-frequency and rare variants in our sets of 25OHD SNPs, increases the "noise" in the lasso regression, since the real effect of these SNPs on 25OHD could be far from the one estimated in the GWAS. Thus, our genetic predictor becomes less accurate. This problem can be overcome by increasing the sample sizes in future 25OHD GWAS.

Finally, the identification of novel genetic determinants of 25OHD in Chapter 5 offers an enhanced set of instruments for vitamin D. To date, more than 70 MR studies have been published assessing the causal role of vitamin D in human diseases and traits[98]. Nevertheless, given the large body of evidence from observational studies linking vitamin D to a plethora of health outcomes, and the shortage of high-quality vitamin D RCTs, we foresee that the newly identified 25OHD SNPs (explaining a larger portion of the variance in 25OHD levels), will serve as instruments in future MR experiments. These MR studies will focus on outcomes not assessed in previous vitamin D MR papers, or will revisit the causal role of vitamin D in previously studied outcomes through MR. At this point, it is important to highlight that many of these newly identified SNPs lie in genes

not directly related to the vitamin D metabolic pathway. Some of these variants pertain to lifestyle traits that could affect vitamin D levels indirectly, for instance through the time spent outdoors. The risk of pleiotropy in MR studies using these variants as instruments is obvious, since some of these traits (e.g. alcohol consumption, educational attainment) represent typical environmental confounders, which bias observational associations with vitamin D. We therefore plan to investigate the hypothesis that, expanding the variants for a metabolite, such as vitamin D, beyond the classic ones directly involved in its metabolic pathway, might lead to a set of instruments which are less reliable and more likely to be invalid.

In conclusion, the findings presented this thesis represent novel contributions to the understanding of the genetic determinants of vitamin D in humans and a stepping stone for further research on vitamin D genomics and their clinical applications.

Chapter 8: Master Reference List

1. Forrest, K.Y., and Stuhldreher, W.L. (2011). Prevalence and correlates of vitamin D deficiency in US adults. Nutr Res 31, 48-54.

2. Bouillon, R., Marcocci, C., Carmeliet, G., Bikle, D., White, J.H., Dawson-Hughes, B., Lips, P., Munns, C.F., Lazaretti-Castro, M., Giustina, A., et al. (2018). Skeletal and extra-skeletal actions of vitamin D: Current evidence and outstanding questions. Endocr Rev.

3. Autier, P., and Gandini, S. (2007). Vitamin D supplementation and total mortality: a meta-analysis of randomized controlled trials. Arch Intern Med 167, 1730-1737.

4. Shea, M.K., Benjamin, E.J., Dupuis, J., Massaro, J.M., Jacques, P.F., D'Agostino, R.B., Sr., Ordovas, J.M., O'Donnell, C.J., Dawson-Hughes, B., Vasan, R.S., et al. (2009). Genetic and non-genetic correlates of vitamins K and D. Eur J Clin Nutr 63, 458-464.

5. Livshits, G., Karasik, D., and Seibel, M.J. (1999). Statistical genetic analysis of plasma levels of vitamin D: familial study. Ann Hum Genet 63, 429-439.

6. Hunter, D., De Lange, M., Snieder, H., MacGregor, A.J., Swaminathan, R., Thakker, R.V., and Spector, T.D. (2001). Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. J Bone Miner Res 16, 371-378.

7. Wang, T.J., Zhang, F., Richards, J.B., Kestenbaum, B., van Meurs, J.B., Berry, D., Kiel, D.P., Streeten, E.A., Ohlsson, C., Koller, D.L., et al. (2010). Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet 376, 180-188.

8. Ahn, J., Yu, K., Stolzenberg-Solomon, R., Simon, K.C., McCullough, M.L., Gallicchio, L., Jacobs, E.J., Ascherio, A., Helzlsouer, K., Jacobs, K.B., et al. (2010). Genome-wide association study of circulating vitamin D levels. Hum Mol Genet 19, 2739-2745.

9. Jiang, X., O'Reilly, P.F., Aschard, H., Hsu, Y.H., Richards, J.B., Dupuis, J., Ingelsson, E., Karasik, D., Pilz, S., Berry, D., et al. (2018). Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. Nat Commun 9, 260.

10. Gratten, J., Wray, N.R., Keller, M.C., and Visscher, P.M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nat Neurosci 17, 782-790.

11. Zheng, H.F., Forgetta, V., Hsu, Y.H., Estrada, K., Rosello-Diez, A., Leo, P.J., Dahia, C.L., Park-Min, K.H., Tobias, J.H., Kooperberg, C., et al. (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature 526, 112-117.

12. Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56-65.

13. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., Y, A.R., H, K.F., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 48, 1443-1448.

14. Mokry, L.E., Ross, S., Ahmad, O.S., Forgetta, V., Smith, G.D., Goltzman, D., Leong, A., Greenwood, C.M., Thanassoulis, G., and Richards, J.B. (2015). Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLoS Med 12, e1001866.

15. Rhead, B., Baarnhielm, M., Gianfrancesco, M., Mok, A., Shao, X., Quach, H., Shen, L., Schaefer, C., Link, J., Gyllenberg, A., et al. (2016). Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. Neurol Genet 2, e97.

16. (1999). Vitamin D supplement in early childhood and risk for Type I (insulin-dependent) diabetes mellitus. The EURODIAB Substudy 2 Study Group. Diabetologia 42, 51-54.

17. Bierschenk, L., Alexander, J., Wasserfall, C., Haller, M., Schatz, D., and Atkinson, M. (2009). Vitamin D levels in subjects with and without type 1 diabetes residing in a solar rich environment. Diabetes Care 32, 1977-1979.

18. Feldman, D., Krishnan, A.V., Swami, S., Giovannucci, E., and Feldman, B.J. (2014). The role of vitamin D in reducing cancer risk and progression. Nat Rev Cancer 14, 342-357.

19. Martineau, A.R., Jolliffe, D.A., Hooper, R.L., Greenberg, L., Aloia, J.F., Bergman, P., Dubnov-Raz, G., Esposito, S., Ganmaa, D., Ginde, A.A., et al. (2017). Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data. BMJ 356, i6583.

20. Pilz, S., Verheyen, N., Grubler, M.R., Tomaschitz, A., and Marz, W. (2016). Vitamin D and cardiovascular disease prevention. Nat Rev Cardiol 13, 404-417.

21. Jia, F., Wang, B., Shan, L., Xu, Z., Staal, W.G., and Du, L. (2015). Core symptoms of autism improved after vitamin D supplementation. Pediatrics 135, e196-198.

22. Mazahery, H., Conlon, C.A., Beck, K.L., Mugridge, O., Kruger, M.C., Stonehouse, W., Camargo, C.A., Jr., Meyer, B.J., Tsang, B., Jones, B., et al. (2019). A Randomised-Controlled Trial of Vitamin D and Omega-3 Long Chain Polyunsaturated Fatty Acids in the Treatment of Core Symptoms of Autism Spectrum Disorder in Children. J Autism Dev Disord 49, 1778-1794.

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

24. Scragg, R. (2018). Emerging Evidence of Thresholds for Beneficial Effects from Vitamin D Supplementation. Nutrients 10.

25. Bolland, M.J., Grey, A., and Avenell, A. (2018). Effects of vitamin D supplementation on musculoskeletal health: a systematic review, meta-analysis, and trial sequential analysis. Lancet Diabetes Endocrinol 6, 847-858.

26. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 37, 658-665.

27. Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N., and Davey Smith, G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 27, 1133-1163.

28. Smith, G.D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32, 1-22.

29. Larsson, S.C., Traylor, M., Mishra, A., Howson, J.M.M., Michaelsson, K., Markus, H.S., and Consortium, M.P.o.t.I.S.G. (2018). Serum 25-Hydroxyvitamin D Concentrations and Ischemic Stroke and Its Subtypes. Stroke 49, 2508-2511.

30. He, Y., Timofeeva, M., Farrington, S.M., Vaughan-Shaw, P., Svinti, V., Walker, M., Zgaga, L., Meng, X., Li, X., Spiliopoulou, A., et al. (2018). Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: a large Mendelian randomisation study. BMC Med 16, 142.

31. Aspelund, T., Grubler, M.R., Smith, A.V., Gudmundsson, E.F., Keppel, M., Cotch, M.F., Harris, T.B., Jorde, R., Grimnes, G., Joakimsen, R., et al. (2019). Effect of Genetically Low 25-Hydroxyvitamin D on Mortality Risk: Mendelian Randomization Analysis in 3 Large European Cohorts. Nutrients 11.

32. Michaelsson, K., Melhus, H., and Larsson, S.C. (2018). Serum 25-Hydroxyvitamin D Concentrations and Major Depression: A Mendelian Randomization Study. Nutrients 10.

33. Bowman, K., Jones, L., Pilling, L.C., Delgado, J., Kuchel, G.A., Ferrucci, L., Fortinsky, R.H., and Melzer, D. (2019). Vitamin D levels and risk of delirium: A mendelian randomization study in the UK Biobank. Neurology 92, e1387-e1394.

34. Lund-Nielsen, J., Vedel-Krogh, S., Kobylecki, C.J., Brynskov, J., Afzal, S., and Nordestgaard, B.G. (2018). Vitamin D and Inflammatory Bowel Disease: Mendelian Randomization Analyses in the Copenhagen Studies and UK Biobank. J Clin Endocrinol Metab 103, 3267-3277.

35. Sun, J.Y., Zhao, M., Hou, Y., Zhang, C., Oh, J., Sun, Z., and Sun, B.L. (2019). Circulating serum vitamin D levels and total body bone mineral density: A Mendelian randomization study. J Cell Mol Med 23, 2268-2271.

36. Jiang, X., Dimou, N.L., Al-Dabhani, K., Lewis, S.J., Martin, R.M., Haycock, P.C., Gunter, M.J., Key, T.J., Eeles, R.A., Muir, K., et al. (2018). Circulating vitamin D concentrations and risk of breast and prostate cancer: a Mendelian randomization study. Int J Epidemiol.

37. Yarmolinsky J, R.C., Lophatananon A, et al. . (November 2018:472696.). Evaluating causal associations between previously reported risk factors and epithelial ovarian cancer: a Mendelian randomization analysis. . BioRxiv.

38. Mai, X.M., Videm, V., Sheehan, N.A., Chen, Y., Langhammer, A., and Sun, Y.Q. (2019). Potential causal associations of serum 25-hydroxyvitamin D with lipids: a Mendelian randomization approach of the HUNT study. Eur J Epidemiol 34, 57-66.

39. Dong, J., Gharahkhani, P., Chow, W.H., Gammon, M.D., Liu, G., Caldas, C., Wu, A.H., Ye, W., Onstad, L., Anderson, L.A., et al. (2019). No Association Between Vitamin D Status and Risk of Barrett's Esophagus or Esophageal Adenocarcinoma: A Mendelian Randomization Study. Clin Gastroenterol Hepatol.

40. Tan, V.Y., Biernacka, K.M., Dudding, T., Bonilla, C., Gilbert, R., Kaplan, R.C., Qibin, Q., Teumer, A., Martin, R.M., Perks, C.M., et al. (2018). Reassessing the Association between Circulating Vitamin D and IGFBP-3: Observational and Mendelian Randomization Estimates from Independent Sources. Cancer Epidemiol Biomarkers Prev 27, 1462-1471.

41. Havdahl, A., Mitchell, R., Paternoster, L., and Davey Smith, G. (2019). Investigating causality in the association between vitamin D status and self-reported tiredness. Sci Rep 9, 2880.

42. Yuri Milaneschi, W.J.P., Michel G Nivard, Hamdi Mbarek, Dorret I Boomsma, Brenda WJH Penninx. (2019). A role for vitamin D and omega-3 fatty acids in major depression? An exploration using genomics. BioRxiv.

43. Libuda, L., Laabs, B.H., Ludwig, C., Buhlmeier, J., Antel, J., Hinney, A., Naaresh, R., Focker, M., Hebebrand, J., Konig, I.R., et al. (2019). Vitamin D and the Risk of Depression: A Causal Relationship? Findings from a Mendelian Randomization Study. Nutrients 11.

44. Liyanage, U.E., Law, M.H., Melanoma Meta-analysis, C., Barrett, J.H., Iles, M.M., and MacGregor, S. (2019). Is there a causal relationship between vitamin D and melanoma risk? : A Mendelian randomization study. Br J Dermatol.

45. Man, L., Zhang, Z., Zhang, M., Zhang, Y., Li, J., Zheng, N., Cao, Y., Chi, M., Chao, Y., Huang, Q., et al. (2015). Association between vitamin D deficiency and insufficiency and

the risk of childhood asthma: evidence from a meta-analysis. Int J Clin Exp Med 8, 5699-5706.

46. Cassim, R., Russell, M.A., Lodge, C.J., Lowe, A.J., Koplin, J.J., and Dharmage, S.C. (2015). The role of circulating 25 hydroxyvitamin D in asthma: a systematic review. Allergy 70, 339-354.

47. Robl, R., Uber, M., Abagge, K.T., Lima, M.N., and Carvalho, V.O. (2016). Serum Vitamin D Levels Not Associated with Atopic Dermatitis Severity. Pediatr Dermatol.

48. Debinska, A., Sikorska-Szaflik, H., Urbanik, M., and Boznanski, A. (2015). The role of vitamin D in atopic dermatitis. Dermatitis 26, 155-161.

49. Cheng, H.M., Kim, S., Park, G.H., Chang, S.E., Bang, S., Won, C.H., Lee, M.W., Choi, J.H., and Moon, K.C. (2014). Low vitamin D levels are associated with atopic dermatitis, but not allergic rhinitis, asthma, or IgE sensitization, in the adult Korean population. J Allergy Clin Immunol 133, 1048-1055.

50. Wang, S.S., Hon, K.L., Kong, A.P., Pong, H.N., Wong, G.W., and Leung, T.F. (2014). Vitamin D deficiency is associated with diagnosis and severity of childhood atopic dermatitis. Pediatr Allergy Immunol 25, 30-35.

51. Hypponen, E., Berry, D.J., Wjst, M., and Power, C. (2009). Serum 25-hydroxyvitamin D and IgE - a significant but nonlinear relationship. Allergy 64, 613-620.

52. Kang, J.W., Kim, J.H., Kim, H.J., Lee, J.G., Yoon, J.H., and Kim, C.H. (2016). Association of serum 25-hydroxyvitamin D with serum IgE levels in Korean adults. Auris Nasus Larynx 43, 84-88.

53. Dogru, M., Kirmizibekmez, H., Yesiltepe Mutlu, R.G., Aktas, A., and Ozturkmen, S. (2014). Clinical effects of vitamin D in children with asthma. Int Arch Allergy Immunol 164, 319-325.

54. Searing, D.A., and Leung, D.Y. (2010). Vitamin D in atopic dermatitis, asthma and allergic diseases. Immunol Allergy Clin North Am 30, 397-409.

55. Arshi, S., Fallahpour, M., Nabavi, M., Bemanian, M.H., Javad-Mousavi, S.A., Nojomi, M., Esmaeilzadeh, H., Molatefi, R., Rekabi, M., Jalali, F., et al. (2014). The effects of vitamin D supplementation on airway functions in mild to moderate persistent asthma. Ann Allergy Asthma Immunol 113, 404-409.

56. Castro, M., King, T.S., Kunselman, S.J., Cabana, M.D., Denlinger, L., Holguin, F., Kazani, S.D., Moore, W.C., Moy, J., Sorkness, C.A., et al. (2014). Effect of vitamin D3 on asthma treatment failures in adults with symptomatic asthma and lower vitamin D levels: the VIDA randomized clinical trial. JAMA 311, 2083-2091.

57. Martineau, A.R., Cates, C.J., Urashima, M., Jensen, M., Griffiths, A.P., Nurmatov, U., Sheikh, A., and Griffiths, C.J. (2016). Vitamin D for the management of asthma. Cochrane Database Syst Rev 9, CD011511.

58. Riverin, B.D., Maguire, J.L., and Li, P. (2015). Vitamin D Supplementation for Childhood Asthma: A Systematic Review and Meta-Analysis. PLoS One 10, e0136841.

59. Luo, J., Liu, D., and Liu, C.T. (2015). Can Vitamin D Supplementation in Addition to Asthma Controllers Improve Clinical Outcomes in Patients With Asthma?: A Meta-Analysis. Medicine (Baltimore) 94, e2185.

60. Kim, G., and Bae, J.H. (2016). Vitamin D and atopic dermatitis: A systematic review and meta-analysis. Nutrition.

61. Bath-Hextall, F.J., Jenkinson, C., Humphreys, R., and Williams, H.C. (2012). Dietary supplements for established atopic eczema. Cochrane Database Syst Rev 2, CD005205.

62. Hata, T.R., Audish, D., Kotol, P., Coda, A., Kabigting, F., Miller, J., Alexandrescu, D., Boguniewicz, M., Taylor, P., Aertker, L., et al. (2014). A randomized controlled double-blind investigation of the effects of vitamin D dietary supplementation in subjects with atopic dermatitis. J Eur Acad Dermatol Venereol 28, 781-789.

63. Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., von Mutius, E., Farrall, M., Lathrop, M., Cookson, W.O., et al. (2010). A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 363, 1211-1221.

64. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). 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.

65. Paternoster, L., Standl, M., Chen, C.M., Ramasamy, A., Bonnelykke, K., Duijts, L., Ferreira, M.A., Alves, A.C., Thyssen, J.P., Albrecht, E., et al. (2012). Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet 44, 187-192.

66. Faye, L.L., Sun, L., Dimitromanolakis, A., and Bull, S.B. (2011). A flexible genome-wide bootstrap method that accounts for ranking and threshold-selection bias in GWAS interpretation and replication study design. Stat Med 30, 1898-1912.

67. Ollier, W., Sprosen, T., and Peakman, T. (2005). UK Biobank: from concept to reality. Pharmacogenomics 6, 639-646.

68. Hollams, E.M., Hart, P.H., Holt, B.J., Serralha, M., Parsons, F., de Klerk, N.H., Zhang, G., Sly, P.D., and Holt, P.G. (2011). Vitamin D and atopy and asthma phenotypes in children: a longitudinal cohort study. Eur Respir J 38, 1320-1327.

69. Goleva, E., Searing, D.A., Jackson, L.P., Richers, B.N., and Leung, D.Y. (2012). Steroid requirements and immune associations with vitamin D are stronger in children than adults with asthma. J Allergy Clin Immunol 129, 1243-1251.

70. Vimaleswaran, K.S., Berry, D.J., Lu, C., Tikkanen, E., Pilz, S., Hiraki, L.T., Cooper, J.D., Dastani, Z., Li, R., Houston, D.K., et al. (2013). Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. PLoS Med 10, e1001383.

71. Higgins, J.P., Thompson, S.G., Deeks, J.J., and Altman, D.G. (2003). Measuring inconsistency in meta-analyses. BMJ 327, 557-560.

72. Greco, M.F., Minelli, C., Sheehan, N.A., and Thompson, J.R. (2015). Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat Med 34, 2926-2940.

73. Brion, M.J., Shakhbazov, K., and Visscher, P.M. (2013). Calculating statistical power in Mendelian randomization studies. Int J Epidemiol 42, 1497-1501.

74. Dastani, Z., Li, R., and Richards, B. (2013). Genetic regulation of vitamin D levels. Calcif Tissue Int 92, 106-117.

75. Thomsen, S.F. (2015). Epidemiology and natural history of atopic diseases. Eur Clin Respir J 2.

76. Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I., et al. (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 44, 659-669.

77. Tobacco, and Genetics, C. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 42, 441-447.

78. Wang, S.S., Hon, K.L., Kong, A.P., Tang, M.F., Sy, H.Y., Chan, J.C., and Leung, T.F. (2014). Eczema phenotypes are associated with multiple vitamin D pathway genes in Chinese children. Allergy 69, 118-124.

79. Silverberg, J.I. (2014). Atopic dermatitis. JAMA Dermatol 150, 1380.

80. Mokry, L.E., Ross, S., Ahmad, O.S., Forgetta, V., Smith, G.D., Leong, A., Greenwood, C.M., Thanassoulis, G., and Richards, J.B. (2015). Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLoS Med 12, e1001866.

81. Brehm, J.M., Celedon, J.C., Soto-Quiros, M.E., Avila, L., Hunninghake, G.M., Forno, E., Laskey, D., Sylvia, J.S., Hollis, B.W., Weiss, S.T., et al. (2009). Serum vitamin D levels and markers of severity of childhood asthma in Costa Rica. Am J Respir Crit Care Med 179, 765-771.

82. Granell, R., Henderson, A.J., Evans, D.M., Smith, G.D., Ness, A.R., Lewis, S., Palmer, T.M., and Sterne, J.A. (2014). Effects of BMI, fat mass, and lean mass on asthma in childhood: a Mendelian randomization study. PLoS Med 11, e1001669.

83. Vimaleswaran, K.S., Cavadino, A., Berry, D.J., LifeLines Cohort Study, i., Jorde, R., Dieffenbach, A.K., Lu, C., Alves, A.C., Heerspink, H.J., Tikkanen, E., et al. (2014). Association of vitamin D status with arterial blood pressure and hypertension risk: a mendelian randomisation study. Lancet Diabetes Endocrinol 2, 719-729.

84. Prentice, A.M., and Jebb, S.A. (2001). Beyond body mass index. Obes Rev 2, 141-147.

85. Lochte, L., Nielsen, K.G., Petersen, P.E., and Platts-Mills, T.A. (2016). Childhood asthma and physical activity: a systematic review with meta-analysis and Graphic Appraisal Tool for Epidemiology assessment. BMC Pediatr 16, 50.

86. Searing, D.A., Zhang, Y., Murphy, J.R., Hauk, P.J., Goleva, E., and Leung, D.Y. (2010). Decreased serum vitamin D levels in children with asthma are associated with increased corticosteroid use. J Allergy Clin Immunol 125, 995-1000.

87. Wang, S. (2009). Epidemiology of vitamin D in health and disease. Nutr Res Rev 22, 188-203.

88. Bosse, Y., Lemire, M., Poon, A.H., Daley, D., He, J.Q., Sandford, A., White, J.H., James, A.L., Musk, A.W., Palmer, L.J., et al. (2009). Asthma and genes encoding components of the vitamin D pathway. Respir Res 10, 98.

89. Papadopoulou, A., Kouis, P., Middleton, N., Kolokotroni, O., Karpathios, T., Nicolaidou, P., and Yiallouros, P.K. (2015). Association of vitamin D receptor gene polymorphisms and vitamin D levels with asthma and atopy in Cypriot adolescents: a case-control study. Multidiscip Respir Med 10, 26.

90. Leung, T.F., Wang, S.S., Tang, M.F., Kong, A.P., Sy, H.Y., Hon, K.L., Chan, J.C., and Wong, G.W. (2015). Childhood asthma and spirometric indices are associated with polymorphic markers of two vitamin D 25-hydroxylase genes. Pediatr Allergy Immunol 26, 375-382.

91. Hallau, J., Hamann, L., Schumann, R.R., Worm, M., and Heine, G. (2016). A Promoter Polymorphism of the Vitamin D Metabolism Gene Cyp24a1 is Associated with Severe Atopic Dermatitis in Adults. Acta Derm Venereol 96, 169-172.

92. Suzuki, H., Makino, Y., Nagata, M., Furuta, J., Enomoto, H., Hirota, T., Tamari, M., and Noguchi, E. (2016). A rare variant in CYP27A1 and its association with atopic dermatitis with high serum total IgE. Allergy 71, 1486-1489.

93. Shaheen, S.O., Rutterford, C., Zuccolo, L., Ring, S.M., Davey Smith, G., Holloway, J.W., and Henderson, A.J. (2014). Prenatal alcohol exposure and childhood atopic disease: a Mendelian randomization approach. J Allergy Clin Immunol 133, 225-232 e221-225.

94. Hysinger, E.B., Roizen, J.D., Mentch, F.D., Vazquez, L., Connolly, J.J., Bradfield, J.P., Almoguera, B., Sleiman, P.M., Allen, J.L., Levine, M.A., et al. (2016). Mendelian randomization analysis demonstrates that low vitamin D is unlikely causative for pediatric asthma. J Allergy Clin Immunol.

95. Zittermann, A., Schleithoff, S.S., Frisch, S., Gotting, C., Kuhn, J., Koertke, H., Kleesiek, K., Tenderich, G., and Koerfer, R. (2009). Circulating calcitriol concentrations and total mortality. Clin Chem 55, 1163-1170.

96. Hong, J., Hatchell, K.E., Bradfield, J.P., Bjonnes, A., Chesi, A., Lai, C.Q., Langefeld, C.D., Lu, L., Lu, Y., Lutsey, P.L., et al. (2018). Transethnic Evaluation Identifies Low-Frequency Loci Associated With 25-Hydroxyvitamin D Concentrations. J Clin Endocrinol Metab 103, 1380-1392.

97. Manson, J.E., Cook, N.R., Lee, I.M., Christen, W., Bassuk, S.S., Mora, S., Gibson, H., Gordon, D., Copeland, T., D'Agostino, D., et al. (2019). Vitamin D Supplements and Prevention of Cancer and Cardiovascular Disease. N Engl J Med 380, 33-44.

98. Xiangrui Meng, X.L., Maria Timofeeva, Yazhou He, Athina, and Spiliopoulou, W.-Q.W., Aliya Gifford, Hongjiang Wu, Timothy Varley, Peter Joshi,Joshua C. Denny, Susan Farrington, Lina Zgaga, Malcolm G. Dunlop, Paul McKeigue, Harry Campbell, Evropi Theodoratou. (2019). Phenome Wide Mendelian Randomization Study of genetically determined Vitamin D on multiple health outcomes using the UK Biobank Study. Int J Epidemiol In press.

Appendices

Appendix 1: Copyright Permissions

As per Elsevier policies, the article presented in Chapter 2, including all Figures and Tables, is included in full in this thesis for non-commercial purposes. Digital object identifier for the original work is provided here: https://doi.org/10.1016/j.ajhg.2017.06.014

The article presented in Chapter 3, including all Figures and Tables, is an open access article reproduced in this thesis for non-commercial purposes, with the original author and source credited, as per the policies of the Creative Commons Attribution License. Digital object identifier for the original works is provided again here: https://doi.org/10.1161/CIRCGENETICS.116.001396

The content presented in Chapter 4, including all Figures and Tables, is an open access article reproduced in this thesis for non-commercial purposes, with the original author and source credited as per the policies of the Creative Commons Attribution License. Digital object identifier for the original works is provided again here:

https://dx.doi.org/10.1371%2Fjournal.pmed.1002294

Appendix 2: Ethics Approval

For Chapters 2, to 5, written and informed consent was obtained for each participant and was approved by each participating sites' regional ethical review board.

Appendix 3: Pharmacogenetics study proposal

Title of the proposal: **Pharmacogenetic study exploring the impact of a genetic variant in the *CYP2R1* gene on the response to vitamin D replacement therapy**

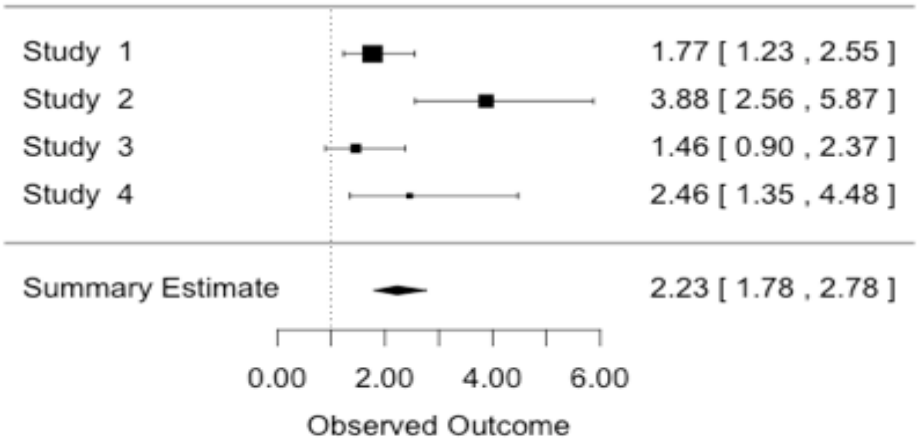
Despoina Manousaki, MD, PhD Candidate, McGill University & Brent Richards, MD, MSc, Professor of Medicine, McGill University

Introduction: A recent Genome-Wide Association Study (GWAS)[1] has shown that a novel low-frequency genetic variant in a gene called *CYP2R1* confers a substantial decrease in the blood levels of 25-hydroxy-vitamin D (25OHD), a biomarker of vitamin D status in humans, and is associated with a two-fold increase in their risk of vitamin D insufficiency[1]. We hypothesize that this genetic variant blocks the hepatic step of hydroxylation of vitamin D, one of the steps in the production of vitamin D's active form. In order to test this hypothesis, we will measure the response in 25OHD levels to oral vitamin D supplementation with cholecalciferol of individuals carrying this variant compared to non-carriers.

Background: Vitamin D deficiency affects up to 50% of otherwise healthy adults with potential public health consequences[2]. Almost half of the variability in the levels of 25OHD has been attributed to genetic factors [3; 4]. In recent years, multiple GWAS of 25OHD levels have been conducted on participants of Europeans ancestry with the largest comprising of 79,366 individuals and have identified six common genetic variants (minor allele frequency (MAF) >5%)[8–11]. In 2017, we led a large meta-analysis of 19 GWAS studies (N=42,326 individuals of European ancestry)[1],

aiming specifically to test the effect of rare and low frequency genetic variants (MAF<5%) on 25OHD levels. This study identified a low-frequency variant (rs117913124, minor allele frequency =2.5%) in a gene coding region of *CYP2R1* (which controls hepatic hydroxylation of cholecalciferol, the form of vitamin D taken by diet or synthetized in the skin), which confers a much larger effect on 25OHD level than the common variant already described in the same locus, and its effect is independent of the effect of any known common vitamin D variant. By analyzing 8,711 individuals from four of the participating studies, it was demonstrated that carriers of one allele of this low-frequency variant have an increased risk of vitamin D insufficiency (OR=2.2, 95% CI 1.8-2.8, P-value=1.25 x 10^{-12}) (Figure 1). Also, carrying one copy of this variant increases significantly the risk of multiple sclerosis, (OR=1.4, 95%CI 1.19-1.64, P=2.63 x 10^{-5}), a disease shown to be strongly associated with low 25OHD levels[5].

Figure 1: Forest plot with the odds ratios of vitamin D insufficiency in carriers of the *CYP2R1* variant from 4 studies, as well as the meta-analytic odds ratio.



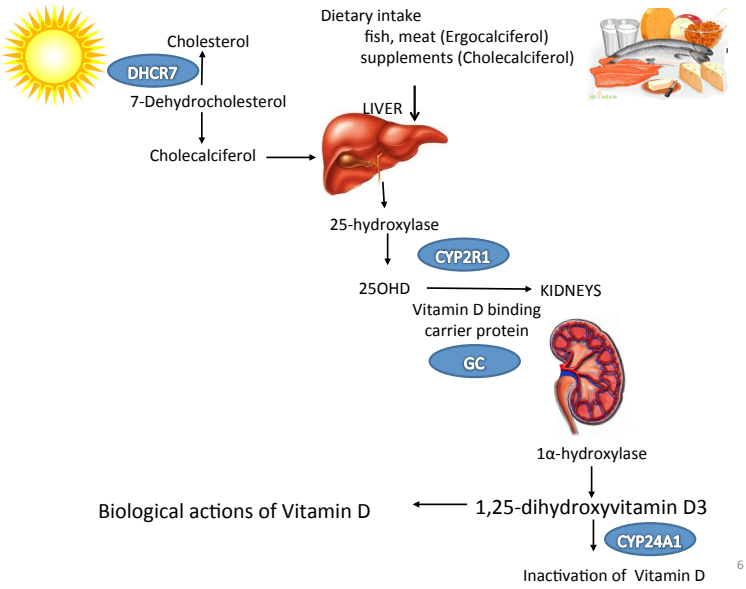


Figure 2: The vitamin D pathway.

CYP2R1 encodes one of the 25 hydoxylases that transform cholecalciferol, (the form of vitamin D produced in the skin after sun exposure or taken from the diet) to 25OHD. Thus *CYP2R1* is involved in a major step of 25OHD synthesis (**Figure 2**). 25OHD is then further metabolized to 1,25 dihydroxyvitamin D, the biological active form of vitamin D.

Given that: 1) the *CYP2R1* encodes the major 25-hydroxylase for vitamin D6 and 2) hepatic hydroxylation is a necessary step in the conversion of dietary vitamin D to its active form, we hypothesize that individuals heterozygous or homozygous for the *CYP2R1* variant are likely to have less of an increase in 25OHD after administration of cholecalciferol, than individuals without this genetic variant. This is because these individuals have a coding genetic variant in a key enzyme, *CYP2R1*, that is associated with lowered 25OHD level and impairment of this enzyme

would lead to less conversion of dietary vitamin D to 25OHD. Given the large effect of this variant on multiple sclerosis, restoring normal 25OHD levels in the carriers could be crucial in preventing this debilitating disease in those individuals. This study may have direct clinical relevance, since carriers of the *CYP2R1* variant have a low level of vitamin D and efforts to increase their 25OHD level are likely to be hindered by this genetic variant. Consequently, these individuals should have their low vitamin D levels treated with the active form of the vitamin D pathway, 1,25 dihydroxyvitamin D, rather than cholecalciferol, which is the current standard of care. Therapy with 1,25 hydroxyvitamin D would enable individuals with the genetic variant to bypass the *CYP2R1* and have the biologic effect of the active form of vitamin D. Importantly this genetic variant affects ~5% of individuals of European descent, which could influence the clinical care of hundreds of thousands of Canadians.

The recent VITAL NEJM article[7] described, in a subgroup of 836 participants taking vitamin D with repeated 25OHD measurements, a change in 25OHD levels of 30 nmol/l at 1 year. If the standard deviation of this delta is 25nmol/l (or below), and assuming that the delta 25OHD of the *CYP2R1* variant carriers is 15 nmol/L instead of 30 nmol/L, (a reasonable assumption, given its profound effect on 25OHD levels), a sample size of 44 *CYP2R1* variant carriers and 44 non-carriers gives our study a statistical power of at least 80%. Clearly, the true effect size of this variant is not known, but this is, of course, the reason why we are doing this study. Based on the above estimates, 40 to 50 participants per group (carriers /non-carriers) could be considered as a satisfying sample size for this study. We estimate that in a sample of 836 VITAL participants, there are 42 carriers

of the *CYP2R1* variant. Therefore, by genotyping the *CYP2R1* low-frequency variant in 836 VITAL participants on daily cholecalciferol, with available baseline and repeated 25OHD measurements, this study is powered to evaluate if individuals carrying this variant do not respond as well to oral cholecalciferol compared to non-carriers. If this is shown, these individuals should likely be treated with the active form of vitamin D (1,25 hydroxyvitamin D).

References

1. Manousaki, D., Dudding, T., Haworth, S., Hsu, Y.H., Liu, C.T., Medina-Gomez, C., Voortman, T., van der Velde, N., Melhus, H., Robinson-Cohen, C., et al. (2017). Low-Frequency Synonymous Coding Variation in CYP2R1 Has Large Effects on Vitamin D Levels and Risk of Multiple Sclerosis. Am J Hum Genet 101, 227-238.

2. Holick, M.F. (2007). Vitamin D deficiency. N Engl J Med 357, 266-281.

3. Shea, M.K., Benjamin, E.J., Dupuis, J., Massaro, J.M., Jacques, P.F., D'Agostino, R.B., Sr., Ordovas, J.M., O'Donnell, C.J., Dawson-Hughes, B., Vasan, R.S., et al. (2009). Genetic and non-genetic correlates of vitamins K and D. Eur J Clin Nutr 63, 458-464.

4. Livshits, G., Karasik, D., and Seibel, M.J. (1999). Statistical genetic analysis of plasma levels of vitamin D: familial study. Ann Hum Genet 63, 429-439.

5. Mokry, L.E., Ross, S., Ahmad, O.S., Forgetta, V., Smith, G.D., Goltzman, D., Leong, A., Greenwood, C.M., Thanassoulis, G., and Richards, J.B. (2015). Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLoS Med 12, e1001866.

6. Cheng, J.B., Levine, M.A., Bell, N.H., Mangelsdorf, D.J., and Russell, D.W. (2004). Genetic

evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. Proc Natl Acad Sci U S A 101, 7711-7715.

7. Manson, J.E., Cook, N.R., Lee, I.M., Christen, W., Bassuk, S.S., Mora, S., Gibson, H., Gordon, D., Copeland, T., D'Agostino, D., et al. (2018). Vitamin D Supplements and Prevention of Cancer and Cardiovascular Disease. N Engl J Med.

Appendix 4: Significant Contributions by the Author to Other Projects

Invited Editorials

Manousaki D, Richards JB. Role of vitamin D in disease through the lens of Mendelian randomization-Evidence from Mendelian randomization challenges the benefits of vitamin D supplementation for disease prevention.Int J Epidemiol. 2019. 1;48(5):1435-1437. doi: 10.1093/ije/dyz183. PMID: 31518416

Manousaki D, Richards JB. Vitamin D deficiency is an etiologic factor for multiple sclerosis Yes. *Multiple Sclerosis Journal* 2018. PMID: 30499750

Manousaki D, Richards JB Low Vitamin D levels as a risk factor for cancer. *BMJ*. 2017. Oct 31;359:j4952. doi: 10.1136/bmj.j4952. PMID: 29089329

Peer-Reviewed Publications

*denotes equal contribution

Despoina Manousaki, Tracie Barnett, Marie-Eve Mathieu, Katerina Maximova, Gabrielle Simoneau, Andrea Benedetti, Jennifer McGrath, Mélanie Henderson. (2018). Tune out and turn in: the influence of television viewing and sleep on lipid profiles in children. International Journal of Obesity (in press)

Vincenzo Forgetta, **Despoina Manousaki,** Roman Istomine, Stephanie Ross, Marie-Catherine Tessier, Luc Marchand, Hui Qi Qu, Jonathan P. Bradfield, Struan F. A. Grant, Hakon Hakonarson, Andrew Paterson, Ciriaco Piccirillo, Constantin Polychronakos, and J Brent Richards (2019) Genetic Variant of Large effect at STK39 Influences Risk of Type 1 Diabetes. Diabetes (2nd revision)

Despoina Manousaki, Anders Kämpe, Vince Forgetta,Riikka E Makitie, Ghalib Bardai, Alexandre Belisle, Rui Li, Outi Makitie, Frank Rauch, J Brent Richards (2019) Increased burden of common risk alleles in children with a significant fracture history (under review) JBMR 2019 (2nd revision)

Julyan Keller-Baruch,Vincenzo Forgetta, **Despoina Manousaki**, Sirui Zhou, J Brent Richards (2019)Genetically decreased serum vascular endothelial growth factor and osteoporosis outcomes: A Mendelian randomization study. JBMR Plus 2019 (in press)

Despoina Manousaki, Johnny Deladoey, Louis Geoffroy, and Patricia Olivier (2017) Continuous Subcutaneous Insulin Infusion in Children: a Pilot Study Validating a Protocol to Avoid Hypoglycemia at Initiation. Frontiers Endocrinol (Lausanne) 2017. Apr 24;8:84 doi: 10.3389/fendo.2017.00084. eCollection 2017.PMID: 28484424

Despoina Manousaki, Frank Rauch, Josée Dubois, Gilles Chabot, Nathalie Alos. (2016). Pediatric Reference Data for Dual X-Ray Absorptiometric Measures of Normal Lumbar Bone Mineral Density in Young Children: Impact of the Mechanostat. J Musculoskelet Neuronal Interact. 2016. Sep 7;16(3):247-55. PMID: 27609039

Despoina Manousaki, Jack Kent, Karin Haack, Sirui Zhou, Pingxing Xie, Celia M. Greenwood, Paul Brassard, Deborah Newman, Shelley Cole, Jason G. Umans, Guy Rouleau, Anthony G. Comuzzie and J. Brent Richards. (2016). Towards Precision Medicine: TBC1D4 Disruption is Common in The Inuit and Leads to Under-Diagnosis of Type 2 diabetes. Diabetes Care 2016. Nov;39(11):1889-1895. Epub 2016 Aug 25. PMID: 27561922

Lauren E. Mokry, Stephanie Ross, John A. Morris, **Despoina Manousaki**, Vince Forgetta, J. Brent Richards (2016). Genetically decreased vitamin D and risk of Alzheimer's disease. Neurology. 2016. Dec 13;87(24):2567-2574. Epub 2016 Nov 16. PMID: 27856775

Despoina Manousaki, Judith Allanson, Lior Wolf, Cheri Deal. (2014). Characterization of Facial Phenotypes of Children with Congenital Hypopituitarism and their Parents: a matched case-control study. American Journal of Medical Genetics 2015 Jul;167(7):1525-33. doi: 10.1002/ajmg.a.37069. Epub 2015 Apr 5. PMID: 25845580

Despoina Manousaki, Cheri Deal, Jean Jacques De Bruycker, Philippe Ovetchkine, Claude Mercier and Nathalie Alos. (2014). A 15-year-old adolescent with a rare pituitary lesion. Endocrinology, Diabetes and Metabolism Case Reports 2014;2014:140010. doi: 10.1530/EDM-14-0010. Epub 2014 May PMID: 24851183

Magiakou MA, **Manousaki D***, Papadaki M, Hadjidakis D, Levidou G, Vakaki M, Papaefstathiou A, Lalioti N, Kanaka-Gantenbein C, Piaditis G, Chrousos GP, Dacou-Voutetakis C. (2010). The efficacy and safety of gonadotropin-releasing hormone analog treatment in childhood and

adolescence: a single center, long-term follow-up study J Clin Endocrinol Metab 2010 Jan;. 95(95:109-17) PMID: 19897682