Genetic Predictors of Circulating 25-Hydroxyvitamin D and Risk of Colorectal Cancer

Linda T Hiraki1*, Conghui Qu2, Carolyn M Hutter2*, John A Baron3, Sonja I Berndt4, Stéphane Bézieau5, Hermann Brenner6, Bette J Caan7, Graham Casey8, Jenny Chang-Claude9, Stephen J Chanock4,8, David V Conti10, David Duggan11, Charles S Fuchs12,13,14, Steven Gallinger15, Edward Giovannucci11,16,17, Tabitha A Harrison2, Richard Hayes18, Aditi Hazra1,14, Brian Henderson8, Michael Hoffmeister6, John L Hopper19, Thomas J Hudson20,21, Mark A Jenkins19, Sébastien Küry22, Loïc Le Marchand23, Mathieu Lemire20, Jing Ma16, JoAnn E Manson14,12, Hongmei Nan24, Polly A Newcomb2,25, Kimmie Ng26,14, John D Potter27, Robert E Schoen27, Fredrick Schumacher8, Daniela Seminara28, Martha L Slattery29, Jean Wactawski-Wende15, Emily White2,25, Kana Wu17, Brent W Zanke31, Peter Kraft1,32, Ulrike Peters2,25, Andrew T Chan33,16

1 – Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health
2 - Public Health Sciences Division, Fred Hutchinson Cancer Research Center
3 - Division of Gastroenterology and Hepatology, UNC School of Medicine
4 - Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health
5 - Service de Génétique Médicale, CHU Nantes
6 - Division of Clinical Epidemiology and Aging Research, German Cancer Research Center
7 - Division of Research, Kaiser Permanente Medical Care Program
8 - Department of Preventive Medicine, University of Southern California, Keck School of Medicine
9 - Division of Cancer Epidemiology, German Cancer Research Center
10 - Department of Preventive Medicine, Keck School of Medicine, University of Southern California
11 - Translational Genomics Research Institute
12 – Department of Medicine, Brigham and Women’s Hospital
13 - Gastrointestinal Malignancy Program, Dana-Farber Cancer Institute
14 - Department of Medicine, Harvard Medical School
15 - Department of Surgery, Toronto General Hospital
16 - Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School
17 - Department of Nutrition, Harvard School of Public Health
18 - Department of Environmental Medicine, Division of Epidemiology, New York University School of Medicine
19 - Melbourne School of Population Health, The University of Melbourne
20 - Ontario Institute for Cancer Research
21 - Departments of Medical Biophysics and Molecular Genetics, University of Toronto
22 - Service de Génétique Médicale, CHU Nantes
23 - Epidemiology Program, Research Cancer Center of Hawai'i
24 - University of Maryland, School of Medicine
25 - Department of Epidemiology, University of Washington School of Public Health
26 - Medical-Oncology, Dana-Farber Cancer Institute
27 - Department of Epidemiology, University of Pittsburgh Medical Center
28 - Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health
29 - Department of Internal Medicine, University of Utah Health Sciences Center
30 - Department of Social and Preventive Medicine, University of Buffalo
31 - Faculty of Medicine, The University of Ottawa
32 - Department of Biostatistics, Harvard School of Public Health
33 - Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School
Vitamin D genetics GECCO: Manuscript

*Current Address
Carolyn M. Hutter
National Cancer Institute
Host Susceptibility Factors Branch
Bethesda, MD

Linda T. Hiraki
SickKids, Research Institute
Genetics & Genome Biology
Toronto, ON Canada

Grant Support
S. Bézieau and S. Kury are affiliated with ASTERISK which was funded by a regional Hospital Clinical Research Program (PHRC) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC).

J.A. Baron, G. Casey, D.V. Conti, J.L. Hopper, M.A. Jenkins, P.A. Newcomb and F.R. Schumacher are affiliated with CCFR which is supported by the National Cancer Institute, National Institutes of Health under RFA # CA-95-011 and through cooperative agreements with members of the Colon Cancer Family Registry and P.I.s. This genome wide scan was supported by the National Cancer Institute, National Institutes of Health by U01 CA122839. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the CFRs, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the CFR. The following Colon CFR centers contributed data to this manuscript and were supported by the following sources: Australasian Colorectal Cancer Family Registry (U01 CA097735), Seattle Colorectal Cancer Family Registry (U01 CA074794) and Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783).

S. Gallinger, T.J. Hudson, M. Lemire and B.W. Zanke are affiliated with OFCCR which is supported by the National Institutes of Health, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783); see CCFR section above. As subset of ARCTIC, OFCCR is supported by a GL2 grant from the Ontario Research Fund, the Canadian Institutes of Health Research, and the Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society Research Institute. Thomas J. Hudson and Brent W. Zanke are recipients of Senior Investigator Awards from the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation.

B. Henderson and L. Le Marchand are affiliated with COLO2&3/MEC. COLO2&3 is supported by the National Cancer Institute (R01 CA60987). MEC is supported by R37 CA54281, P01 CA033619, and R01 CA63464.

H. Brenner, J. Chang-Claude and M. Hoffmeister are affiliated with DACHS which was supported by grants from the German Research Council (Deutsche Forschungsgemeinschaft, BR
Vitamin D genetics GECCO: Manuscript

1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814).

B.J. Caan, J.D. Potter and M.L. Slattery are affiliated with DALs which was supported by the National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (R01 CA48998 to M.L.S.).

T.A. Harrison, C.M. Hutter, U. Peters and C. Qu are affiliated with GECCO which is supported by National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088). Funding for the genome-wide scan of DALs, PLCO, and WHI was provided by the National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (R01 CA059045).

A.T. Chan, C.S. Fuchs, E.L. Giovannucci, L.T. Hiraki, A. Hazra, P. Kraft, H. Nan, K. Ng and K. Wu are affiliated with HPFS, NHS and PHS: HPFS was supported by the National Institutes of Health (P01 CA 055075., UM1 CA167552, R01 137178, and P50 CA 127003), NHS by the National Institutes of Health (R01 137178, P50 CA 127003, and P01 CA 087969) and PHS by the National Institutes of Health (CA42182).

S.I. Berndt, S.J. Chanock, R.B. Hayes and R.E. Schoen are affiliated with PLCO which was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) prostate cancer scan, supported by the Intramural Research Program of the National Cancer Institute. The datasets used in this analysis were accessed with appropriate approval through the dbGaP online resource (http://www.cgems.cancer.gov/data_access.html) through dbGaP accession number 000207v.1p1.e1. (National Cancer Institute (2009) Cancer Genetic Markers of Susceptibility (CGEMS) data website. http://cgems.cancer.gov/data_access.html; Yeager et al. 2007) Control samples were also genotyped as part of the GWAS of Lung Cancer and Smoking (Landi et al. 2009). Funding for this work was provided through the National Institutes of Health, Genes, Environment and Health Initiative [NIH GEI] (Z01 CP 010200). The human subjects participating in the GWAS are derived from the Prostate, Lung, Colon and Ovarian Screening Trial and the study is supported by intramural resources of the National Cancer Institute. Assistance with genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies, GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01 HG 004438). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000093.

E. White is affiliated with VITAL which is supported in part by the National Institutes of Health (K05 CA154337) from the National Cancer Institute and Office of Dietary Supplements.
Vitamin D genetics GECCO: Manuscript

J. Manson and J. Wactawski-Wende are affiliated with WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

Acknowledgements

ASTERISK: We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students.

DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Renate Hettler-Jensen, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance.

GECCO: The authors would like to thank all those at the GECCO Coordinating Center for helping bring together the data and the people that made this project possible.

HPFS, NHS, PHS: We would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS, and PHS under the supervision of Dr. Immaculata De Vivo and Dr. David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. We would like to thank the participants and staff of the Nurses’ Health Study, the Health Professionals Follow-Up Study, and Physicians’ Health Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY.

PLCO: The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff or the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Barbara O’Brien and staff, Westat, Inc., and Drs. Bill Kopp, Wen Shao, and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: https://cleo.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

Corresponding author:
Andrew T Chan
Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School
Gastroenterology Associates
Vitamin D genetics GECCO: Manuscript

55 Fruit Street
Boston, MA 02114-2696
Phone: 617-726-3212
Fax: 617-724-6832
Email: ACHAN@partners.org

The authors have no conflicts of interest to disclose.
Abstract

Background: Experimental evidence has demonstrated an anti-neoplastic role for vitamin D in the colon and higher circulating 25-hydroxyvitamin D (25[OH]D) levels are consistently associated with a lower risk of colorectal cancer (CRC). Genome-wide association studies have identified loci associated with levels of circulating 25(OH)D. The identified SNPs from four gene regions, collectively explain approximately 5% of the variance in circulating 25(OH)D.

Methods: We investigated whether six polymorphisms in GC, CYP2R1, CYP24A1 and DHCR7/NADSYN1, genes previously shown to be associated with circulating 25(OH)D levels, were associated with CRC risk in 10,061 cases and 12,768 controls drawn from 13 studies included in the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and Colon Cancer Family Registry (CCFR). We performed a meta-analysis of crude and multivariate-adjusted logistic regression models to calculate odds ratios and associated confidence intervals for SNPs individually, SNPs simultaneously, and for a vitamin D additive genetic risk score (GRS).

Results: We did not observe a statistically significant association between the 25(OH)D associated SNPs and CRC marginally, conditionally, or as a GRS, or for colon or rectal cancer separately or combined.

Conclusions: Our findings do not support an association between SNPs associated with circulating 25(OH)D and risk of CRC. Additional work is warranted to investigate the complex relationship between 25(OH)D and CRC risk.

Impact: There was no association observed between genetic markers of circulating 25(OH)D and CRC. These genetic markers account for a small proportion of the variance in 25(OH)D.
Vitamin D genetics GECCO: Manuscript

Introduction

Colorectal cancer (CRC) is the second leading cause of cancer death in men and women in the United States. It is estimated that in 2012, a total of 143,460 U.S. men and women will be diagnosed with cancer of the colon and rectum (1). Family history is a strong risk factor for colorectal cancer(2, 3), which is consistent with the existence of shared etiologic and genetic determinants among relatives. Known genetic mutations account for about 30-50% of the familial risk(4); much of the remaining familial aggregation is unexplained. Genome-wide association studies (GWAS) of sporadic CRC have identified at least 20 independent loci statistically significantly associated with risk (5-12). However, these variants cumulatively explain only a very small fraction of CRC risk (13).

Beyond inherited risk, there is a large body of evidence supporting the role of non-genetic factors, including vitamin D status, in the etiology of CRC. The first human evidence to suggest an association between vitamin D and CRC was the ecologic study by Garland and Garland based on data collected in the 1950’s and 60’s which showed a strong inverse association between colon cancer mortality and solar ultraviolet B (UVB) radiation exposure in the U.S. (14). Since then, most but not all case-control and cohort studies have found an inverse association between intake of vitamin D (both diet and supplements) and CRC risk (15-18), with even stronger associations observed using directly measured circulating 25(OH)D (19-22), an integrated biomarker of vitamin D status (23). Randomized clinical trials of vitamin D supplementation, including the Women’s Health Initiative(24) and the British Oxford Trial(25), have not demonstrated reductions in colorectal cancer incidence. However, these trials have generally tested low doses of vitamin D and each included less than seven years of follow-up, which is likely insufficient to shown an effect on cancer incidence, particularly in light of the
Vitamin D genetics GECCO: Manuscript

long latency of disease. Similarly, large meta-analyses of clinical trials have not shown robust evidence for a protective role of vitamin D in the development of colorectal cancer(26) despite some indication of a preventative role in the development of adenomas(27).

Circulating 25(OH)D levels are a function of dietary sources and exposure of the skin to sunlight, specifically UVB rays. In addition to environmental determinants, twin and family studies suggest that genetic factors contribute substantially to circulating vitamin D levels, with heritability estimates ranging from 43 to 80% (28-31).

Two published GWAS of 25(OH)D have uncovered SNPs significantly associated with lower circulating 25(OH)D levels in four gene regions that appear to have functional relevance: GC (group-specific component vitamin D binding protein); CYP2R1 (cytochrome P450, family 2, subfamily R, polypeptide 1, encoding C-25 hydroxylase that converts vitamin D to the active ligand for the vitamin D receptor; DHCR7/NADSYN1 (7-dehydrocholesterol [7-DHC] reductase/nicotinamide adenine dinucleotide synthetase(32)) with roles in the synthetic vitamin D pathway(33, 34); and CYP24A1 (encoding 24-hydroxylase involved in the degradation of both 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D) (33, 34). Both the GC SNP rs2282679 and the DHCR7/NADSYN1 SNPs are located in intronic regions, with the GC SNP demonstrating the largest magnitude of association with 25(OH)D <75 nmol/L (OR 1.63 (1.53–1.73)(34). The SNP rs10741657 is proximal to the CYP2R1 gene and rs6013897 is proximal to CYP24A1 yet the precise associations with gene expression are yet to be determined.

We investigated the association between these SNPs previously identified as associated with 25(OH)D, and risk of CRC in 13 cohorts that are part of the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and the Colon Cancer Family Registry (CCFR). Within a subset of participants from NHS, HPFS and PHS with measured pre-diagnostic plasma
Vitamin D genetics GECCO: Manuscript

25(OH)D levels, we also examined the joint effects of plasma 25(OH)D and 25(OH)D-associated SNPs on CRC risk.

Materials and Methods

Study Population

The analysis included a total of 10,061 cases and 12,768 controls of European ancestry drawn from 13 studies within GECCO and CCFR. Details on the studies are provided in Table 1, and are described in detail in the Supplemental Note and Supplemental Table 1. In brief, each study defined CRC cases as colorectal adenocarcinoma, confirmed by medical records, pathologic reports, or death certificates. All participants provided informed consent and studies were approved by their respective Institutional Review Boards. None of the studies included in GECCO or CCFR contributed subjects to the any previous GWAS of 25(OH)D with the exception of a subset of the NHS subjects (n=1,342) who participated in the validation stage of one study (33).

Genotyping, Quality Assurance/Quality Control and Imputation

We used genotype data from GECCO and CCFR. GECCO consisted of participants within the French Association Study Evaluating RISK for sporadic colorectal cancer (ASTERISK); Hawaii Colorectal Cancer Studies 2 and 3 (Colo2&3); Darmkrebs: Chancen der Verhütung durch Screening (DACHS); Diet, Activity, and Lifestyle Study (DALS); Health Professionals Follow-up Study (HPFS); Multiethnic Cohort (MEC); Nurses’ Health Study (NHS); Ontario Familial Colorectal Cancer Registry (OFCCR); Physician’s Health Study (PHS); Prostate, Lung, Colorectal Cancer, and Ovarian Cancer Screening Trial (PLCO); VITamins And
Vitamin D genetics GECCO: Manuscript

Lifestyle (VITAL); and the Women’s Health Initiative (WHI). Phase-one genotyping on a total of 1,709 colon cancer cases and 4,214 controls from PLCO, WHI, and DALS (PLCO Set 1, WHI Set 1, and DALS Set 1) was done using Illumina HumanHap 550K, 610K, or combined Illumina 300K and 240K, and has been described previously (12). A total of 650 CRC cases and 522 controls from OFCCR are included in GECCO from previous genotyping using Affymetrix platforms (35). A total of 5,540 CRC cases and 5,425 controls from ASTERISK, Colo2&3, DACHS, DALS Set 2, MEC, PMH, PLCO Set 2, VITAL, and WHI Set 2 were successfully genotyped using Illumina HumanCytoSNP. A total of 2,004 CRC cases and 2,244 controls from HPFS (2 sets), NHS (2 sets), and PHS (2 sets) were successfully genotyped using Illumina HumanOmniExpress. The CCFR included a population-based case-control set of participants from sites in USA, Canada, and Australia successfully genotyped using Illumina Human1M or Human1M-Duo (36).

DNA was extracted from samples of white blood cells or, in the case of a subset of NHS, HPFS, DACHS, MEC, and PLCO samples, and all VITAL samples from buccal cells using conventional methods (37). All studies included 1 to 6% blinded duplicates to monitor quality of the genotyping. All individual-level genotype data were managed centrally at University of Southern California (CCFR), the Ontario Institute for Cancer Research (OFCCR), the University of Washington (HPFS, NHS, and PHS), or the GECCO and CCFR Coordinating Center (CC) at the Fred Hutchinson Cancer Research Center (all other studies) to ensure consistent quality assurance and quality control (QA/QC) and statistical analysis. Details on the QA/QC can be found in Supplemental Table 2. In brief, samples were excluded based on call rate, heterozygosity, unexpected duplicates, gender discrepancy, and unexpectedly high identity-by-descent or unexpected concordance (> 65%) with another individual. For missing SNP data, all
Vitamin D genetics GECCO: Manuscript

GECCO studies were imputed to HapMap II release 24, with the exception of OFCCR, which was imputed to HapMap II release 22. CCFR was imputed using IMPUTE (38), OFCCR was imputed using BEAGLE (39), and all other studies were imputed using MACH (40). All SNPs met quality-control measures for Hardy Weinberg Equilibrium in controls (HWE, \( p \geq 10^{-4} \)), and minor allele frequency (MAF \( \geq 1\% \)) or imputation \( R^2 > 0.3 \).

**Dietary and Lifestyle Factors**

Dietary information, including calcium, folate, fiber and alcohol intake, was available for Colo2&3, DAL5, HPFS, MEC, NHS, PLCO I, PLCO II, VITAL, WHI; calcium, folate and alcohol was available in PHS; and calcium and alcohol in ASTERISK and DACHS. Regular use of non-steroidal anti-inflammatory drugs (NSAIDs) was available for CCFR, Colo2&3, DACHS, HPFS, MEC, NHS and VITAL. All studies collected data on smoking status, red meat consumption, physical activity, body mass index, and hormone replacement therapy in post-menopausal women with the exception of ASTERISK. ASTERISK was restricted to cases with no family history of colorectal cancer. We adopted a flexible approach to retrospective covariate harmonization as previously described (41, 42).

**Laboratory Assessment of 25(OH)D**

In previous studies, we measured plasma levels of 25(OH)D in a subset of the cases and controls with genetic data that were nested within the NHS, HPFS, and PHS (total cases \( n=672 \) and total controls \( n=909 \)) using a radioimmunosorbent assay in the laboratory of Dr. Bruce W. Hollis (Medical University of South Carolina, Charleston, SC). The median intra-assay coefficient of variation from blinded quality-control samples was 11.8% in NHS, 10.1% in
Vitamin D genetics GECCO: Manuscript

HPFS, and 13.8% in PHS. Cases and their controls were analyzed in the same batch, and laboratory personnel were blinded to case, control, and quality-control status (21, 43, 44).

**Statistical Analyses**

The statistical analyses of the GECCO and CCFR samples were conducted at a central data analysis center on individual-level data to ensure a consistent analytical approach. For each study, we estimated the association between each SNP and risk for CRC by calculating betas, odds ratios (ORs), standard errors, 95% confidence intervals (CIs), and p-values using log-additive genetic models relating the genotype dose (0, 1 or 2 copies of the allele) to risk of CRC. For imputed SNPs, we used the dosage (expected number of copies of the minor allele) when testing associations, which has been shown to give unbiased estimates (45). We also created a genetic risk score (GRS), comprised of four SNPs from four distinct gene regions to ensure no single gene was over represented in the score using an allelic scoring system based on summing the number of risk alleles (previously associated with lower 25(OH)D), yielding a possible range of 0-8 alleles to derive estimates of allelic OR.

Minimally adjusted models included covariates for age, sex (when appropriate), center (when appropriate), smoking status, batch effects (ASTERISK only), and the first three principal components from EIGENSTRAT to account for population substructure. Multivariate models were additionally adjusted for family history of CRC, BMI, NSAID use, alcohol use, dietary calcium, folate and red meat intake, sedentary status, and hormone replacement therapy based on covariate availability. We repeated the minimally adjusted model analyses stratified by anatomical site (colon and rectum).
We conducted inverse-variance weighted, fixed-effects meta-analysis to combine beta estimates and standard errors from log-additive models across individual studies. We chose to focus on fixed-effects to improve power and assessed heterogeneity across studies utilizing random effects models (46).

For analyses of the joint effect of plasma 25(OH)D and our GRS comprised of 25(OH)D-associated SNPs, we included the 672 cases and 909 controls in NHS, HPFS, and PHS among whom we had previously measured pre-diagnostic levels of 25(OH)D and also had genotype data (21, 44). We calculated ORs and 95% CI for CRC comparing extreme quartiles of 25(OH)D defined according to cohort-specific cutpoints determined by the distribution in controls (44). We compared the GRS-associated risk for CRC across categories of high versus low vitamin D levels and quartiles of vitamin D, as well as tested for multiplicative interactions between GRS and a 1ng/mL increase in 25(OH)D and high/low vitamin D using a product term in the model and assessing its significance by the Wald method.

We used PLINK, R (47, 48) and SAS 9.2 (SAS Institute Inc, Cary, NC) to conduct the statistical analyses. We estimated our power to detect an association between a GRS and CRC using the method of Tosteson et al. (49). These calculations account for the strength of association between the vitamin D SNPs and circulating 25(OH)D.

Results

Our study included 10,061 CRC cases and 12,768 controls. Overall 53% were female, and the mean age at CRC diagnosis was 64.0 years (+/- 9.6 SD), 55.0% past/current smokers and with risk allele frequencies ranging from 16 – 61%. Table 1 summarizes the characteristics of the studies included in the analyses. Analyses of each individual SNP in models first minimally and
then fully adjusted, did not demonstrate a statistically significant association with CRC risk (Table 2). In analyses stratified by anatomic site each of the four SNP associations remained non-significant (Table 3), in tests of associations with cancers of the colon and rectum.

We considered the possibility that a combination of SNPs associated with circulating 25(OH)D in prior GWAS may be associated with risk of CRC. However, an analysis of a GRS comprised of the risk alleles from the four SNPs associated with plasma 25(OH)D and CRC risk did not demonstrate any significant association (Table 4). The figure 1 forest plot depicts the ORs and 95% CIs of the GRS and CRC association of the individual GECCO and CCFR studies showing a balanced distribution of study-specific odds ratios around the null value of 1.0.

Among the subset of NHS, HPFS and PHS subjects with measured prediagnostic plasma 25(OH)D levels (N=672 cases and N=909 controls), we observed that a 1 allele change in our GRS was associated with a mean 1.5ng/mL decrease in 25(OH)D (p-value <0.0001). Compared with the lowest quartile of 25(OH)D, the highest quartile of 25(OH)D was associated with an OR of 0.66 (95% CI 0.47, 0.92) for CRC. We did not observe a differential association of GRS with risk of CRC according to strata of quartile levels of plasma 25(OH)D or according to high or low levels of 25(OH)D (p-heterogeneity>0.05).

**Discussion**

Consistent evidence from epidemiologic studies supports an inverse association between circulating 25(OH)D, the best integrated biomarker of vitamin D status, and risk of CRC. However, contrary to expectation, we did not observe a statistically significant association between SNPs associated with circulating 25(OH)D and CRC, marginally or in an additive GRS.
A number of epidemiologic studies have reported inverse associations between 25(OH)D and CRC. A meta-analysis of five nested case–control studies reported a pooled odds ratio of 0.49 (95% CI; 0.35–0.68) for CRC comparing the highest quintile (median 37 ng/mL) of 25(OH)D with the lowest (6 ng/mL) (19). Another recent systematic review of nine studies observed pooled RRs for CRC of 0.67 (95% CI, 0.54 to 0.80) comparing extreme quintiles of 25(OH)D (22). Overall, the estimated OR of CRC for a 10ng/mL increase in circulating 25(OH)D was 0.74 (95% CI, 0.63 to 0.89) with the relationship appearing approximately linear (22). Several mechanisms could explain an anti-cancer benefit for vitamin D: reduction of cell proliferation; inhibition of angiogenesis; promotion of cell differentiation; and stimulation of apoptosis (23, 50-57). Vitamin D also has an anti-inflammatory effect, reducing PTGS-2 (COX-2) expression and decreasing levels of the inflammatory marker C-reactive protein (58-60).

Prior clinical trials testing the association between vitamin D and cancer have been null. In a 5-year British placebo-controlled trial with cancer assessed as a secondary outcome, 100,000 IU of vitamin D3 every four months over five years was not associated with CRC incidence (RR=1.02, 95% CI 0.60-1.74)(25). Similarly, among 36,000 women in the WHI calcium-vitamin D trial, a combination of calcium (1000 mg/day) plus low-dose vitamin D3 (400 IU/d) for a mean of 7 years did not reduce CRC incidence (RR=1.08, 95% CI 0.86-1.34)(24). However, the interpretation of these null results is tempered by several important limitations. First, the relatively low doses of vitamin D used were probably inadequate to yield a substantial contrast between the treatment and placebo groups. Second, the duration of follow-up was probably too short to observe an influence on incidence of cancer. Observational data suggest that any influence of calcium and vitamin D intake on CRC risk could require at least 10 years to emerge, consistent with our understanding of the prolonged dwell time of the adenoma-carcinoma
Vitamin D genetics GECCO: Manuscript

pathway(18). On the other hand, a Nebraska population-based placebo-controlled trial of calcium alone or calcium plus vitamin D₃ (1100 IU/d) observed a significantly lower cancer incidence among those supplemented with calcium and vitamin D over just four years of treatment (61). However, follow-up of total cancers was a secondary outcome and there were only a small number of CRC cases, limiting the interpretation of these results.

Over 90% of circulating 25(OH)D is protein-bound with the GC encoded vitamin D binding protein being the major carrier of 25(OH)D. Less than 1% of vitamin D circulates in its unbound form (62). Vitamin D binding protein is a multifunctional protein that also binds fatty acids, and may have immune functions independent of its role as a carrier of vitamin D(63). Prior studies have observed that unbound 25(OH)D was more strongly related to bone mineral density(64), parathyroid hormone levels among hemodialyzed patients(65), than total 25(OH)D, thereby implicating a role for vitamin D binding protein in modifying the biologic activity of circulating vitamin D. The available estimates of the association between 25(OH)D and CRC, as well as genetic markers of 25(OH)D, are based solely on total circulating 25(OH)D levels (19, 22, 33, 34). It is unclear how these estimates might change when accounting for vitamin D binding protein levels or by individually examining free and protein-bound 25(OH)D.

Prior studies have examined individual SNPs in CYP24A1 or GC (66, 67) in association with risk of CRC. A prior DALS multicenter population-based case-control study of 1,600 CRC cases found a statistically significant association between one CYP24A1 polymorphism and overall risk of colon cancer, particularly for proximal colon cancer, as well as an association between three CYP24A1 polymorphisms and distal colon cancer (67). However, the correlation between these CYP24A1 genetic markers investigated by the DALS study and the CYP24A1 marker examined in this full combined analysis study is very low (r² <0.1). Because our a priori
hypothesis was that SNPs most strongly associated with 25(OH)D levels would be associated with CRC, we did not consider total genetic variation in CYP24A1 with CRC risk. It is possible that alternative CYP24A1 SNPs may be associated with CRC through mechanisms independent of 25(OH) D levels. A prior study of the CCFR cohort of 1,750 sibships found no evidence for associations between GC and the risk of CRC, and no evidence for modification of the association by calcium and/or vitamin D intake (66). Studies of additional genes that are not significantly associated with circulating 25(OH)D but are implicated in the vitamin D pathway, including the vitamin D receptor (VDR) SNPs FokI and BsmI, have yielded inconsistent results (68-71). Notably, among over 10,000 men with prostate cancer, a greater number of low 25(OH)D SNPs were associated with a decreased risk of aggressive prostate cancer (OR 0.66; 95% CI, 0.44-0.98 for 6–8 vs. 0–1 alleles.; Ptrend = 0.003)(72).

The lack of association that we observed between genetic markers associated with circulating 25(OH)D and CRC is consistent with prior clinical trials of vitamin D and CRC and would, at least initially, seem to argue against a causal association between vitamin D and CRC(73). However, prior work has demonstrated that these 4 SNPs, though correlated with circulating 25(OH)D, explain only a small fraction (5%) of the variance in circulating 25(OH)(74). Recently, a Scottish case-control study observed a significant association between direct plasma measurements of 25(OH)D and CRC risk, yet failed to replicate the association using an instrumental-variable method of mendelian randomization (MR) with the same four genome-wide significant risk loci examined in our analysis (75). The investigators attributed these inconsistent results to a presumed weak correlation between these SNPs and 25(OH)D, as well as a limited sample size of 2,001 cases of CRC and 2,237 controls. Given our significantly larger sample size of approximately 10,000 cases and 12,500 controls and assuming a correlation
between our GRS and 25(OH)D of \( r=0.17 \), with a 10 ng/ml increase in 25(OH)D associated with OR=0.74 for CRC(22), our power to detect a 1-allele change in our GRS is 96% (significance level of 0.05). However, if the true magnitude of association with a 10ng/mL increase in 25(OH)D is in fact an OR of 0.85 for CRC, we would have only had 56% power to detect a 1-allele change in our GRS.

We are not certain of the precise pathway or biological mediators by which 25(OH)D influences CRC risk. Our GRS assumes that each included SNP would be associated with increased CRC risk according to their observed association with lower 25(OH)D. If this assumption is invalid, combining the alleles into this GRS would reduce our power to detect associations with CRC. If we remove the GC SNP and repeat our power calculation, our observed correlation between our GRS and 25(OH)D becomes \( r=0.11 \), resulting in 68% power to detect a 1 allele change in our proxy score.

We acknowledge some limitations. First, our study includes only populations of European descent, which limits the generalizability of our findings. However, the circulating 25(OH)D SNPs that we examined were identified in GWAS of populations of European descent, and so the underlying genetic associations should hold in our study population. Moreover, limiting our analysis to European descent populations minimizes the potential for confounding by population structure. Second, if these SNPs are correlated with another locus that influences the risk of CRC, this could confound our results(73). Third, despite our large sample size, we had limited power to detect associations between individual SNPs and risk of CRC.

In conclusion, our findings do not support an association between SNPs associated with circulating 25(OH)D and risk of CRC. This may be due to the fact that these SNPs account for only a small portion of the variance observed in circulating 25(OH)D levels and that those alleles
Vitamin D genetics GECCO: Manuscript

associated with low circulating levels of 25(OH)D may not affect CRC risk in the same
direction. Future studies are needed to examine the role of unbound and protein-bound 25(OH)D,
along with other biomarkers of the vitamin D pathway, in the development of CRC.
References

Vitamin D genetics GECCO: Manuscript


Vitamin D genetics GECCO: Manuscript

Vitamin D genetics GECCO: Manuscript

47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-75.
Vitamin D genetics GECCO: Manuscript


**Table 1. Characteristics of colorectal cancer patients in included study populations**

<table>
<thead>
<tr>
<th>Study*</th>
<th>Abbreviation</th>
<th>Number of Cases</th>
<th>Number of Controls</th>
<th>% Female</th>
<th>Mean age diagnosis</th>
<th>% CRC family history</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontario Familial Colorectal Cancer Registry</td>
<td>OFCCR</td>
<td>650</td>
<td>522</td>
<td>52</td>
<td>62.0</td>
<td>20.1</td>
</tr>
<tr>
<td>Colon Cancer Family Registry</td>
<td>CCFR</td>
<td>1,171</td>
<td>983</td>
<td>50</td>
<td>54.2</td>
<td>16.9</td>
</tr>
<tr>
<td>Diet, Activity, and Lifestyle Study</td>
<td>DALS</td>
<td>1,116</td>
<td>1174</td>
<td>45</td>
<td>63.9</td>
<td>13.3</td>
</tr>
<tr>
<td>A case-control study from the University of Hawai’i</td>
<td>COLO 2&amp;3</td>
<td>87</td>
<td>125</td>
<td>45</td>
<td>65.2</td>
<td>11.8</td>
</tr>
<tr>
<td>Multiethnic Cohort</td>
<td>MEC</td>
<td>328</td>
<td>346</td>
<td>46</td>
<td>63.0</td>
<td>12.2</td>
</tr>
<tr>
<td>Darmkrebs: Chancen der Verhütung durch Screening</td>
<td>DACHS</td>
<td>1,710</td>
<td>1,708</td>
<td>41</td>
<td>68.6</td>
<td>12.5</td>
</tr>
<tr>
<td>Prostate, Lung, Colorectal, and Ovarian</td>
<td>PLCO</td>
<td>1,019</td>
<td>2,391</td>
<td>31</td>
<td>64.0</td>
<td>11.4</td>
</tr>
<tr>
<td>Women’s Health Initiative</td>
<td>WHI</td>
<td>1,476</td>
<td>2,538</td>
<td>100</td>
<td>65.8</td>
<td>17.6</td>
</tr>
<tr>
<td>Association Study Evaluation RISK for sporadic colorectal cancer</td>
<td>ASTERISK</td>
<td>892</td>
<td>947</td>
<td>41</td>
<td>65.2</td>
<td>NA</td>
</tr>
<tr>
<td>VITamins And Lifestyle Study</td>
<td>VITAL</td>
<td>285</td>
<td>288</td>
<td>48</td>
<td>66.5</td>
<td>13.6</td>
</tr>
<tr>
<td>Health Professionals Follow-up Study</td>
<td>HPFS</td>
<td>403</td>
<td>402</td>
<td>0</td>
<td>65.2</td>
<td>17.3</td>
</tr>
<tr>
<td>Nurses’ Health Study</td>
<td>NHS</td>
<td>549</td>
<td>955</td>
<td>100</td>
<td>59.8</td>
<td>15.2</td>
</tr>
<tr>
<td>Physicians’ Health Study</td>
<td>PHS</td>
<td>375</td>
<td>389</td>
<td>0</td>
<td>58.9</td>
<td>NA</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>10,061</strong></td>
<td><strong>12,768</strong></td>
<td><strong>53</strong></td>
<td><strong>64.0</strong></td>
<td><strong>13.0</strong></td>
</tr>
</tbody>
</table>

* DALS Set 2, Colo2&3, DACHS, MEC, PLCO Set 2, WHI Set 2, ASTERISK, and VITAL were genotyped on the Illumina CytoSNP BeadChip. WHI Set 1 was genotyped using Illumina 550K and 550K duo platforms; PLCO Set 1 was genotyped using Illumina 610K and 550K platforms; DALS set 1 was genotyped using Illumina 610K and 550K platforms; OFCCR was genotyped using Affymetrix GeneChip Human Mapping 100K and 500K Array Set and a 10K non-synonymous SNP chip; CCFR was genotyped using Illumina 1M, 1MDuo and 1M-Omni platforms; and HPFS, NHS, and PHS were genotyped on the OmniExpress platform.
Table 2: Association between five 25(OH)D associated SNPs and CRC among 10,061 cases and 12,768 controls.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position (Nearest Gene)</th>
<th>Major/Minor Allele Frequency</th>
<th>Age, Sex and PCA-Adjusted OR (95% CI)**</th>
<th>p-value</th>
<th>Multivariable-Adjusted OR (95% CI)**</th>
<th>p-value</th>
<th>N studies genotyped†</th>
<th>Mean R²</th>
<th>p-value for Heterogeneity‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2282679</td>
<td>GC</td>
<td>T/G</td>
<td>1.03 (0.99, 1.08)</td>
<td>0.14</td>
<td>1.02 (0.77, 1.09)</td>
<td>0.14</td>
<td>11</td>
<td>1.00</td>
<td>0.55</td>
</tr>
<tr>
<td>rs10741657</td>
<td>CYP2RI</td>
<td>G'/A</td>
<td>0.98 (0.94, 1.02)</td>
<td>0.22</td>
<td>0.97 (0.93, 1.02)</td>
<td>0.23</td>
<td>0</td>
<td>0.99</td>
<td>0.73</td>
</tr>
<tr>
<td>rs12785878</td>
<td>DHCR7/NADSYN1</td>
<td>T/G</td>
<td>1.03 (0.99, 1.08)</td>
<td>0.15</td>
<td>1.04 (0.99, 1.10)</td>
<td>0.10</td>
<td>0</td>
<td>0.99</td>
<td>0.73</td>
</tr>
<tr>
<td>rs11234027</td>
<td>DHCR7/NADSYN1</td>
<td>G/A</td>
<td>1.01 (0.96, 1.07)</td>
<td>0.63</td>
<td>1.03 (0.97, 1.10)</td>
<td>0.29</td>
<td>9</td>
<td>0.99</td>
<td>0.19</td>
</tr>
<tr>
<td>rs6013897</td>
<td>CYP24A1</td>
<td>T/A</td>
<td>0.97 (0.92, 1.02)</td>
<td>0.30</td>
<td>0.98 (0.92, 1.04)</td>
<td>0.46</td>
<td>0</td>
<td>0.88</td>
<td>0.74</td>
</tr>
</tbody>
</table>

| Abbreviations: CRC=colorectal cancer, SNP=single-nucleotide polymorphism, OR=odds ratio, CI=confidence interval, N studies genotyped=number of studies with directly genotyped in all the studies, Mean r²=average imputation R2 value across all the studies which the SNP was imputed |
| * Allele associated with decreased circulating 25(OH)D in prior GWAS (Ahn et al. HMG 2010; Wang et al. Lancet 2010.). |
| **Odds ratio calculated in reference to the allele associated with decreased 25(OH)D. Multivariable models included available covariates age, sex (when appropriate), center (when appropriate), smoking status, batch effects, three principal components from EIGENSTRAT, family history of CRC, BMI, NSAID use, alcohol use, dietary calcium, folate and red meat intake, sedentary status, and hormone replacement therapy. |
| † Number of studies directly genotyping and imputing SNPs varies for each SNP. |
| ‡ p-value for heterogeneity across study in the multivariable-adjusted models. |

Table 3: Association between five 25(OH)D associated SNPs and CRC stratified by site

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position (Nearest Gene)</th>
<th>Age, Sex and PCA-Adjusted OR (95% CI)**</th>
<th>p-value</th>
<th>N Studies*</th>
<th>p-value for Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2282679</td>
<td>GC</td>
<td>1.03 (0.98, 1.08)</td>
<td>0.23</td>
<td>12</td>
<td>0.29</td>
</tr>
<tr>
<td>rs10741657</td>
<td>CYP2RI</td>
<td>0.98 (0.94, 1.02)</td>
<td>0.35</td>
<td>12</td>
<td>0.57</td>
</tr>
<tr>
<td>rs12785878</td>
<td>DHCR7/NADSYN1</td>
<td>1.05 (1.00, 1.10)</td>
<td>0.06</td>
<td>12</td>
<td>0.21</td>
</tr>
<tr>
<td>rs11234027</td>
<td>DHCR7/NADSYN1</td>
<td>1.02 (0.97, 1.08)</td>
<td>0.30</td>
<td>12</td>
<td>0.39</td>
</tr>
<tr>
<td>rs6013897</td>
<td>CYP24A1</td>
<td>0.96 (0.91, 1.02)</td>
<td>0.21</td>
<td>12</td>
<td>0.75</td>
</tr>
</tbody>
</table>

<p>| Abbreviations: CRC=colorectal cancer, SNP=single-nucleotide polymorphism, OR=odds ratio, CI=confidence interval |
| *A total of 7220 colon cancer and 2308 rectal cancer cases were included in each site-specific analysis with 12,768 controls. |
| **Odds ratio calculated in reference to the allele associated with decreased 25(OH)D. |</p>
<table>
<thead>
<tr>
<th>SNP</th>
<th>Position (Nearest Gene)</th>
<th>N Cases/ Controls</th>
<th>Major/ Minor Allele</th>
<th>Age, Sex and PCA-Adjusted OR (95% CI)**</th>
<th>p-value</th>
<th>N studies</th>
<th>p-value for Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2282679</td>
<td>GC</td>
<td>10,128/12,768</td>
<td>T/G *</td>
<td>1.03 (0.99, 1.08)</td>
<td>0.13</td>
<td>13</td>
<td>0.20</td>
</tr>
<tr>
<td>rs10741657</td>
<td>CYP2R1</td>
<td>10,128/12,768</td>
<td>G*/A</td>
<td>0.98 (0.94, 1.02)</td>
<td>0.22</td>
<td>13</td>
<td>0.77</td>
</tr>
<tr>
<td>rs12785878</td>
<td>DHCR7/NADSYN1</td>
<td>10,128/12,768</td>
<td>T/G *</td>
<td>1.03 (0.99, 1.08)</td>
<td>0.16</td>
<td>13</td>
<td>0.10</td>
</tr>
<tr>
<td>rs6013897</td>
<td>CYP24A1</td>
<td>10,128/12,768</td>
<td>T/A *</td>
<td>0.97 (0.93, 1.03)</td>
<td>0.33</td>
<td>13</td>
<td>0.89</td>
</tr>
<tr>
<td>Score</td>
<td>-</td>
<td>10,128/12,768</td>
<td>-</td>
<td>1.00 (0.98, 1.03)</td>
<td>0.72</td>
<td>13</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Abbreviations: CRC=colorectal cancer, SNP=single-nucleotide polymorphism, OR=odds ratio, CI=confidence interval

* Allele associated with decreased circulating 25(OH)D in prior GWAS (Ahn et al. HMG 2010; Wang et al. Lancet 2010.).

**Odds ratio calculated in reference to the allele associated with decreased 25(OH)D.
Figure 1: Forest plot of the genetic risk score and CRC for individual studies and meta-analysis of all studies (allelic OR, 95% CI)
Figure 1

<table>
<thead>
<tr>
<th>Study</th>
<th>OR</th>
<th>95%CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASTERISK</td>
<td>1.03</td>
<td>(0.95-1.12)</td>
<td>5.08e-01</td>
</tr>
<tr>
<td>COLO23</td>
<td>1.11</td>
<td>(0.89-1.37)</td>
<td>3.58e-01</td>
</tr>
<tr>
<td>CCFR</td>
<td>1.02</td>
<td>(0.95-1.09)</td>
<td>5.72e-01</td>
</tr>
<tr>
<td>DACHS</td>
<td>0.99</td>
<td>(0.94-1.05)</td>
<td>8.16e-01</td>
</tr>
<tr>
<td>DALS</td>
<td>1.03</td>
<td>(0.96-1.10)</td>
<td>4.14e-01</td>
</tr>
<tr>
<td>HPFS</td>
<td>1.02</td>
<td>(0.92-1.15)</td>
<td>6.69e-01</td>
</tr>
<tr>
<td>MEC</td>
<td>0.92</td>
<td>(0.82-1.04)</td>
<td>2.04e-01</td>
</tr>
<tr>
<td>NHS</td>
<td>0.93</td>
<td>(0.85-1.04)</td>
<td>8.83e-02</td>
</tr>
<tr>
<td>OFCCCR</td>
<td>1.04</td>
<td>(0.94-1.14)</td>
<td>4.65e-01</td>
</tr>
<tr>
<td>PHS</td>
<td>0.89</td>
<td>(0.79-1.00)</td>
<td>4.85e-02</td>
</tr>
<tr>
<td>PLCO</td>
<td>0.97</td>
<td>(0.91-1.03)</td>
<td>3.68e-01</td>
</tr>
<tr>
<td>VITAL</td>
<td>1.02</td>
<td>(0.89-1.17)</td>
<td>7.66e-01</td>
</tr>
<tr>
<td>WHI</td>
<td>1.05</td>
<td>(1.00-1.11)</td>
<td>5.63e-02</td>
</tr>
<tr>
<td>Summary</td>
<td>1.00</td>
<td>(0.98-1.03)</td>
<td>7.25e-01</td>
</tr>
</tbody>
</table>

p-value for heterogeneity = 0.11