UNIVERSITY OF CALIFORNIA

Los Angeles

Dairy and Milk Intake, Genetic Predispositions, and Circulating Proteins: Unveiling the Multifaceted Landscape of Colorectal Cancer Risk and Progression

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy in Epidemiology

by

Sihao Han

2024

ABSTRACT OF THE DISSERTATION

Dairy and Milk Intake, Genetic Predispositions, and Circulating Proteins: Unveiling the Multifaceted Landscape of Colorectal Cancer Risk and Progression

by

Sihao Han

Doctor of Philosophy in Epidemiology

University of California, Los Angeles, 2024

Professor Zuo-Feng Zhang, Co-Chair

Assistant Professor Brian Z. Huang, Co-Chair

Background: Colorectal cancer (CRC) remains a significant global health challenge, influenced by both dietary factors and genetic predispositions. Dairy/milk consumption has been suggested to be involved in CRC pathogenesis, yet the underlying mechanisms, particularly in ethnically diverse populations, remain poorly understood. This research aims to unravel the complex relationships between dairy/milk consumption, genetic polymorphisms related to dairy/milk digestion, and the influence of circulating proteins on CRC risk and progression.

Objectives and Specific Aims: The study aims to: 1) Assess the association between dairy/milk intake (including lactose, calcium, and vitamin D) and CRC incidence and mortality across race/ethnicity; 2) Investigate the associations between genetic polymorphisms in *LCT*, *MCM6*, *CASR*, and *VDR* genes and CRC risk, including gene-diet interactions; 3) Explore the causal relationship between circulating proteins and CRC risk.

Methods: Data from the Multiethnic Cohort study, involving 215,634 participants, were analyzed for dietary associations. Genetic analyses involved candidate SNPs on a sub-cohort of 70,000 participants. Proteomic Mendelian Randomization (MR) analyses leveraged summary-level statistics from various GWAS studies.

Results: Higher consumption of dairy/milk and their key components was linked to a 13-17% reduced CRC risk. There was a weak association between dairy/milk intake and mortality among CRC patients. While no evident racial/ethnic heterogeneity was observed in incidence, significant differences were noted across racial/ethnic groups in mortality. No interaction was found between dairy/milk consumption and lactase persistence status. Several SNPs within *LCT, MCM6, CASR,* and *VDR* were associated with CRC risk across the different racial/ethnic groups. Elevated genetically determined lactase-phlorizin hydrolase levels were inversely associated with CRC risk. Four unique proteins (GREM1, LPH, PDE5A, LIMA1) were associated with CRC risk in the *cis*-MR analyses, while 15 additional proteins were identified in the analyses using all (*cis*+*trans*) protein quantitative trait loci.

Conclusions: This study highlights the protective effects of dairy/milk intake against CRC risk, the genetic factors influencing CRC susceptibility, and the potential role of circulating proteins in CRC carcinogenesis. These findings emphasize the importance of race/ethnicity-specific dietary guidelines and genetic risk stratification in CRC prevention, suggesting targeted public health interventions for effective CRC management.

The dissertation of Sihao Han is approved.

Jian Yu Rao

Roch Arnaud Kibsa Nianogo

Brian Z. Huang Committee Co-Chair

Zuo-Feng Zhang, Committee Co-Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

LIST OF TABLES ..... **vii**

LIST OF FIGURES ..... **ii**

ACKNOWLEDGEMENTS ..... **vi**

VITA ..... **ix**

CHAPTER 1. BACKGROUND **1**

1.1 Colorectal Cancer 1
1.1.1 Overview 1
1.1.2 Epidemiology 1
1.1.3 Pathogenesis 3
1.1.4 Risk Factors 4

1.2 Gaps in Knowledge 8

CHAPTER 2. RESEARCH OBJECTIVES **9**

2.1 Rationale and Objectives 9

2.2 Specific Aims and Hypotheses 9

CHAPTER 3. SPECIFIC AIM 1A – COHORT STUDY **11**

3.1 Introduction 11

3.2 Methods 12
3.2.1 Study population 12
3.2.2 Exposure assessment 13
3.2.3 Outcome assessment 14
3.2.4 Statistical analyses 14

3.3 Results 16
3.3.1 Baseline characteristics 16
3.3.2 Total dairy products, milk, lactose, calcium, and vitamin D intakes and colorectal cancer risk among the overall population 17
3.3.3 Stratified analyses by race and ethnicity groups, sex, and age groups 17
3.3.4 Stratified analysis by tumor site 18

3.4 Discussion 19

CHAPTER 4. SPECIFIC AIM 1B – SURVIVAL ANALYSIS **22**

4.1 Introduction 22

4.2 Methods 22
4.2.1 Study population 22
4.2.2 Exposure assessment 23
4.2.3 Outcome assessment 24
4.2.4 Statistical analyses 24

4.3 Results 26
4.3.1 Baseline characteristics 26
4.3.2 Total dairy products, milk, lactose, calcium, and vitamin D intakes and survival among the overall population 26
4.3.3 Stratified analyses by race and ethnicity groups, sex, and age group 27
4.3.4 Stratified analyses by tumor site, tumor stage, and tumor grade 29

4.4 Discussion 30
CHAPTER 5. SPECIFIC AIM 2A – CANDIDATE SNPS ASSOCIATION STUDY 33
5.1 Introduction 33
5.2 Methods 35
5.2.1 Study population 35
5.2.2 SNPs selection 36
5.2.3 Assessment of other variables 37
5.2.4 Outcome assessment 37
5.2.5 Statistical analyses 38
5.3 Results 39
5.3.1 rs4988235 and CRC 39
5.3.2 G×E interaction 40
5.3.3 Single TagSNP association 41
5.4 Discussion 41
CHAPTER 6. SPECIFIC AIM 2B – MENDELIAN RANDOMIZATION STUDY 44
6.1 Introduction 44
6.2 Methods 46
6.2.1 Study design 46
6.2.2 Genetic instruments 47
6.2.3 Outcome assessment 49
6.2.4 Statistical analyses 50
6.3 Results 52
6.3.1 FinnGen Dataset 52
6.3.2 PLCO Dataset 53
6.3.3 Pan-UK Biobank Dataset 53
6.3.4 Meta-Analysis Combining FinnGen, PLCO, and Pan-UK Biobank Results 54
6.3.5 CRC Subtype-Specific MR Analyses 54
6.4 Discussion 55
CHAPTER 7. SPECIFIC AIM 3 – PROTEOME-WIDE MENDELIAN RANDOMIZATION STUDY 62
7.1 Introduction 62
7.2 Methods 63
7.2.1 Data sources for plasma proteins 63
7.2.2 Data source for CRC 64
7.2.3 Statistical analyses 65
7.3 Results 68
7.4 Discussion 71
CHAPTER 8. CONCLUSIONS AND PUBLIC HEALTH IMPLICATIONS 75
TABLES 78
FIGURES 135
REFERENCES 177

LIST OF TABLES

Table 1-1. Research gaps and corresponding dissertation aims **78**

Table 3-1. Baseline characteristics of participants according to intake of all dairy products, Multiethnic Cohort Study, 1993–2019 ..... **79**

Table 3-2. Relative risk of colorectal cancer according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019 ..... **81**

Table 3-2 (*continue*). Relative risk of colorectal cancer according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019 ..... **83**

Table 3-3. Relative risk of colorectal cancer according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different age groups, Multiethnic Cohort Study, 1993-2019 .. **85**

Table 3-3 (*continue*). Relative risk of colorectal cancer according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different age groups, Multiethnic Cohort Study, 1993-2019 ..... **87**

Table 4-1. Baseline characteristics of CRC patients according to intake of total dairy products, Multiethnic Cohort Study, 1993–2019 ..... **89**

Table 4-2. Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019 ..... **91**

Table 4-2 (*continue*). Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019 ..... **93**

Table 4-3. Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019 ..... **95**

Table 4-3 (*continue*). Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019 ..... **97**

Table 4-4. Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different age groups, Multiethnic Cohort Study, 1993-2019..... **99**

Table 4-5. Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different age groups, Multiethnic Cohort Study, 1993-2019 ..... **101**

Table 4-6. Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor stage groups, Multiethnic Cohort Study, 1993-2019..... **103**

Table 4-6 (*continue*). Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor stage groups, Multiethnic Cohort Study, 1993-2019..... **105**

Table 4-7. Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor stage groups, Multiethnic Cohort Study, 1993-2019..... **107**

Table 4-8. Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor grade groups, Multiethnic Cohort Study, 1993-2019..... **109**

Table 4-9. Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor grade groups, Multiethnic Cohort Study, 1993-2019..... **111**

Table 5-1. Baseline characteristics of participants according to genotype of rs4988235, Multiethnic Cohort Study **113**

Table 5-2. Genotype distribution of rs4988235 according to race and ethnicity group..... **116**

Table 5-3. Association between rs4988235 and the risk of CRC **117**

Table 5-4. Overall and race-specific association of dairy and milk intake on CRC risk by genotypes of rs4988235 ..... **118**

Table 5-5. Significant associations (before FDR correction) between candidate SNP and risk of CRC among participants in the Multiethnic Cohort Study **119**

Table 6-1. Summary of GWAS datasets used for LPH levels and CRC **122**

Table 6-2. Characteristics of genetic instruments for elevated LPH levels from the GWAS identified in the GWAS Catalog **123**

Table 6-3. Summary of four genetic instruments and their proxies (where necessary) from the FinnGen, PLCO, and Pan-UK Biobank GWAS on CRC **124**

Table 6-4. Associations of genetically predicted elevated LPH Levels and CRC in the FinnGen, PLCO, and Pan-UK Biobank datasets **125**

Table 6-5. Associations of genetically predicted elevated LPH levels and colon cancer in the FinnGen, PLCO, and Pan-UK Biobank datasets **126**

Table 6-6. Associations of genetically predicted elevated LPH levels and rectal cancer in the FinnGen, PLCO, and Pan-UK Biobank datasets **127**

Table 6-7. Meta-analysis results for the association between elevated LPH levels and CRC, colon cancer, and rectal cancer **128**

Table 7-1. Summary of GWAS datasets for plasma proteins and CRC **129**

Table 7-2. MR analysis results for the association between plasma proteins and CRC using *cis* and *cis+trans* pQTLs **130**

Table 7-3. Colocalization analysis of *cis*-pQTLs and CRC **132**

Table 7-4. Steiger filtering analysis of causal directionality test between plasma proteins and CRC **133**

LIST OF FIGURES

Figure 1-1. Anatomy of the Gastrointestinal System..... **135**

Figure 1-2. Global Incidence and Mortality of leading cancers in 2022 ..... **136**

Figure 1-3. Ten leading cancer types for the estimated new cancer cases and deaths by sex, United States, 2024 ..... **137**

Figure 1-4. Incidence and mortality of colorectal cancer by age and sex worldwide..... **138**

Figure 1-5. Pathways of colorectal carcinogenesis **139**

Figure 3-1. Flowchart of the study..... **140**

Figure 3-2. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of CRC among participants in the Multiethnic Cohort **141**

Figure 3-3. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of CRC across race and ethnicity groups in the Multiethnic Cohort..... **142**

Figure 3-4. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of CRC in men and women in the Multiethnic Cohort **143**

Figure 3-5. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of CRC across age groups in the Multiethnic Cohort**. 144**

Figure 3-6. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of colon and rectal cancer among participants in the Multiethnic Cohort..... **145**

Figure 4-1. Flowchart of the study..... **146**

Figure 4-2. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and survival among CRC patients in the Multiethnic Cohort..... **147**

Figure 4-3. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the overall survival among CRC patients across race and ethnicity groups in the Multiethnic Cohort **148**

Figure 4-4. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the CRC-specific survival among CRC patients across race and ethnicity groups in the Multiethnic Cohort **149**

Figure 4-5. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and overall survival among CRC patients in men and women in the Multiethnic Cohort **150**

Figure 4-6. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and CRC-specific survival among CRC patients in men and women in the Multiethnic Cohort **151**

Figure 4-7. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the overall survival among colon and rectal cancer patients in the Multiethnic Cohort **152**

Figure 4-8. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the CRC-specific survival among colon and rectal cancer patients in the Multiethnic Cohort **153**

Figure 5-1. Flowchart of the study **154**

Figure 5-2. Association between rs4988235 and risk of CRC among participants in the Multiethnic cohort **155**

Figure 6-1. Schematic overview of the study design **156**

Figure 6-2. Scatter plots of the IVW and MR-Egger methods investigating the effect of elevated LPH levels on CRC in the FinnGen, PLCO, and Pan-UK Biobank datasets **157**

Figure 6-3. Forest plots of the IVW estimates on the association between genetically predicted LPH levels and CRC risk for each genetic instrument in the FinnGen, PLCO, and Pan-UK Biobank datasets **158**

Figure 6-4. Leave-one-out analyses for the MR analysis on LPH levels and CRC risk in the FinnGen, PLCO, and Pan-UK Biobank datasets **159**

Figure 6-5. Meta-analysis results for the association of LPH levels with CRC risk **160**

Figure 6-6. Scatter plots of the IVW and MR-Egger methods investigating the effect of elevated LPH levels on colon cancer in the FinnGen, PLCO, and Pan-UK Biobank datasets **161**

Figure 6-7. Forest plots of the IVW estimates on the association between genetically predicted LPH levels and colon cancer risk for each genetic instrument in the FinnGen, PLCO, and Pan-UK Biobank datasets..... **162**

Figure 6-8. Leave-one-out analyses for the MR analysis on LPH levels and colon cancer risk in the FinnGen, PLCO, and Pan-UK Biobank datasets **163**

Figure 6-9. Scatter plots of the IVW and MR-Egger methods investigating the effect of elevated LPH levels on rectal cancer in the FinnGen, PLCO, and Pan-UK Biobank datasets **164**

Figure 6-10. Forest plots of the IVW estimate on the association between genetically predicted LPH levels and rectal cancer risk for each genetic instrument in the FinnGen, PLCO, and Pan-UK Biobank datasets..... **165**

Figure 6-11. Leave-one-out analyses for the MR analysis on LPH levels and rectal cancer risk in the FinnGen, PLCO, and Pan-UK Biobank datasets **166**

Figure 6-12. Meta-analysis results for the association of elevated LPH levels with colon cancer risk using *cis*-MR and MR using all genetic variants **167**

Figure 6-13. Meta-analysis results for the association of elevated LPH levels with rectal cancer risk using *cis*-MR and MR using all genetic variants **168**

Figure 7-1. Schematic overview of the study design **169**

Figure 7-2. Volcano plot of the *cis*-only MR analyses for the association between plasma proteins and CRC risk..... **170**

Figure 7-3. Volcano plot of the MR analyses with all pQTLs (*cis*+*trans*) for the association between plasma proteins and CRC risk **171**

Figure 7-4. PPI network for *cis*-prioritized proteins associated with CRC risk **172**

Figure 7-5. PPI network for all MR significant proteins associated with CRC risk..... **173**

Figure 7-6. GO enrichment analysis for *cis*-MR prioritized proteins..... **174**

Figure 7-7. GO enrichment analysis for *cis*+*trans* MR prioritized proteins **175**

Figure 7-8. KEGG enrichment analysis for *cis*-MR prioritized proteins..... **176**

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to everyone who has supported me throughout my PhD journey, especially during these challenging and uncertain times. Your collective encouragement and belief in my potential have been a constant source of motivation and strength, helping me navigate through the highs and lows of this academic endeavor.

My deepest appreciation goes to Dr. Zuo-Feng Zhang, my academic advisor and co-chair of my dissertation committee. Your unwavering support and invaluable guidance have been instrumental in my academic and personal growth. Your mentorship has extended beyond the realm of academia, enriching my life in countless ways. I am profoundly grateful for the countless hours you have invested in my development and the opportunities you have provided for me to excel.

I am equally grateful to Dr. Brian Z. Huang, co-chair of my dissertation committee. Your friendship, insightful wisdom, and unwavering support have been indispensable to my work. Thank you for everything. Your ability to offer both critical feedback and steadfast encouragement has been a unique and invaluable combination, helping me to refine my research and grow as a scholar.

I would also like to extend my sincere thanks to Dr. Jian-Yu Rao and Dr. Roch A. Nianogo, members of my dissertation committee. Your continuous support and constructive feedback have significantly shaped my dissertation. Your expertise and dedication to my academic success have been truly inspiring, pushing me to achieve a higher standard of excellence.

A special thank you to Dr. Shira Shafir, under whom I have had the pleasure of being a Teaching Assistant numerous times. Your exemplary teaching has been a source of inspiration,

and you will always be a role model to me. The insights and skills I have gained under your guidance have been invaluable, and I am honored to have had the opportunity to learn from you.

I am also grateful to Dr. Julia Heck and Dr. Marissa Seamans for providing me with invaluable working opportunities that have greatly contributed to my professional development. Your mentorship and the experiences you have offered have enriched my understanding of the field and expanded my professional horizons.

I would like to extend my appreciation to Joy Miller for creating an environment conducive to my research and for providing the necessary resources. Your support has been crucial in facilitating my research endeavors, and I am deeply thankful for your commitment to fostering a productive and supportive environment.

I owe a debt of gratitude to all the professors I have worked with and learned from, including Dr. Onyebuchi A. Arah, Dr. Beate Ritz, Dr. Donatello Telesca, Dr. Su Yon Jung, Dr. Noel Weiss, Dr. Edward Giovannucci, and many others. Your expertise and mentorship have profoundly impacted my academic journey. The knowledge and insights I have gained from each of you have been instrumental in shaping my research and intellectual growth.

My appreciation also goes out to my peers, members of the Cancer Molecular Epidemiology Group, colleagues, and co-authors who have enriched my graduate experience through their support and thoughtful discussions. Your intellectual contributions have made this journey more enriching and enjoyable, and I am grateful for the collaborative spirit that you have fostered.

To my friends, parents, and all family members, your endless love, patience, and encouragement have been my stronghold. Special thanks to Dr. Jiemin Yao, Eachan Chu, Dr.

Hiroyoshi Iwata, Dr. Mingyan Zhang, Dr. Fang Fang, Dr. Claire Euiyoung Kim, Dr. Anqi Wang, Yili Xu, Dr. Qianxi Feng, Ziyuan Wang, Cory T. Lumsdaine, Gina Eujeong Nam, and Dr. Christina Hoven. Your unwavering support and belief in my abilities have been the bedrock upon which I have built my success.

I am thankful for the department, school, and university that provided me with funding and resources, including fellowship, summer award, travel grants, and the funding support from Dr. Brian Z. Huang and Dr. Zuo-Feng Zhang. These resources have been indispensable in allowing me to pursue my research and academic goals without financial burden.

I am grateful to the participants from various studies, including the Multiethnic Cohort Study (MEC), The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, UK Biobank, Fenland study, FinnGen Study, and deCODE study, for their invaluable contributions to my research. Your collaboration and the data provided have been essential to the progress and success of my dissertation work.

Lastly, I would like to remember Dr. Janet Sinsheimer, who served on my committee and was an exceptionally kind and dedicated mentor. Her passing last year was a great loss, and I dedicate a moment to honor her memory. Her legacy of kindness, brilliance, and dedication to students will continue to inspire me and many others.

Finally, I would like to extend a small thank you to myself for enduring and thriving through this challenging but rewarding journey. This achievement would not have been possible without your perseverance and dedication. You are the best, and your hard work and resilience have truly paid off in the end.

VITA

EDUCATION

2018	M.S in Epidemiology, T.H. Chan School of Public Health, Harvard University
2015	M.P.H in Epidemiology & Biostatistics, Brown School, Washington University in St. Louis
2013	M.B.B.S in Clinical Medicine, Wuhan University
2012	B.S in Cognitive Psychology, Central China Normal University

PROFESSIONAL EMPLOYMENT (SELECTED)

2020-2024	Teaching Fellow/Associate/Assistant, University of California, Los Angeles
2019-2024	Research Assistant, University of California, Los Angeles
2017-2018	Research Assistant, Brigham and Women's Hospital & Harvard University

SELECTED PUBLICATIONS

1. **Han S.**, Yao J., Yamazaki H., Streicher S.A., Rao J., Nianogo R.A., Zhang Z., Huang B.Z., 2024. Genetically determined circulating lactase/phlorizin hydrolase concentrations and risk of colorectal cancer: a two-sample Mendelian randomization study. *Nutrients*. 16(6):808

2. Jung, S.Y., Scott, P.A., Papp, J.C., Sobel, E.M., Pellegrini, M., Yu, H., **Han, S.** and Zhang, Z.F., 2021. Genome-wide association analysis of proinflammatory cytokines and gene–lifestyle interaction for invasive breast cancer risk: the WHI dbGaP study. *Cancer Prevention Research*, *14*(1), pp.41-54.

3. Jung, S.Y., Mancuso, N., **Han, S.** and Zhang, Z.F., 2020. The role of genetically determined glycemic traits in breast cancer: a Mendelian randomization study. *Frontiers in Genetics*, 11, p.1076.

4. Yang, B., Zhou, Z.H., Chen, L., Cui, X., Hou, J.Y., Fan, K.J., **Han, S.H.**, Li, P., Yi, S.Q. and Liu, Y., 2018. Prognostic significance of NFIA and NFIB in esophageal squamous carcinoma and esophagogastric junction adenocarcinoma. *Cancer medicine*, 7(5), pp.1756-1765.

CHAPTER 1. BACKGROUND

1.1 Colorectal Cancer

1.1.1 Overview

Colorectal cancer (CRC), commonly referred to as bowel cancer, develops within the large intestine, consisting of the colon and rectum (Figure 1-1). This malignancy originates from the uncontrolled proliferation of abnormal cells in the mucosal lining of the colon or rectum [1]. CRC typically starts as benign polyps in the colorectum's inner lining [1]. Although most polyps remain noncancerous, some may evolve into cancer over time. CRC continues to pose a significant public health challenge worldwide, with a substantial impact on morbidity and mortality. This underscores the urgency of the CRC related research to advance our understanding of the etiology and clinical management of this disease.

1.1.2 Epidemiology

Incidence and Mortality

Despite advances in understanding of CRC, it remains the third most common cancers globally, with 1.93 million new diagnosed cases, and it ranks as the second leading cause of cancer-related deaths resulting in nearly 0.9 million deaths annually (Figure 1-2) [2]. It accounts for about 10% of all cancers diagnosed and of cancer-related deaths worldwide [2]. The incidence rates in developed countries are higher than developing countries, and most likely associated with western style diet and the availability of improved screening modality and early detection methods [3]. The ongoing westernization in many countries is expected to increase the burden of CRC to an estimated 3.2 million cases by 2040 [4].

In the United States, CRC is the third most common cancer in both men and women (Figure 1-3) [5]. There are estimated to be 152,810 newly diagnosed CRC cases in 2024, including 106,590 colon and 46,220 rectal cancer cases [5]. Projected mortality figures suggest 53,010 deaths from CRC, making it the third leading cause of cancer death in men and fourth in women [5].

The incidence and mortality of CRC increase steadily along with the age and most of cases and deaths occur after 50 years old [1]. Men have higher age-adjusted CRC rates than women, with 1.3 to 1.5 times higher incidence and mortality rate, a disparity that becomes more pronounced after age 50 (Figure 1-4) [1, 3]. This disparity can be partly attributed to variations in cumulative environmental exposure, sex hormones, and lifestyle factors [6-8]. While there is an overall decline in CRC incidence and mortality due to the population-based screening [9], the rise in cases among younger adults is concerning, because it is often associated with a more advanced stage at diagnosis and less favorable histological features [10].

Survival

CRC survival are significantly influenced by the stage at diagnosed, with early-stage detection yielding considerably higher survival rates [1, 2, 11-14]. The five-year survival rate of CRC exceeds 90% when CRC is diagnosed at an early stage in contrast to a mere 11-15% for late-stage diagnoses [13, 14]. Globally, the cumulative risk of colon cancer mortality between age 0 and 74 is 0.65% for males and 0.45% for females [2, 13].

In the United States, more than 1.4 million individuals were living with a history of CRC as of January 1st, 2022, according to the Surveillance, Epidemiology, and End Results (SEER) cancer registries [14]. This population is expected to grow, driven by demographic shifts and an aging population.

The 5-year relative survival rate for CRC in the U.S. has increased from 50% in the mid-1970s to 65% in 2017 [14]. This progress is largely attributed to advancements in treatment modalities, such as targeted chemotherapies and cancer-directed surgeries, alongside enhancements in early detection techniques and broader healthcare access [15-17]. The survival rate for colon cancer (64%) is slightly lower than rectal cancer (67%) [14].

Despite a recent uptick in early-onset CRC cases defined as diagnosis before the age of 50, research has not identified a significant disparity in 5-year cancer-specific survival rates between early-onset CRC, often associated with more advanced stage at diagnosis, and later-onset cases across all stages [1, 18].

1.1.3 Pathogenesis

Generally, there are three primary pathways of pathogenesis for CRC (Figure 1-5) [1].

Adenoma–carcinoma sequence pathway

This pathway accounts for over 80% of CRC cases [1, 3]. The accumulation of genetic aberrations and epigenetic alterations drives the malignant transformation of normal colorectal epithelium cells, evolving into adenoma (polyp), and eventually CRC over a decade or more [1]. Mutations in the tumor suppressor gene *APC* may play an important role of initiation, followed by deregulation of Wnt signaling pathway and subsequent neoplastic cells [19-21]. Mutations in the proto-oncogene *KRAS* and the ensuing function loss of the tumor suppressor gene *TP53* are believed to be involved in the progression process from adenoma to invasive CRC [22]. This classic adenoma–carcinoma sequence paradigm is also associated with chromosomal instability (CIN) [23].

Serrated pathway

Representing 10-20% of CRCs, this pathway starts with serrated polyps, mostly sessile serrated adenomas, evolving into CRC [3, 24]. Compared to the traditional adenoma-carcinoma pathway, this pathway is often initiated by mutations in the oncogene *BRAF* instead of the inactivation of *APC* [25]. The activation of mitogen-activated protein kinase (MAPK) pathway due to *BRAF* mutations leads to serrated polyps and then CRC with the facilitation by methylation in CpG island methylator phenotype (CIMP) cells [24, 25]. This pathway is more associated with microsatellite instability-high (MSI-H) tumors and high levels of CIMP [26-28].

Inflammatory pathway

A less common pathway, accounting for fewer than 2% of CRCs, involves chronic inflammation leading to a progression from no dysplasia to a high degree of dysplasia and ultimately colorectal carcinoma [1, 29]. Unlike the other pathways, this one is characterized by multifocality of dysplastic lesions, with *TP53* mutations occurring early and *APC* mutations less frequent and later in the process [29-34]. Chronic inflammation and oxidative stress are thought to drive neoplastic transformation in this pathway, with conditions like ulcerative colitis (UC) significantly increasing CRC risk [29, 35, 36].

1.1.4 Risk Factors

The incidence of CRC is significantly influenced by a confluence of genetic, environmental, and lifestyle factors [1, 37].

Hereditary and Genetic

Hereditary colorectal cancer syndromes contribute to approximately 5% of CRC cases [1, 38], such as Lynch Syndrome (Hereditary nonpolyposis CRC), characterized by MSI resulting from dysfunction of the DNA mismatch repair system [3, 38], and Familial adenomatous polyposis (FAP), marked by numerous adenomatous polyps due to inherited germline mutations [39]. In addition, individuals with a family history of CRC or inflammatory bowel disease (IBD) in relatives face a considerably increased risk [40-43], with meta-analysis indicating a 2.24 to 3.97-fold increased risk among these populations [44]. Researchers have also successfully identified several susceptibility genes associated with CRC risk through genome-wide association studies (GWAS) [45, 46].

Environmental factors

Sporadic cases, accounting for 60% to 65% of all CRC patients [1], often arise from somatic genetic or epigenetic abnormalities caused by modifiable risk factors rather than hereditary predispositions, underscoring the pivotal role of environmental risk factors [47].

Several environmental factors have been shown to be associated with the increased risk of CRC, including unhealthy dietary pattern, sedentary lifestyle and lower physical activity, smoking, alcohol intake, obesity, and so on [1, 3, 38, 48-51].

Unhealthy dietary patterns or western-style diet, particularly those rich in red or processed meats, sugary beverages, and refined grains, are linked to an increased risk of CRC [1, 52]. Conversely, a prudent dietary pattern characterized by high intakes of fruits, vegetables, whole grains, nuts, fish, and dairy products is associated with a reduced risk [1, 52]. A meta-analysis of 28 studies have highlighted that a Western-style diet increases CRC risk by 25%, whereas a

prudent dietary pattern can reduce the risk by 19% [53]. The relationship between diet and CRC may involve multiple biological mechanisms mediated by diverse nutrients, including calcium, vitamin D, fiber, folic acid, carbohydrates, total fat, saturated fatty acid, and iron.

In a meta-analysis of prospective observational studies, researchers found that the risk of CRC decreased 8% for each 300mg/day increase in total calcium intake [54], although randomized clinical trials (RCT) have not consistently supported this association, partly due to short follow-up periods [55]. Experimental evidence suggested that calcium can mitigate carcinogenic effects by binding fatty acids and bile acids and activating calcium-sensing receptors (CaSR), which inhibit proliferation of colonic cells and promote cell differentiation and apoptosis [56, 57].

Dairy products, despite some inconsistent epidemiological findings, are predominantly reported to have a protective effect against CRC, attributed to its components such as calcium, vitamin D, lactoferrin, and conjugated linoleic acid [58-65]. A recent systematic review revealed that a higher intake of dairy products and milk is associated with an approximately 20% reduction in CRC risk compared to lower consumption levels [66]. In addition, a pooled analysis of cohort studies further supported the protective role of dairy consumption, even when controlling for calcium's effect, suggesting additional calcium-independent mechanisms of CRC prevention [60].

Physical inactivity or sedentary lifestyle can reduce metabolic rates and gut motility, which may lead to an inflammatory situation, insulin resistance, and metabolic dysregulation, thereby elevating the risk of CRC [40, 67]. Albeit an optimal intensity of activity has not been established, empirical evidence indicated that exercise by the amount of 5 metabolic equivalents of task (MET) hours weekly can reduce 8% risk of colon cancer [1, 48].

Tobacco smoking has been an established risk factor of CRC because of involvement with more than 60 carcinogens that can induce genetic or epigenetic aberrations, such as polycyclic

aromatic hydrocarbon (PAHs), N-nitrosamines, and aromatic amines [68]. Meta-analysis of studies suggested that the risk of CRC increased 6% for 5 pack-years and 26% for 30 pack-years [69].

The International Agency for Research on Cancer (IARC) has classified both ethanol in alcoholic beverages and its metabolite acetaldehyde as group 1 carcinogenic to humans [70]. Meta-analyses indicated that both incidence and mortality of CRC are positively associated with alcohol consumption [50, 71]. Notably, even light drinking (less than 1 alcoholic drink per day) has been shown to be associated with 4% increased CRC risk [71].

Many observational studies consistently suggested that excessive adiposity is associated with increased risk of CRC with different measurements [1, 51, 72, 73]. Importantly, abdominal fatness appears to have a more significant correlation with CRC risk compared to overall body fatness [1, 74, 75]. It is noteworthy that the association of CRC risk with obesity is more pronounced in women than in men, suggesting a potential sex-specific differential impact of obesity on CRC susceptibility [76].

The role of the gut microbiota in CRC development is increasingly recognized [77-83]. Evidence indicated that the impact of diet pattern on gut microbial community could perform an important role in the development of CRC [77, 84-87]. For example, the alteration of dietary pattern may change not only microbial composition but also gene expression in the intestinal microbiome [77, 88]. The difference of microbial characteristics between CRC patients and healthy individuals further supports that microbiome ought to be considered as a factor to contribute the onset of colorectal carcinogenesis [40, 81].

1.2 Gaps in Knowledge

Given the complex etiology of CRC involving multifaceted interactions between dietary and genetic factors, there remains a substantial research gap in fully understanding these interactions, especially across diverse ethnic groups (Table 1-1). Although existing research has highlighted the potential protective roles of dairy and milk consumption against CRC, the underlying mechanisms of these associations remain elusive. Moreover, the heterogeneity of these associations among different ethnicities calls for further investigation.

In addition, the exploration of genetic polymorphisms associated with dairy and milk digestion pathways, such as those within the *LCT*, *MCM6*, *CASR*, and *VDR* genes, is either incomplete or predominantly focused on European populations. This oversight leaves a pronounced gap in our knowledge, especially considering the potential for genetic variations to influence individual responses to dietary components and the risk of CRC.

Furthermore, while emerging evidence points to a critical role for circulating proteins in CRC carcinogenesis, the causal links between these proteins and CRC risk are yet to be fully explored. Mendelian Randomization offers a promising method for research to unveil these connections but has yet to be fully leveraged.

Addressing these gaps is critical for enhancing our understanding of CRC pathogenesis and could significantly influence public health outcomes by informing more targeted dietary recommendations, improving risk stratification through genetic and proteomic profiling, and potentially uncovering new therapeutic targets for CRC prevention and treatment.

CHAPTER 2. RESEARCH OBJECTIVES

2.1 Rationale and Objectives

The primary goal of this dissertation is to delve into the multifaceted relationships between lifestyle/dietary factors, genetic predispositions, circulating proteins, and the risk and progression of CRC, with a special focus on the roles of dairy/milk consumption, alongside genetic polymorphisms associated with dairy/milk digestion. This comprehensive exploration aims to shed light on how these factors contribute to CRC development and progression, particularly across different ethnic groups. By unraveling these relationships, this research endeavors to enhance our understanding how diet, genetics, and proteins influence CRC, potentially paving the way for tailored prevention, early detection, and treatment strategies, addressing critical gaps in our current understanding of CRC etiology and progression.

2.2 Specific Aims and Hypotheses

Aim 1: To assess the association between dairy/milk intake, including their key components (lactose, calcium, vitamin D), and CRC risk and survival, and further investigate the racial heterogeneity effects in these associations.

- **Hypothesis 1:** Higher consumption of dairy/milk products, including their key components (lactose, calcium, vitamin D), might be associated with reduced CRC incidence and mortality.
- **Hypothesis 2:** The relationship between pre-diagnostic dairy/milk intake, including their key components (lactose, calcium, vitamin D), and CRC incidence and mortality might vary across different racial and ethnic groups.

Aim 2: To examine the relationship of genetic polymorphisms related to dairy/milk digestion with CRC risk, focusing on candidate SNPs within the *LCT*, *MCM6*, *CASR*, and *VDR* genes, as well as the G×E interaction with dairy/milk intake, and further investigate the causal relationship between genetically predicted LPH level and CRC incidence.

- **Hypothesis 1:** SNPs in genes related to dairy/milk digestion pathways (*LCT*, *MCM6*, *CASR*, *VDR*) might be associated with CRC risk, with the magnitude and direction of these associations potentially varying across different racial and ethnic groups.
- **Hypothesis 2:** The relationship between dairy/milk intake and CRC risk might be heterogeneous across lactase persistence status.
- **Hypothesis 3:** Elevated genetically predicted LPH level, indicative of lactose digestion capability, might be linked to a reduced risk of CRC.

Aim 3: To explore the causal relationship between genetically predicted level of circulating proteins and CRC risk.

- **Hypothesis:** Circulating proteins might perform a critical role in the carcinogenesis of CRC.

CHAPTER 3. SPECIFIC AIM 1A – COHORT STUDY

3.1 Introduction

Colorectal cancer (CRC) is one of the most common forms of cancer in the digestive system. Dietary factors are considered to play a crucial role in the pathogenesis of CRC, with multiple studies linking the consumption of specific foods and nutrients to CRC risk [89, 90]. Research has suggested that dairy products, particularly milk, may offer protection against colorectal neoplasia [90]. Many epidemiologic studies have investigated the association between dairy product consumption and CRC risk [91-118]. A meta-analysis of 24 prospective cohort studies revealed a protective effect of dairy, including milk, against CRC incidence [119]. However, most studies were conducted predominantly in Europeans or Whites, with limited research involving ethnic minorities [107-109], underscoring a significant gap in the literature.

Notably, an earlier analysis within the Multiethnic Cohort Study (MEC), which included 2,110 CRC cases diagnosed up to December 31, 2001, observed an inverse association between total dairy products intake and CRC risk in both men and women, with milk specifically reducing CRC risk in men [107]. The study, however, was underpowered to detect ethnicity-specific associations. Given the potential heterogeneity of CRC risk across different race and ethnicity groups [120], understanding ethnic-specific associations between risk factors and CRC could help elucidate disparities and inform more nuanced public health recommendations.

The protective association between dairy product and CRC may perhaps be attributed to several key nutrients in dairy products, including calcium [121, 122], vitamin D [123, 124], and lactose [99]. Additional constituents in dairy, including butyric acid, conjugated linoleic acid,

sphingolipids, and lactoferrin, have also been identified as potential contributors to its role in CRC prevention [61, 77, 125, 126]. However, limited epidemiologic studies were conducted in ethnic diverse populations.

In this study, our aim was to elucidate the effects of total dairy intake and its principal nutrients on CRC risk within a multiethnic cohort, with a particular focus on racial and ethnic minorities previously underrepresented in research. Our primary objective was to utilize over 18 additional years of follow-up (from 1993 to 2019) to conduct updated and ethnicity-specific analyses on dairy and milk intake and CRC risk in the MEC. Secondly, we aimed to evaluate the potential roles of individual key nutritional components in dairy and milk, specifically calcium, vitamin D, and lactose, in CRC risk.

3.2 Methods

3.2.1 Study population

The MEC is a prospective cohort study initiated to explore risk factors and disparities in cancer and other chronic conditions [127]. It includes >215,000 individuals from Los Angeles and Hawaii aged 45-75 at enrollment between 1993 and 1996. The cohort consists of mainly five racial and ethnic groups: African Americans, Japanese Americans, Native Hawaiians, Latinos, and Whites. At cohort entry, participants completed a self-administrated and comprehensive 26-page baseline questionnaire covering demographic characteristics, anthropometric measurements, personal and family history of medical conditions, medication use, lifestyle factors, and dietary intake.

The study was approved by the Institutional Review Boards of the University of Hawaii and the University of Southern California, with all participants providing informed consent at cohort entry.

In this study, participants were excluded if they (Figure 3-1): (1) had prevalent CRC at baseline (n=2,673); (2) had invalid histology indicating non-CRC (n=17); (3) did not belong to the five main racial and ethnic groups (n=11,618); (4) had implausible dates recorded, such as missing dates of study entry/exit, date of death before date of diagnosis, or follow-up time less than zero (n=27); or (5) provided invalid diet information (calorie intake less than zero) (n=8,655).

3.2.2 Exposure assessment

Study participants' dietary intakes were assessed at baseline using a self-reported validated quantitative food frequency questionnaire (QFFQ) on over 180 food items [127-129]. The QFFQ was specifically developed for the MEC, including specific food items that were in traditional diets of particular racial and ethnic groups [127]. For each food and beverage item, participants were asked to report the frequency (eight categories for foods and nine for beverages), and portion size (three (in a few instances four) choices based on typical serving sizes for each single food or grouping of foods) of consumption. These data were then analyzed using the food composition table developed by the Cancer Research Center of Hawaii to compute individual daily food and nutrient intakes [127]. Additionally, the baseline questionnaire collected information on participants' use of vitamin and mineral supplements.

In our analysis, we employed intake densities (intake per 1,000 kcal per day) for evaluating the consumptions of five dairy-related food items: total dairy products, milk, lactose, calcium, and

vitamin D. This was computed by dividing the absolute intake of the food item or nutrient by total daily caloric intake, and then multiplying by 1,000. Previous studies suggested that intake densities tended to yield better correlations than using absolute intake amounts, minimizing the potential measurement error from dietary recall [128, 130]. For vitamin D and calcium intake, we included total intakes from both foods and supplements, where supplemental vitamin D intake was from multivitamin supplements and supplemental calcium intake was from both calcium supplements and multivitamins.

3.2.3 Outcome assessment

Study participants were followed from cohort entry until the earliest occurrence of CRC diagnosis, death, or the conclusion of the follow-up period on December 31, 2019. Information on mortality was obtained from Hawaii and California state death certificates and the National Death Index. Incident CRC cases were ascertained by linkage to the statewide Surveillance, Epidemiology, and End Results (SEER) Program tumor registries of Hawaii and California. These cases were further classified based on anatomic subsites, utilizing the International Classification of Diseases for Oncology, third edition (ICD-O-3) codes: C18.0-C18.7 for colon, and C19.9 and C20.9 for rectum cancer.

3.2.4 Statistical analyses

We used Cox proportional hazards regression to estimate hazard ratios (reported as relative risks [RRs]) and 95% confidence intervals (CIs) for the associations between each of the five dairy-related food items (total dairy products, milk, lactose, calcium, and vitamin D) and CRC risk. Each food item was assessed in a separate model. The proportional hazards assumption was assessed

using the Schoenfeld residuals. Intakes of total dairy products, milk, lactose, calcium, and vitamin D were analyzed as cohort-specific quintiles and as continuous variables (per standard deviation [SD] increase). To assess trends in increasing quintiles, we ran a separate model where the quintile variable was treated as a continuous measure for each exposure of interest.

All Cox regression models included age as the time metric, with gender, race and ethnicity, and 10-year age group at baseline (e.g., aged<50, 50-60, 60-70, ≥70) as strata variables. The fully adjusted models further included the following covariates assessed at baseline: family history of colorectal cancer (yes vs. no), history of intestinal polyps (yes vs. no), education (≤12th grade, vocational/some college, ≥college graduate), body mass index (BMI, <25, 25-30, ≥30 kg/m^2), diabetes (yes vs. no), smoking status and pack-years (never smoker, former smoker <20 pack-years, former smoker ≥20 pack-years, current smoker <20 pack-years, current smoker ≥20 pack-years), alcohol consumption (never, <15, 15-30, ≥30 g/day), quintiles of physical activity (MET-hours/day), use of nonsteroidal anti-inflammatory drugs (yes vs. no), regular use of multivitamins (yes vs. no), log-transformed total energy intake (kcal/day), quintiles of red meat density intake (g/1,000 kcal/day), quintiles of processed meat density intake (g/1,000 kcal/day), quintiles of dietary fiber density intake (g/1,000 kcal/day), quintiles of folate intake (μg DFE/day), and use of hormone replacement therapy (male non-user, female missing menopause information, premenopausal female, postmenopausal female ever user, postmenopausal women never user/unknown use).

For all exposures, we performed stratified analyses by racial and ethnic group (Whites, African Americans, Native Hawaiians, Japanese Americans, Latinos), sex (men and women), age group (<50, 50-59, 60-69, ≥70 years old), and tumor site (colon and rectum). Heterogeneity across

groups was assessed by likelihood ratio test comparing the model with and without interaction terms for the subgroup variable and exposure.

All statistical tests were two-sided, with the level of significance pre-determined at $P < 0.05$. We performed all analyses using R version 4.3.0 (The R Foundation for Statistical Computing) [131].

3.3 Results

3.3.1 Baseline characteristics

After exclusions, there were 192,644 individuals included for analyses, including 46,977 Whites (24.3%), 32,765 African Americans (17.0%), 13,744 Native Hawaiians (7.1%), 53,960 Japanese Americans (28.0%), and 45,178 Latinos (23.5%). During an average follow-up time of 20.0 years, a total of 6,575 incidence CRC cases were identified among eligible participants.

The baseline characteristics of the study population, stratified by quintile of total dairy product intake, are presented in Table 3-1. Compared to those with lower dairy product intake, participants with higher intake were more likely to be older and never smokers, and more likely to use nonsteroidal anti-inflammatory drugs (NSAID), multivitamins, and hormone replacement therapy (post-menopausal women). In addition, they were less likely to have family history of CRC and history of colorectal polyps. They also consumed less red meat and processed meat and more dietary fiber and folate than those with lower intake of total dairy products.

3.3.2 Total dairy products, milk, lactose, calcium, and vitamin D intakes and colorectal cancer risk among the overall population

Results from the multivariate-adjusted Cox regression models demonstrated inverse associations between total dairy products intake and CRC risk in the study population (Figure 3-2). Among all MEC participants, the highest quintile of total dairy products consumption was associated with a decreased risk of CRC (RR_{Q5vsQ1} = 0.86, 95% CI: 0.79-0.94), and a significant trend was observed across quintiles (p_{trend} = 0.001). Similar inverse associations were also observed between milk intake and CRC incidence (Figure 3-2), comparing individuals whose milk consumption was in the highest quintile to the lowest (RR_{Q5vsQ1} = 0.83, 95% CI: 0.76-0.91).

In terms of key components in the dairy and milk (Figure 3-2), higher lactose intake was associated with a decrease in CRC risk (RR_{Q5vsQ1} = 0.87, 95% CI: 0.80-0.95). Similar patterns were observed for calcium and vitamin D intake, higher quintile consumptions were associated with a decreased risk of CRC (calcium: RR_{Q5vsQ1} = 0.82, 95% CI: 0.74-0.92; vitamin D: RR_{Q5vsQ1} = 0.87, 95% CI: 0.76-1.00).

3.3.3 Stratified analyses by race and ethnicity groups, sex, and age groups

Stratified analyses by race and ethnicity groups (Table 3-2 and Figure 3-3) demonstrated statistically significant associations of intakes in total dairy products, milk, calcium on CRC risk among Latinos (total dairy products: RR_{Q5vsQ1} = 0.77, 95% CI: 0.62-0.95; milk: RR_{Q5vsQ1} = 0.72, 95% CI: 0.59-0.89; calcium: RR_{Q5vsQ1} = 0.71, 95% CI: 0.54-0.94; vitamin D: RR_{Q4vsQ1} = 0.74, 95% CI: 0.57-0.91). However, no heterogeneity was observed across race and ethnicity groups for all exposures.

Evidence of heterogeneity was observed across sex (Figure 3-4) in the association between lactose intake and CRC risk, with results demonstrating an inverse association between higher lactose intake and CRC risk in men but not in women (men: RR_{Q5vsQ1} = 0.85, 95% CI: 0.75-0.97; women: RR_{Q5vsQ1} = 0.90, 95% CI: 0.79-1.01; $p_{heterogeneity}$ = 0.033). Although statistically significant heterogeneity was detected across sex for vitamin D intake ($p_{heterogeneity}$ = 0.036), neither sex specific group showed significant association with CRC risk. Results were homogenous across sex for the associations between total dairy products, milk, and calcium intakes and CRC risk.

Regarding the analyses stratified by age groups (Table 3-3 and Figure 3-5), albeit significant associations between intakes of total dairy products, milk, lactose, calcium, and vitamin D and CRC risk were observed for certain age groups, there was no significant heterogeneity across age groups for all exposures.

3.3.4 Stratified analysis by tumor site

Results from analyses by CRC subtypes (Figure 3-6) indicated inverse associations between higher intake of total dairy products and colon cancer but not rectal cancer risk (colon cancer: RR_{Q5vsQ1} = 0.86, 95% CI: 0.77-0.95, p_{trend} = 0.009). For milk intake, there was an inverse association between higher intake and colon cancer risk (RR_{Q5vsQ1} = 0.85, 95% CI: 0.77-0.94), and the protective effect was stronger for rectal cancer, with a 0.77 RR comparing the highest quintile of milk consumption to the lowest (95% CI: 0.65-0.93; p_{trend} = 0.011).

Regarding the key nutrients in dairy products (Figure 3-6), an inverse association between higher lactose intake and colon cancer risk was observed (RR_{Q5vsQ1} = 0.88, 95% CI: 0.79-0.97). For total calcium intake, statistically significant inverse associations were detected in both colon

and rectal cancer (colon: RR_{Q5vsQ1} = 0.84, 95% CI: 0.74-0.95; rectum: RR_{Q5vsQ1} = 0.77, 95% CI: 0.91-0.97). In addition, for total vitamin D intakes, statistically significant inverse associations were only detected in rectal cancer (RR_{Q5vsQ1} = 0.74, 95% CI: 0.56-0.97) but not in colon cancer.

3.4 Discussion

In this study, we investigated the relationship between intakes of dairy products and milk and CRC risk in a racially and ethnically diverse prospective cohort. We further explored the role specific dairy-related nutrients (lactose, calcium, and vitamin D) in CRC risk and progression. Our analysis revealed an inverse relationship of total dairy products and milk intakes with CRC risk. Compared to individuals consuming the lowest quintile, the reductions of CRC incidence among those in the highest quintile were 14% for dairy intake and 17% for milk intake, respectively. Similar reductions in CRC risk were also observed among participants with higher consumptions of lactose, calcium, and vitamin D. While the associations were consistent across racial and ethnic groups, Latino group demonstrated stronger inverse associations for dairy, milk, and calcium consumptions.

The findings in our study align with previous analyses in the MEC, with the current study benefiting from 18 additional years of follow-up and triple the number of CRC cases [107]. This allowed us to perform race- and ethnicity-specific analyses with greater statistical power. Moreover, our findings were consistent with previous epidemiologic studies on the relationship between dairy and milk intakes and the risk of CRC. A recent meta-analysis of 24 prospective cohort studies on CRC incidence found a protective effect of dairy products and milk against CRC development [119]. In addition, meta-analyses on the relationship between calcium and vitamin D intake indicated inverse associations of these nutrients with CRC development [89, 132]. Our

research contributes to this body of evidence by examining the associations across races and ethnicities, sex, age groups and subtypes of CRC, suggesting that protective effects of dairy and milk on CRC development are homogeneous across populations.

The potential underlying mechanisms linking higher dairy and milk consumption with reduced CRC risk may be due to key components in dairy and milk, such as lactose, calcium, and vitamin D [77]. Calcium's protective effects can be attributed to its capacity to bind secondary bile acids and ionized fatty acids, as well as activating certain signaling pathways though the calcium-sensing receptor (CaSR) [122, 133]. Vitamin D modulates molecular pathways relevant to CRC development, including the downregulation of the *COX-2* gene, the upregulation of 15-hydroxyprostaglandin dehydrogenase (15-PDGH), and interference with β-catenin-mediated gene transcription [121, 123]. Moreover, it is suggested that butyrate, a metabolite fermented by gut bacteria from lactose [134], plays an important role in antitumor activities [135]. Many epidemiologic studies suggested an inverse relationship between calcium [136], vitamin D [132] and CRC incidence. However, studies on lactose's relationship with CRC have been limited [94, 99, 137] and have yielded inconclusive results, likely due to small sample sizes.

By using data from the MEC, one of the strengths of our study was the ability to use a large and racially and ethnically diverse prospective cohort to minimize recall and selection bias and compare associations across multiple underrepresented minority groups. Moreover, we were able to update our results from our previous MEC study and provide further information on the associations of total dairy products and milk intakes with CRC incidence. In addition, cancer case ascertainment through linkage to SEER cancer registries allowed us to achieve near complete capture of CRC cases. Lastly, by studying the association between key nutrients in dairy products

and CRC risk, we were able to conduct a comprehensive assessment of the potential mechanisms involved in the relationship between dairy products and milk and CRC development.

However, our study had some limitations. First, the dietary intakes were assessed at baseline as time-invariant variables, despite the possibility of changes over the follow-up period. Nevertheless, previous MEC analyses identified no significant differences in total calcium and vitamin D among CRC cases diagnosed within the first 2 years follow-up and those diagnosed later [107]. Second, the self-reported information on food intake might lead to measurement error for the exposure. However, results from a calibration study in the MEC indicated that the use of nutrient densities calculated from total energy intake could improve the quality of the exposure assessment [128]. In addition, we were unable to track participants who moved out of the state during the follow-up. Lastly, it is worth noting that the total vitamin D intake in our study might not be an ideal indicator for vitamin D exposure in humans. Though we considered vitamin D from both food and supplements, the unavailability of data on other vitamin D sources such as ultraviolet exposure undermined our ability to accurately measure vitamin D exposure.

In conclusion, our study provides support for a protective role of dairy products against CRC risk and suggests that milk, lactose, calcium, and vitamin D may contribute to this protective effect. Our findings further indicate that the effects are homogenous across racial and ethnic groups, highlighting the importance of dairy, milk, and their key components in CRC prevention for all races and ethnicities. These findings also underscore the significance of dairy consumption in CRC prevention strategies and call for further research into the specific components of dairy that contribute to these protective mechanisms.

CHAPTER 4. SPECIFIC AIM 1B – SURVIVAL ANALYSIS

4.1 Introduction

Dairy and milk, along with their essential components, might influence not only CRC risk but also survival. While direct link between dairy/milk consumption CRC survival have not been established, a meta-analysis of 7 prospective cohort studies on CRC mortality indicated an inverse association between overall dairy intake and death from CRC, albeit without distinguishing effects by dairy product types [119]. Prior research has predominantly focused on White populations, with limited attention to underrepresented groups. Given the observed heterogeneity in CRC survival among racial and ethnic minorities and unclear link between dairy/milk intake and CRC survival [120], it becomes imperative to examine the potential role of dairy/milk consumption in contributing to racial disparities in CRC survival. Therefore, this study aimed to investigate the impacts of dairy/milk intake and its key components on CRC survival within a multiethnic cohort and evaluate the heterogeneity of the effects across different racial and ethnic groups.

.

4.2 Methods

4.2.1 Study population

The MEC (introduced in the previous chapter) is a prospective cohort study established to investigate risk factors and disparities in cancer and other chronic diseases [127]. It includes >215,000 individuals from Los Angeles and Hawaii aged 45-75 at enrollment between 1993 and 1996. The cohort consist of mainly five racial and ethnic groups: Africa Americans, Japanese Americans, Native Hawaiians, Latinos, and Whites [127].

In the MEC, incident CRC cases were identified through linkage to the statewide Surveillance, Epidemiology, and End Results (SEER) Program tumor registries of Hawaii and California. For this study (Figure 4-1), a total of 6,575 incident CRC cases were included as our study population. These CRC patients were further excluded due to the same age of diagnosis and death or exit (n=45). The study was approved by the Institutional Review Boards of the University of Hawaii and the University of Southern California. All study participants provided informed consent at cohort entry.

4.2.2 Exposure assessment

Study participants' dietary intakes were assessed at baseline using a self-reported validated quantitative food frequency questionnaire (QFFQ) on over 180 food items [127-129]. The QFFQ was specifically developed for the MEC, including specific food items that were in traditional diets of particular racial and ethnic groups [127]. For each food and beverage item, participants were asked to report the frequency (eight categories for foods and nine for beverages), and portion size (three (in a few instances four) choices based on typical serving sizes for each single food or grouping of foods) of consumption. These data were then analyzed using the food composition table developed by the Cancer Research Center of Hawaii to compute individual daily food and nutrient intakes [127]. Additionally, the baseline questionnaire collected information on participants' use of vitamin and mineral supplements. Tumor characteristics including tumor stage, tumor grade, and primary treatment were obtained from the statewide SEER Program tumor registries in California and Hawaii.

In our analysis, we employed intake densities (intake per 1,000 kcal per day) for evaluating the consumptions of five dairy-related food items: total dairy products, milk, lactose, calcium, and

vitamin D. This was computed by dividing the absolute intake of the food item or nutrient by total daily caloric intake, and then multiplying by 1,000. Previous studies suggested that intake densities tended to yield better correlations than using absolute intake amounts, minimizing the potential measurement error from dietary recall [128, 130]. For vitamin D and calcium intake, we included total intakes from both foods and supplements, where supplemental vitamin D intake was from multivitamin supplements and supplemental calcium intake was from both calcium supplements and multivitamins.

4.2.3 Outcome assessment

Study participants were followed from time of CRC diagnosis until the earliest occurrence of death or the conclusion of the follow-up period on December 31, 2019. Information on mortality of study participants was attained from state death certificates in Hawaii and California and the National Death Index for the period since age at CRC diagnosis through December 31, 2019. For our study, we consider deaths specifically from CRC and deaths from all causes, including deaths from cardiovascular diseases and cancers, as well as deaths from other causes, such as accidents and suicides, as the outcomes of interest.

4.2.4 Statistical analyses

To examine the association of consumptions of dairy products, milk, lactose, calcium, and vitamin D on CRC-specific and all-causes death, Cox proportional hazards regression models with age as the time metric (age at CRC diagnosis to age of death or age of December 31, 2019) were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs). The proportional hazards

assumption was assessed using the Schoenfeld residuals. Intakes of dairy products, milk, lactose, calcium, and vitamin D were analyzed as cohort-specific quintiles and as continuous variables per standard deviation (SD). To assess trends in increasing quintiles, we ran a separate model where the quintile variable was treated as a continuous measure for each exposure of interest.

All regression models were adjusted for sex and ethnicity at baseline as strata variables. The following covariates were further included in the model: age at CRC diagnosis in 10 year age groups, family history of colorectal cancer (yes vs. no), history of intestinal polyps (yes vs. no), education (≤12th grade, vocational/some college, ≥college graduate), body mass index (BMI, <25, 25-30, ≥30 kg/m^2), smoking status and pack-years (never smoker, former smoker <20 pack-years, former smoker ≥20 pack-years, current smoker <20 pack-years, current smoker ≥20 pack-years), quintiles of physical activity (MET-hours/day), use of nonsteroidal anti-inflammatory drugs (NSAID, yes vs. no), log-transformed total energy intake (kcal/day), comorbidities including heart disease, stroke, and hypertension (not having any of the three diseases, had one of the disease, had ≥2 diseases), radiation treatment for CRC (yes, no, unknow), and chemotherapy for CRC treatment (yes, no, unknown). Participants' BMI and smoking status were treated as time-varying variables. That is, the most updated information of these variables from Qx1 to Qx4 before CRC diagnosis.

For all exposures, we performed stratified analyses by race and ethnicity group (Whites, African Americans, Native Hawaiians, Japanese Americans, Latinos), sex (men and women), age group (40-59, 60-69, ≥70 years old), tumor site (colon and rectum), tumor stage (in-situ/non-invasive, localized, regional, distant, unknown), and tumor grade (1, 2, 3, 4, 9). Heterogeneity across groups was assessed by likelihood ratio test comparing the model with and without interaction terms for the subgroup variable and exposure.

We performed all analyses using R version 4.3.0 (The R Foundation for Statistical Computing) [131]. All statistical tests were two-sided, with the level of significance pre-determined at $P < 0.05$.

4.3 Results

4.3.1 Baseline characteristics

After exclusion, our dataset consisted of 6,530 CRC cases, with 1,310 Whites (20.0%), 1,250 African Americans (19.0%), 409 Native Hawaiians (6.3%), 2,237 Japanese Americans (34.0%), and 1,324 Latinos (20.0%). During an average follow-up time of 7.7 years, a total of 1,814 CRC-specific deaths and 4,437 deaths from any causes were identified among eligible participants.

The baseline characteristics of the study population according to total dairy products intake is presented in Table 4-1. Compared to those with lower dairy products intake, participants with higher intake were more likely to be female, to be never smoker, and to use NSAIDs. A higher proportion of White CRC cases had the highest quintile of dairy intake compared to other racial and ethnic groups, and individuals who consumed more dairy products were older.

4.3.2 Total dairy products, milk, lactose, calcium, and vitamin D intakes and survival among the overall population

In our analysis, we consistently observed decreased, although non-significant, risks of all-causes death among CRC patients with the highest quintile of intakes of total dairy products (HR_{Q5vsQ1} = 0.97, 95% CI: 0.87-1.08), milk (HR_{Q5vsQ1} = 0.93, 95% CI: 0.84-1.03), lactose

(HR_{Q5vsQ1} = 0.93, 95% CI: 0.84-1.03), calcium (HR_{Q5vsQ1} = 0.97, 95% CI: 0.88-1.09), and vitamin D (HR_{Q5vsQ1} = 0.99, 95% CI: 0.89-1.09), compared to those in the lowest quintile of intakes (Figure 4-2A).

We found a similar pattern for CRC-specific mortality, where individuals in the highest quintile intakes of total dairy products (HR_{Q5vsQ1} = 0.89, 95% CI: 0.75-1.04), milk (HR_{Q5vsQ1} = 0.86, 95% CI: 0.73-1.01), and vitamin D (HR_{Q5vsQ1} = 0.93, 95% CI: 0.79-1.09) experienced better survival, compared to those in the lowest quintile of intakes, though these were also not statistically significant. However, those with the highest quintile of lactose intake showed a marginally significant reduced risk of CRC-specific death (HR_{Q5vsQ1} = 0.85, 95% CI 0.72-0.99). No significant association was found between the highest quintile of calcium intake and CRC-specific death (HR_{Q5vsQ1} = 1.00, 95% CI: 0.84-1.19) (Figure 4-2B).

4.3.3 Stratified analyses by race and ethnicity groups, sex, and age group

For all-cause deaths among CRC patients, stratified analyses by race and ethnicity groups showed inverse associations of total dairy products, milk, and lactose intake among Whites (total dairy products: HR_{Q5vsQ1} = 0.78, 95% CI: 0.61-1.00; milk: HR_{Q5vsQ1} = 0.79, 95% CI: 0.63-0.99; lactose: HR_{Q5vsQ1} = 0.75, 95% CI: 0.59-0.95), while a positive association of calcium intake (HR_{Q5vsQ1} = 1.42, 95% CI: 1.12-1.81) was noted among African Americans (Table 4-2 and Figure 4-3). Test of heterogeneity results revealed significant differences across race and ethnicity groups for all exposures except vitamin D, with notable heterogeneities for the intakes of total dairy products ($p_{heterogeneity}$ = 0.029), milk ($p_{heterogeneity}$ = 0.022), lactose ($p_{heterogeneity}$ = 0.014), and calcium ($p_{heterogeneity}$ < 0.001).

For CRC-specific deaths, stratified analyses by race and ethnicity groups revealed inverse associations of milk (HR_{Q5vsQ1} = 0.70, 95% CI: 0.49-0.99), lactose (HR_{Q5vsQ1} = 0.65, 95% CI: 0.45-0.93), and vitamin D (HR_{Q5vsQ1} = 0.60, 95% CI: 0.41-0.88) intakes with the risk of CRC-specific deaths among Whites. Among Japanese Americans, a higher milk intake was associated with a decreased risk of CRC-specific death (HR_{Q5vsQ1} = 0.72, 95% CI: 0.54-0.97). Conversely, African Americans with the highest quintiles of milk and calcium intake showed statistically significant increased risks of CRC-specific death (milk: HR_{Q5vsQ1} = 1.53, 95% CI: 1.05-1.88, calcium: HR_{Q5vsQ1} = 1.91, 95% CI: 1.33-2.75) (Table 4-3 and Figure 4-4). Notable heterogeneities across race and ethnicity groups were observed for lactose ($p_{heterogeneity}$ = 0.029) and calcium ($p_{heterogeneity}$ = 0.001) intakes.

Sex-specific analyses (Figure 4-5) revealed no significant association for all exposures in relation to the risk of all-causes death, while results from test of heterogeneity suggested significant heterogeneity between men and women for calcium intakes ($p_{heterogeneity}$ = 0.008). On the other hand, sex-specific analyses (Figure 4-6) showed an inverse association between lactose intake and CRC-specific death among men (HR_{Q5vsQ1} = 0.75, 95% CI: 0.60-0.95), with no significant associations observed among women. Significant heterogeneity between sexes was detected only for calcium intake ($p_{heterogeneity}$ = 0.005).

Analyses stratified by age groups showed significant associations between intakes of total dairy total dairy products (HR_{Q5vsQ1} = 0.77, 95% CI: 0.62-0.97), lactose (HR_{Q5vsQ1} = 0.78, 95% CI: 0.62-0.98), and calcium (HR_{Q5vsQ1} = 0.72, 95% CI: 0.57-0.91) and the risk of all-cause death among CRC patients for the 60-69 years age group (Table 4-4). Significant heterogeneity across age groups was observed only for calcium ($p_{heterogeneity}$ = 0.011). Additionally, significant associations were detected for intakes of total dairy products (HR_{Q5vsQ1} = 0.68, 95% CI: 0.49-0.95),

milk (HR_{Q5vsQ1} = 0.65, 95% CI: 0.46-0.91), lactose (HR_{Q5vsQ1} = 0.68, 95% CI: 0.48-0.95), and calcium (HR_{Q5vsQ1} = 0.64, 95% CI: 0.45-0.92) on CRC-specific death within the same age group, without significant heterogeneity across age groups (Table 4-5).

4.3.4 Stratified analyses by tumor site, tumor stage, and tumor grade

Analyses by tumor sites revealed no associations between any dairy related exposures and overall survival among both colon and rectal cancer patients (Figure 4-7). However, we observed inverse associations of higher intakes of total dairy products (HR_{Q5vsQ1} = 0.68, 95% CI: 0.48-0.96), milk (HR_{Q5vsQ1} = 0.62, 95% CI: 0.44-0.87), and lactose (HR_{Q5vsQ1} = 0.64, 95% CI: 0.45-0.89) with the risk of CRC specific deaths among rectal cancer patients (Figure 4-8).

Stratified analyses by tumor stage suggested an inverse association between higher calcium intake and the risk of all-cause death among patients with localized stage (HR_{Q5vsQ1} = 0.82, 95% CI: 0.68-0.98), while a positive association among those with distant stage (HR_{Q5vsQ1} = 1.41, 95% CI: 1.06-1.88) (Table 4-6). A reduced risk of CRC-specific deaths was observed for those with the highest quintile of milk intake (HR_{Q5vsQ1} = 0.74, 95% CI: 0.57-0.97), whereas an increased risk was detected for those with the highest quintile of calcium intake (HR_{Q5vsQ1} = 1.40, 95% CI: 1.02-1.92) (Table 4-7).

In the analyses stratified by tumor grade, similar patterns were observed between all-cause and CRC-specific deaths (Table 4-8 and Table 4-9). Higher intakes of lactose and calcium were inversely associated with both all-causes death (HR_{Q5vsQ1} = 0.84, 95% CI: 0.74-0.96, calcium: HR_{Q5vsQ1} = 0.82, 95% CI: 0.72-0.95) and CRC-specific (lactose: HR_{Q5vsQ1} = 0.79, 95% CI: 0.64-0.98, calcium: HR_{Q5vsQ1} = 0.80, 95% CI: 0.64-1.00) among CRC patients with grade II tumors.

Conversely, we observed positive associations of calcium consumption with both all-cause death (HR_{Q5vsQ1} = 1.53, 95% CI: 1.11-2.10) and CRC-specific death (HR_{Q5vsQ1} = 1.72, 95% CI: 1.13-2.62) and among CRC patients with grade III tumors.

4.4 Discussion

In this large multiethnic cohort consisting of five races and ethnicities, the consumption of total dairy products, milk, lactose, calcium, and vitamin D did not show statistically significantly associations with the risk of all-cause death among CRC patients. However, milk and lactose intake were marginally inversely associated with CRC-specific death. Notably, most effect estimates were consistently below 1, suggesting a potential protective role of dairy/milk intake against mortality among CRC patients.

While no significant associations were found between dairy product intakes and survival among all CRC patients, stratified analyses indicated that higher intakes were linked to better survival among specific subgroups: men, Whites, individuals aged 60-69, those diagnosed with rectal cancer, those with regional and distant stage CRC, and those at cancer grade II. Previous research has suggested that sociodemographic characteristics, cancer stage, and tumor site may modify the relationships between dietary exposures and CRC survival [122, 138]. To further elucidate these heterogeneous effects, future research should investigate why these associations vary across specific populations both in epidemiological and laboratory settings.

Dairy/milk consumption may contribute to improved survival among CRC patients due to their rich content of dietary calcium, vitamin D, and other beneficial components like conjugated linoleic acid, which has been shown to inhibit CRC cell growth in vitro [61, 139, 140]. Calcium is

hypothesized to enhance CRC prognosis not only through binding bile and fatty acids, reducing their toxicity, but also by influencing colonocyte behaviors such as proliferation, differentiation, and apoptosis [139, 141]. While most of our findings supported these mechanisms, not all associations reached statistical significance. However, it is noteworthy that our results indicated that higher milk and calcium intakes were associated with poorer survival among specific groups (e.g., African Americans). A possible explanation for this could be the association between milk consumption and elevated levels of insulin-like growth factor-I (IGF-I) [142], which has been linked to increased CRC mortality [143-145] due to its role in promoting cell proliferation and angiogenesis [146-148].

By leveraging the MEC study, our study was able to explore associations between dietary factors and CRC survival across multiple ethnic groups, especially underrepresented minorities. In addition, the use of baseline dietary information could help minimize the issues of recall bias and reverse causality. The availability of detailed information on tumor stage and differentiation grade allowed us to evaluate heterogenous effects of dietary intakes on survival across these groups. However, several limitations warrant mention. First, dietary intake was assessed before CRC diagnosis, which may not accurately reflect consumption patterns at diagnosis or thereafter, leading to potential misclassification bias. Second, the self-reported information on food intake might potentially lead to error for the measurement of exposure. Nevertheless, results from a calibration study indicated that the use of nutrient densities by taking into account the total energy intake improved the quality of the exposure assessment [128]. Lastly, since the outcome ascertainment was based on linkage to the state registries, we were unable to obtain CRC incidence and survival information for residents who moved to states other than California and Hawaii during the follow-up period.

In conclusion, our study did not find compelling evidence to support a significant link between the intakes of dairy, milk, lactose, calcium, and vitamin D and survival among CRC patients in the MEC study. However, we observed significant heterogeneity in these effects across different racial and ethnic groups. This highlights the need for further research with larger sample sizes to elucidate the impacts of these dietary factors on CRC progression and survival. Future studies should also consider the potential differences among diverse populations to better understand the role of nutrition in cancer outcomes.

CHAPTER 5. SPECIFIC AIM 2A – CANDIDATE SNPS ASSOCIATION STUDY

5.1 Introduction

Dietary factors play a crucial role in CRC pathogenesis [89], with evidence suggesting that dairy products, especially milk, may confer a protective effect against colorectal neoplasia [90]. A meta-analysis of 24 prospective cohort studies indicated a reduction in CRC risk associated with dairy and milk consumption [119]. Despite these findings, the mechanisms underlying this protective effect remain elusive. It is hypothesized that the beneficial impacts of dairy and milk on CRC prevention may be attributed to key constituents such as calcium [3, 4], vitamin D [5, 6], and lactose [7], implicating dairy/milk digestion pathways in CRC risk modulation.

Several genes have been identified to be associated with dairy/milk digestion pathways, including the calcium-sensing receptor (*CASR*) gene, the vitamin D receptor (*VDR*) gene, the lactase (*LCT*) gene, and the minichromosome maintenance complex component 6 (*MCM6*) gene [133, 149-152]. The *CASR* gene (chromosome 3: 136597196-136634013, GRCh37) provides instructions of producing calcium-sensing receptor proteins that are attached by the molecules of calcium to regulate the level of calcium in the blood [149]. The *VDR* gene (chromosome 12: 48235320-48298777, GRCh37) encodes vitamin D3 receptor that functions as a receptor for the secondary bile acid and lithocholic acid and controls the regulation of various metabolic pathways [150]. Additionally, the *LCT* (chromosome 2: 136545420-136594754) and *MCM6* (chromosome 2: 136597196-136634013) genes are central to lactose metabolism. The *LCT* gene encodes lactase-phlorizin hydrolase (LPH), which is a pivotal enzyme in the human body that helps hydrolyze lactose into glucose and galactose, thereby determining an individual's lactase persistence (LP) or lactase non-persistence (LNP) status [133]. The *MCM6* gene, a regulatory region 14kb upstream

from the *LCT* gene, is responsible for the transcriptional regulations of the *LCT* gene [133, 151, 152].

Research has explored the links between genes involved in dairy and milk digestion pathways and CRC risk. For example, studies suggested that both the *CASR* and *VDR* are associated with reduced risk of CRC, indicating their potential role in CRC prevention [153-155]. Although the direct association between the *LCT* and CRC risk has been less extensively studied, significant attention has been given to one single nucleotide polymorphism (SNP), rs4988235, and its relationship to CRC risk [156-160]. However, it's noteworthy that the existing research primarily focused on a limited selection of SNPs and has predominantly been conducted within European or White populations, underscoring a notable gap in the literature regarding the investigation of additional SNPs and the exploration of these genetic associations across diverse ethnic groups.

In this study, we conducted a prospective cohort study within in the Multiethnic Cohort (MEC), aiming to scrutinize the associations between candidate SNPs in four key genes (*LCT*, *MCM6*, *CASR*, and *VDR*) that are instrumental in dairy and milk digestion. Our study has a particular focus on ethnic minorities, which have historically been underrepresented in research. Our investigation was guided by three primary objectives: 1) to examine the overall and race-specific associations between rs4988235 and CRC risk; 2) to explore whether the effect of dairy and milk intake on CRC risk is modified by the genotypes of rs4988235; 3) to evaluate the associations between race-specific TagSNPs within the four genes (*LCT*, *MCM6*, *CASR*, and *VDR*) and CRC risk, thereby broadening our understanding of genetic basis of CRC across diverse populations.

5.2 Methods

5.2.1 Study population

The MEC is a prospective cohort that was established to investigate risk factors and disparities in cancer and other chronic diseases [127]. Individuals aged 45-75 years old living in Los Angeles and Hawaii from 1993 to 1996 were identified through driver's licenses records, voter registration lists, and Health Care Financing Administration (HCFA) files [127]. Selected individuals were then invited to participate in the study by a mailed invitation letter and a self-administered, comprehensive 26-page questionnaire which included information on demographic characteristics, anthropometric measurements, personal and family history of medical conditions, medication use, lifestyle factors, and dietary intake [127]. During the recruitment process, more than 215,000 individuals enrolled in the study, and the majority of them were from five racial/ethnic groups: African Americans, Japanese Americans, Nativa Hawaiians, Latinos, and Whites.

Ten years after the cohort entry (2001-2006), a sub-cohort of around 70,000 participants was established to examine potential roles of biomarkers and genetic factors in carcinogenesis and disease development [161]. Biospecimen samples, including 40cc blood and 20ml urine, were collected for each subject at a fating state. For individuals who refused the blood draw, mouthwash sample were collected [162]. After collection, biospecimen samples were then transported on ice to the University of Hawaii or University of Southern California laboratory for registration and processing. Genotyping was performed in blood specimens using various Illumina genotyping platforms, such as Illumina 1M, Illumina 660W, Illumina 610K, HumanOmni2.5-4 v1, HumanCytoSNP-12 v2, OmniExpress, and Multi-Ethnic Genotyping Array (MEGA) chip

(Illumina, San Diego, CA, USA). The study was approved by the Institutional Review Boards of the University of Hawaii and the University of Southern California. All study participants provided informed consent at cohort entry.

5.2.2 SNPs selection

We focused initially on a key SNP located on the gene *MCM6*, rs4988235, known for its close connection with milk consumption [156, 159, 163]. For additional SNPs in the *LCT*, *MCM6*, *CASR*, and *VDR* genes, we identified race specific TagSNPs to address the challenge of heterogenous linkage-disequilibrium (LD) patterns and allele frequencies across ethnic groups. These TagSNPs were selected utilizing the Tagger function in HaploView (www.broad.mit.edu/mpg/tagger, accessed on 13 April 2023) [164], capturing most of the common genetic variants in the four genes.

The targeted genomic regions were defined as 10kb before the gene start site and 5kb after the gene end site [165]. LD patterns and allele frequencies of SNPs within each genetic region were obtained using genetic reference populations of Utah residents with Northern and Western European ancestry (CEU), African Ancestry in Southwest US (ASW), Japanese in Tokyo, Japan (JPT), and Mexican Ancestry in Los Angeles, California (MXL) for Whites, African Americans, Japanese Americans, and Latinos, respectively. Given the absence of genetic reference panel for Native Hawaiian, we were unable to choose TagSNPs for this ethnic group.

Applying the selection criteria of a pairwise $r^2 \geq 0.8$ and a minor allele frequency ≥5%, we identified 120 TagSNPs for Whites (4 SNPs on *LCT*, 4 SNPs on *MCM6*, 41 SNPs on *CASR*, 71 SNPs on *VDR*), 279 TagSNPs for African Americans (37 SNPs on *LCT*, 16 SNPs on *MCM6*, 60

SNPs on *CASR*, 166 SNPs on *VDR*), 103 TagSNPs for Japanese Americans (16 SNPs on *LCT*, 7 SNPs on *MCM6*, 17 SNPs on *CASR*, 63 SNPs on *VDR*), and 139 TagSNPs for Latinos (14 SNPs on *LCT*, 10 SNPs on *MCM6*, 34 SNPs on *CASR*, 81 SNPs on *VDR*). Genotype information for all TagSNPs were successfully obtained for each ethnicity group (Figure 5-1).

5.2.3 Assessment of other variables

Information on participants' demographics, medical history, family history of CRC, dietary intakes and behaviors, use of alcohol and tobacco, and use of medications were obtained from the baseline questionnaire [127]. Particularly, dietary intakes were assessed using a validated quantitative food frequency questionnaire (QFFQ) on over 180 food items [127-129], and the daily food and nutrient intakes were calculated for each study participant using the food composition table developed by the Cancer Research Center of Hawaii [127]. For this study, intake densities, computed as the absolute intake of the food item or nutrient divided by total calories per day, and multiplied by 1,000, were used for food and vitamin consumptions.

5.2.4 Outcome assessment

Study participants were followed from cohort entry until time of CRC diagnosis, death, or end of follow-up on December 31, 2019, whichever was the earliest. Incident CRC cases were ascertained by linkage to the statewide Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in California and Hawaii using the International Classification of Diseases for Oncology, third edition (ICD-O-3) codes (C18.0-C18.7 for colon, and C19.9 and

C20.9 for rectum cancer). Information on mortality information of study participants was attained from state death certificates in Hawaii and California and the National Death Index (NDI).

5.2.5 Statistical analyses

The Cox proportional hazards regression models with age as the time metric were used to estimate the hazard ratios (reported as relative risks [RRs]) with 95% confidence intervals (CIs) for the associations between each selected SNP and CRC risk for each ethnic group. The allele tested in the genotyping was considered the effective allele, and the other allele was defined as the reference allele. We categorized the SNP assuming one of the three genetic models of inheritance: the additive model, the dominant model (zero effective allele + one effective allele vs. two reference alleles), and the recessive model (zero effective allele vs. one effective allele + two effective alleles). All regression models were adjusted for sex as strata variable, and we further adjusted for race/ethnicity as the strata variable for analyses in the overall population. To account for population stratification, the top ten ancestry-informative eigenvectors (PC1-PC10) from the principal component decomposition of the genotype matrix were also adjusted for in the models. P values were corrected for multiple hypotheses testing with the false discovery rate (FDR) procedures [166].

To investigate the gene-environment (G×E) interactions between dairy/milk consumption and rs4988235, stratified analyses were conducted to estimate the overall and race-specific hazard ratios (reported as RRs) and 95% CIs for the association between dairy/milk consumption and CRC risk across different genotypes of rs4988235. The regression models were adjusted for sex, ethnicity and 10-year age group at baseline as strata variables, and the following covariates assessed at baseline were further included as covariates: PC1-PC10, family history of colorectal

cancer (yes vs. no), history of intestinal polyps (yes vs. no), education (≤12th grade, vocational/some college, ≥college graduate), body mass index (BMI, <25, 25-30, ≥30 kg/m2), diabetes (yes vs. no), smoking status and pack-years (never smoker, former smoker <20 pack-years, former smoker ≥20 pack-years, current smoker <20 pack-years, current smoker ≥20 pack-years), alcohol consumption (never, <15, 15-30, ≥30 g/day), quintiles of physical activity (MET-hours/day), use of nonsteroidal anti-inflammatory drugs (yes vs. no), regular use of multivitamins (yes vs. no), log-transformed total energy intake (kcal/day), quintiles of red meat density intake (g/1,000 kcal/day), quintiles of processed meat density intake (g/1,000 kcal/day), quintiles of dietary fiber density intake (g/1,000 kcal/day), and quintiles of folate intake (μg DFE/day). Heterogeneity across groups was assessed by likelihood ratio test comparing the model with and without interaction terms for the subgroup variable and exposure.

All statistical tests were two-sided, with the level of significance pre-determined at $P < 0.05$. We performed all analyses using R version 4.1.3 (The R Foundation for Statistical Computing) [131].

5.3 Results

5.3.1 rs4988235 and CRC

Japanese Americans were excluded from this analysis due to low effective allele frequency (EAF) of rs4988235 in this population (Table 5-1 and 5-2). Our findings (Table 5-3 and Figure 5-2) indicated a reduced risk of CRC with the addition of each A allele in the overall population (RR = 0.86, 95% CI 0.77, 0.96, P value = 0.007). Individuals with AA or GA genotypes were also observed to have lower risk of CRC (RR = 0.83, 95% CI 0.71, 0.96, P value = 0.011) compared to those with GG genotypes.

Stratified analyses by ethnicities (Table 5-3 and Figure 5-2) revealed that the protective association of the A allele with CRC risk was specifically pronounced among Latinos (RR = 0.80, 95% CI 0.67, 0.97, P value = 0.020). Moreover, a reduction in CRC risk was observed among individuals with the AA genotype relative to those carrying GG or GA genotypes within both African American (RR = 0.76, 95% CI 0.58, 0.98, P value = 0.037) and Latinos (RR = 0.79, 95% CI 0.63, 0.98, P value = 0.030), but not in Whites. Notably, our analysis suggested statistically significant heterogeneity in the effects of rs4988235 on CRC risk across ethnic groups under all three genetic models ($P_{heterogeneity,additive}$ = 0.030, $P_{heterogeneity,dominant}$ = 0.013, $P_{heterogeneity,recessive}$ = 0.016).

5.3.2 G×E interaction

Employing either dominant or recessive models revealed no statistically significant associations between dairy/milk consumption and CRC risk across rs4988235 genotypes in the overall population (Table 5-4).

Race-specific evaluations yielded no statistically significant links between dairy intake and CRC risk across rs4988235 genotypes for any ethnic group. However, a notable association was discerned between milk consumption and increased CRC risk among African Americans with GA or AA genotypes (RR = 1.20, 95% CI 1.01, 1.43, P value = 0.038) and among Whites with GG genotype (RR = 1.37, 95% CI 1.05, 1.88, P value = 0.020).

Heterogeneity analysis suggested that the effects of milk intake on CRC risk were statistically significantly different among Whites with GG genotype compared to those with GA or AA genotype ($P_{heterogeneity.recessive}$ = 0.040).

5.3.3 Single TagSNP association

Under the additive genetic model, the T allele of rs3820790 on the *LCT* gene was found to be associated with increased risk of CRC (RR = 1.39, 95% CI 1.14, 1.69, P_{FDR} = 0.037) among African Americans (Table 5-5). While several SNPs demonstrated association with CRC risk across different ethnic groups, their P values did not withstand FDR correction.

When applying the dominant genetic model, after FDR adjustment, no SNP remained statistically significant in Whites. Among the African American population, several SNPs were linked to CRC risk, but only rs309132 on the *MCM6* gene retained statistical significance post-FDR correction (RR= 0.64, 95% CI 0.50, 0.83, P_{FDR} = 0.016). In Japanese Americans, the TT genotype of rs115319101 (on the *CASR* gene) was statistically significantly associated with a decreased CRC risk, maintaining significance after FDR adjustment (RR= 0.14, 95% CI 0.05, 0.45, P_{FDR} = 0.017).

Using the recessive genetic model revealed no SNP statistically significantly associated with CRC incidence in each ethnic population after FDR correction.

5.4 Discussion

Our study explored the link between the rs4988235 and CRC risk, alongside examining how dairy and milk consumption impacts CRC risk among individuals with different rs4988235 genotypes. We observed a statistically significant reduction in CRC risk associated with the A allele in the overall population. The GxE interaction analyses, utilizing an additive genetic model, indicated significant but marginally heterogeneous effects of dairy and milk intake on CRC risk across rs4988235 genotypes. Race-specific GxE interaction analyses revealed a notable

association between milk consumption and increased CRC risk specifically in African Americans with GA or AA genotypes and in Whites with the GG genotype. Moreover, this effect of milk consumption on CRC risk showed heterogeneity among Whites under a recessive genetic model.

Further investigation into TagSNPs within genes crucial for dairy/milk digestion identified a SNP, rs3820790 on the *LCT*, which was positively associated with CRC risk in African Americans. Additionally, the presence of two effective alleles of the SNP rs309132 on the *MCM6* was associated with a reduced risk of CRC in the same population. Conversely, the TT genotype of rs115319101 on the *CASR* demonstrated a protective effect against CRC risk. When employing a recessive genetic model, no significant associations were observed.

The SNP rs4988235 is a well-studied genetic variant associated with the expression or activity of lactase-phlorizin hydrolase (LPH). Despite mixed findings in prior research across various populations [157, 159, 160], the A allele (associated with increased level of LPH) was inversely associated with CRC risk among the overall population. Although the underlying biological mechanism linking LPH levels and CRC was not fully understood, this protective mechanism might be elucidated by dairy and milk digestion pathway [167-169], including effects from lactose [7], calcium [122, 170, 171], and vitamin D [123, 172].

Prior research suggested that CASR might play an important role in mediating the anticarcinogenic effect of calcium on CRC [173-178]. Several signaling pathways in cell growth and differentiation are activated when calcium binds to CASR, including promotion of E-cadherin expression, suppression of β-catenin/T cell factor activation, and activation of the p38 mitogen-activated protein kinase cascade [173, 179].

It is widely recognized that vitamin D induces differentiation and apoptosis in normal and tumor colonic cells through binding with VDR [176, 180]. Studies have identified polymorphisms

in the *VDR* gene were associated with the structure change and mRNA stability of VDR [181, 182], which consequently influence the binding of vitamin D and its anti-proliferative effects [183].

Our study has several strengths. Firstly, we leveraged the MEC to assess the effect of rs4988235 on CRC risk across various ethnic groups, offering a comprehensive view of candidate SNP analyses in populations often underrepresented in research. In addition, the availability of detailed dietary and lifestyle information facilitated our exploration of GxE interaction between dairy/milk intakes and rs4988235, shedding light on the genetic influences on dietary effects on CRC development. Furthermore, the prospective study design enabled us to examine genetic markers related to the etiology and early detection of CRC.

However, this study also encountered some limitations. First, the use of race-specific TagSNPs constrained our ability to draw comparisons across ethnic groups. Second, the lack of information on rare genetic variants limited our capability of probing gene-level associations with CRC risk. Lastly, Furthermore, the relatively small number of CRC cases within each ethnic group might have limited the power of our race-specific analyses. Nonetheless, our findings may serve as valuable descriptive and hypothesis-generating insights.

In conclusion, our research demonstrated a protective effect of the rs4988235 on CRC risk among the overall population. Furthermore, we delved into the relationships between TagSNPs within genes implicated in dairy/milk digestion and CRC risk across ethnic groups. These findings underscore the importance of conducting further research with more extensive cohorts to enhance our understanding of the intricate mechanisms by which dairy/milk consumption may influence colorectal carcinogenesis.

CHAPTER 6. SPECIFIC AIM 2B – MENDELIAN RANDOMIZATION STUDY

6.1 Introduction

Lactase-phlorizin hydrolase (LPH) is a pivotal enzyme in the human body that helps hydrolyze lactose, the main carbohydrate in milk, into glucose and galactose [133, 184]. The reduced expression or activity of LPH, known as lactase non-persistence (LNP), leads to a clinical condition called lactose intolerance, in which milk and other dairy products cannot be properly digested. Individuals with lactose intolerance experience symptoms such as abdominal pain, bloating, diarrhea, nausea, and vomiting after consumption of milk and other dairy products [133, 184]. Genetically, LPH is encoded by the lactase gene (*LCT*) on chromosome 2. Genetic expression of *LCT* has been found to be regulated by single nucleotide polymorphisms (SNPs) located on the gene *MCM6,* a regulatory region 14 kb upstream from the *LCT* gene [133, 151, 152]. Specifically, the SNP rs4988235 on *MCM6* confers the LNP phenotype.

Diminished LPH levels or activity, leading to lactose maldigestion, are linked to decreased calcium [185] and vitamin D intake [186], along with a reduced abundance of beneficial gut bacteria, Bifidobacterium [187]. Observational studies have reported that reduced calcium [188, 189] and vitamin D [121, 190] intake are associated with increased CRC risk, suggesting protective roles of calcium and vitamin D in CRC development. In addition, clinical studies have shown that dietary intake of Bifidobacterium modulates gut microbiota towards CRC prevention [191]. Given LPH's pivotal role in milk digestion and its downstream influence on crucial nutrient absorption and gut microbiota composition, it may also have a significant impact on CRC susceptibility. In addition, LPH could potentially serve as a potential candidate biomarker for CRC risk stratification or a druggable target for CRC treatment, as several other circulating proteins associated with CRC

risk have been implemented for these purposes [192-195]. Yet, the specific role of LPH in the development of CRC remains unclear, highlighting the need for detailed studies exploring this potential association.

There has been no research directly studying the relationship between LPH levels and CRC risk in the medical literature. Instead, previous epidemiologic studies have investigated this relationship using LNP status, LPH-related SNPs, and dietary milk intake as proxies for LPH levels [196-202]. However, these studies have several limitations, including exposure misclassification, residual confounding, and reverse causality. For instance, lactase persistence/non-persistence status was often binarily defined by individual genotype. However, the negative impacts of lactose maldigestion among lactase-non-persistent individuals are actually determined by continuous residual LPH expression levels [133, 203, 204]. In addition, CRC patients undergoing adjuvant 5-fluorouracil chemotherapy can develop secondary lactose intolerance due to gastrointestinal damage [205, 206], disrupting small intestine enzyme and transporter functions [207]. Consequently, the potential for reverse causation (i.e., CRC leading to reduced LPH levels and thus milk intake) remains plausible.

To circumvent these challenges, we utilize Mendelian Randomization (MR) analysis, an innovative method that employs genetic variants as instrumental variables (IVs) for LPH levels [208]. The random assignment of these variants during meiosis helps mitigate confounding bias and reverse causality issues, offering a robust means to explore potential causality [208-210]. While conventional genome-wide MR studies encompass both *cis*-variants (i.e., located near the gene of interest) and *trans*-variants (i.e., often located on different chromosomes), there is a rising trend in *cis*-MR studies that exclusively use *cis*-variants as IVs, especially in contexts where protein expression is a key consideration [211-215]. The appeal of *cis*-MR studies has grown due

to their potential for drug target identification and validation [213, 215]. In our study, we focus on continuous LPH levels as the exposure, selecting both *cis*- and *trans*-variants associated with LPH levels from a large-scale genome-wide association study (GWAS). We then use sets of (1) only *cis*-variants and (2) combined *cis*- and *trans*-variants as separate IVs in our MR analyses.

This study leverages MR to probe the potential causal influence of genetically determined elevated LPH levels on the risk of CRC and its subtypes, namely colon and rectal cancer. Utilizing publicly accessible summary-level GWAS data from three large-scale, independent cohorts of European ancestry, we seek to enhance our understanding of the genetic underpinnings of CRC and inform future preventive strategies.

6.2 Methods

6.2.1 Study design

Our study utilized a two-sample MR approach, using genetic variants as IVs, to investigate whether there is a causal relationship between elevated LPH levels and the risk of CRC. The MR analyses rest on three fundamental assumptions: (1) the Relevance assumption establishes that the genetic IVs are associated with the exposure (e.g., LPH levels); (2) the Independence assumption states that the genetic IVs have no correlation with potential confounders; and (3) the Exclusion restriction assumption dictates that the genetic IVs could only affect the outcome of interest (e.g., CRC) via the exposure (i.e., no horizontal pleiotropy where genetic IVs can affect multiple outcomes) [216].

The schematic overview of our study design is presented in Figure 6-1. Our process commenced with the selection of genetic instruments for LPH levels from the GWAS Catalog

[217], followed by the extraction of summary statistics of these selected genetic instruments from prior GWAS of CRC risk performed in three independent cohorts: the FinnGen Study, the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) Atlas Project, and the Pan-UK Biobank. Each cohort had prior ethical approvals, negating the need for additional approvals for this study.

To assess the causal effect of elevated LPH levels on CRC risk, we primarily conducted two-sample MR analyses in each cohort using a *cis*-variant for LPH levels. The results from the three cohorts were subsequently integrated using meta-analysis. For validation, MR analyses incorporating all variants (*cis*- + *trans*-) were also performed. Further, this identical workflow was used for the analysis of CRC subtypes (i.e., colon and rectal cancer). Our study followed the Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization (STROBE-MR) reporting guidelines [218].

6.2.2 Genetic instruments

Genetic instruments for LPH levels were retrieved from the NHGRI-EBI GWAS Catalog, a large-scale open GWAS database collaboratively developed by the European Bioinformatics Institute (EBI) and the Human Genome Research Institute (NHGRI) with over 24,000 traits (www.ebi.ac.uk/gwas, accessed on 2 May 2023) [217]. Our focus was on the Fenland Study data from the GWAS Catalog, which offers the largest and most recent GWAS of LPH levels (GWAS Catalog accession ID: GCST90248315). Summaries of the study are listed in Table 6-1. The Fenland Study consisted of 10,708 genotyped participants of European ancestry who were recruited from general practice surgeries in the Cambridgeshire region of the UK from 2005 to 2015 [219]. Genotyping was conducted using three different arrays (Affymetrix UK Biobank

Axiom array [Affymetrix, Santa Clara, CA, USA], Illumina Infinium Core Exome 24v1 [Illumina, San Diego, CA, USA], and Affymetrix SNP5.0 [Affymetrix, Santa Clara, CA, USA]), and levels for each protein target were measured using the rank-based inverse normal-transformed aptamer abundance method [219]. GWAS analysis was then performed using the transformed protein levels, with the residuals used as input for the genetic association analyses [220]. The beta coefficients for each protein target, representing one standard deviation (SD) change in normalized plasma abundance of protein per effect allele of the SNPs, were estimated, adjusting for age, sex, sample collection site, and the first ten principal components [219]. Our study selected SNPs associated with LPH levels at the genome-wide significant threshold of $p < 5 \times 10^{-8}$ [221]. Correlated SNPs were excluded according to measures of linkage disequilibrium (LD) $r^2 < 0.1$ and minor allele frequency (MAF) > 0.01 based on the European populations from the 1000 Genomes phase 3 reference panel using the SNPclip online tool (https://ldlink.nih.gov/, accessed on 4 May 2023) [222].

Following exclusions, our analysis included four variants, one *cis*-variant (rs4988235) and three *trans*-variants (rs516246, rs532436, and rs641476), that were used as genetic instruments to genetically predict LPH levels. The characteristics of the genetic instruments for elevated LPH levels included in our study are presented in Table 6-2. Four independent SNPs associated with *MCM6* (rs4988235), *FUT2* (rs516246), *ABO* (rs532436), and *GAREM1* (rs641476) were selected based on the genome-wide significance level ($p < 5 \times 10^{-8}$) and LD-based pruning ($r^2 < 0.1$). Overall, the four selected SNPs accounted for 36.42% of the observed variance in elevated LPH levels, with the *cis*-variant rs4988235 contributing the majority of the variance.

To assess the strength of the genetic instruments selected, we calculated R^2 (the percent variation in LPH levels explained by the genetic instrument) and the Cragg–Donald *F*-statistics

(the strength of the association between the genetic instrument and LPH levels) for each LPH-associated SNP using the formula: $R^2 = \beta^2 \times 2 \times EAF \times (1 - EAF)$ and $F = R^2 \times (N - 2)/(1 - R^2)$, where EAF denotes the effect allele frequency of the SNP and N represents the sample size of the exposure GWAS [223, 224]. A *F*-statistic greater than 10 indicates strong genetic instruments for the MR analyses [225]. The F-statistics for the four SNPs ranged from 87.01 to 5340.06, underscoring their strength as genetic instruments for MR analyses.

6.2.3 Outcome assessment

A summary of the GWAS datasets for CRC is presented in Table 6-1. Summary-level data pertaining to the association of SNPs with CRC were obtained from three publicly available GWAS: (1) the FinnGen Study (available at https://www.finngen.fi/en/access_results, accessed on 2 May 2023); (2) the PLCO Atlas Project (available at https://exploregwas.cancer.gov/plco-atlas/#/gwas/summary, accessed on 2 May 2023); and (3) the Pan-UK Biobank (available at https://pan.ukbb.broadinstitute.org/, accessed on 2 May 2023). Detailed information for these studies was reported in the original publications [226-228]. CRC cases were identified by: (1) ICD-10 codes C18–C20 in the FinnGen Study; (2) ICD-O-2 codes 180, 182–189, 199, 209, 212, and 218 in the PLCO Atlas; and (3) self-report through verbal interview with a trained nurse in the Pan-UK Biobank [226-228]. To minimize population stratification bias, only GWAS results from individuals of European ancestry were included.

All genetic association estimates between the SNPs and CRC were calculated using logistic regression comparing cases and controls, adjusting for age, sex, and genetic principal components (the first ten in the FinnGen consortium and Pan-UK Biobank, and the first twenty in the PLCO Atlas). In addition, some studies also included study-relevant covariates in their logistic regression

models, such as age^2 (in the Pan-UK Biobank), study center (in the PLCO Atlas), and genotyping batch (in the FinnGen Study).

We extracted estimates (e.g., effective alleles, beta coefficients, standard errors, and p-values) for the associations between the selected genetic instruments and the risk of CRC and CRC subtypes (colon and rectal cancer) from the FinnGen, PLCO Atlas, and Pan-UK Biobank GWAS. For SNPs not available in these GWAS, we identified proxy SNPs in linkage disequilibrium ($r^2 > 0.7$ within a ±500,000 base pairs window) based on the European populations from the 1000 Genomes phase 3 reference panel utilizing the LDProxy online tool (https://ldlink.nih.gov/, accessed on 4 May 2023) [222]. All four genetic instruments were found in the PLCO and Pan-UK Biobank datasets. Rs532436 was not available in the FinnGen dataset, and thus we used the proxy SNP rs635634, which was in high linkage disequilibrium with rs532436 ($r^2 = 0.99$). Details of the genetic association between the SNPs and the risk of CRC are presented in Table 6-3.

6.2.4 Statistical analyses

Effect alleles were defined for each SNP as the allele contributing to increased LPH levels. We performed strand alignment to harmonize the relationships between genetic instruments and CRC, as well as between LPH levels and CRC for the same allele. We primarily performed the Wald ratio two-sample *cis*-MR using rs4988235 as the genetic instrument. For validation, we then employed the inverse-variance weighted (IVW) two-sample MR across all four genetic instruments. The IVW method assumes that all SNPs are valid instruments and that horizontal pleiotropic effects are absent or balanced, constraining the intercepts to zero [229]. The Cochran's Q statistic and I^2 index were used to test for the presence of heterogeneity, which is an indicator of

whether the IVW estimates on LPH levels and CRC risk are different across different genetic variants [230].

Further enhancing the robustness of our investigation, we performed a series of sensitivity MR analyses, including penalized IVW, robust IVW, penalized robust IVW, MR–Egger, weighted median, mode-based estimation, and MR–Lasso. The robust IVW method uses robust regression to downweight outliers, while the penalized IVW method improves the robustness of the estimates by penalizing the weights of genetic instruments with heterogeneous causal estimates for the outcome [231, 232]. The penalized robust IVW method further provides robustness both to outliers and to data points with high leverage through robust regression [231]. The MR–Egger method allows the inclusion of horizontal pleiotropic SNPs and provides a bias-corrected exposure-outcome effect estimate, with a deviating intercept indicating mean pleiotropic effects [233]. Despite relaxing the exclusion restriction assumption, MR–Egger mandates the InSIDE (Instrument Strength Independent of Direct Effect) assumption, which requires that the associations of the genetic instruments with the exposure and the direct effects of the genetic instruments on the outcome are independent [234]. Consequently, we also incorporated MR analyses that do not require the InSIDE assumption (e.g., weighted median and the mode-based estimation) [233, 235]. To assess the distortions of the IVW estimate from any heterogeneity or horizontal pleiotropy, MR–Lasso was used to detect and remove pleiotropic outliers [236].

The effect estimates of genetically predicted LPH on CRC and its subtypes were reported as odds ratios (ORs), along with their 95% confidence intervals (CIs), per one SD increase in normalized plasma abundance of LPH. Each SNP's association was plotted against its corresponding effect on CRC risk. To evaluate the potential influence of a single SNP on MR results, iterative leave-one-out analyses were executed [234].

All of the primary and sensitivity MR analyses were conducted separately within each of the three outcome data sources (i.e., FinnGenn, PLCO Atlas, and Pan-UK Biobank). For comparison and consolidation of effect estimates from varying data sources, we utilized meta-analysis with fixed effects models to integrate the IVW estimates across the three cohorts. The degree of heterogeneity between the IVW estimates was quantified using the I^2 index and Cochran Q statistics [237].

All statistical tests were two-sided, with the level of significance predetermined at $p < 0.05$. We performed all analyses using R version 4.1.2 (The R Foundation for Statistical Computing) [131]. We used the "MendelianRandomization" package [238] for MR analyses and the "meta" package for meta-analyses [239].

6.3 Results

6.3.1 FinnGen Dataset

The FinnGen GWAS summary statistics on CRC consisted of 6509 CRC cases and 287,137 controls. Using only the *cis*-variant rs4988235 as the genetic instrument, the FinnGen dataset showed that genetically determined higher levels of LPH were associated with decreased odds of CRC (OR per SD higher normalized plasma abundance of LPH: 0.91 [95% CI, 0.88–0.95], $p < 0.001$) (Table 6-4). The IVW estimate from the MR analysis using all LPH-associated genetic variants showed similar results as the *cis*-MR analysis (OR: 0.92 [95% CI, 0.88–0.95], $p < 0.001$) (Table 6-4 and Figure 6-2A). Results for sensitivity analyses were presented in Table 6-4 and Figures 6-2 – 6-4. Little heterogeneity across SNPs was evidenced by Cochran's Q statistics (Q = 2.5, $p = 0.482$), and sensitivity analyses produced consistent results. There was no evidence of horizontal pleiotropy according to the MR–Egger results ($P_{Egger\text{-}intercept} = 0.552$). Based on the

leave-one-out analysis (Figure 6-4A), the primary influence on the effect came from the SNP rs4988235 on *MCM6*, which is the most well-characterized SNP responsible for LPH synthesis and the only *cis*-variant selected in the GWAS for LPH levels [133, 240].

6.3.2 PLCO Dataset

The PLCO GWAS dataset included 2065 CRC participants and 67,500 controls. The PLCO dataset illustrated a non-significant association between genetically determined elevated LPH levels and CRC risk in the *cis*-MR (OR: 0.92 [95% CI, 0.85–1.00], $p = 0.063$) (Table 6-4). Similar results were found in the MR analysis including all genetic instruments (OR: 0.94 [95% CI, 0.85–1.03], $p = 0.170$), whereas the confidence interval was slightly wider than that in the *cis*-MR (Table 6-4 and Figure 6-2B). Table 6-4 and Figures 6-2 – 6-4 show the results from the sensitivity analyses. With penalized robust IVW, the association became significant (OR: 0.94 [95% CI, 0.90–0.98], $p = 0.002$), indicating the presence of potential outliers. Results from the MR–Egger, weighted median, and mode-based estimation analyses did not provide strong evidence for horizontal pleiotropic effects among the SNPs (Table 6-4). The leave-one-out analysis plot suggested that the MR IVW estimates were largely influenced by rs4988235, which was consistent with results in the FinnGen dataset (Figure 6-4B).

6.3.3 Pan-UK Biobank Dataset

There were 592 CRC cases and 419,881 controls in the Pan-UK Biobank. The *cis*-MR Wald ratio did not provide evidence supporting the effect of genetically determined elevated LPH levels on CRC risk in the Pan-UK Biobank dataset (OR: 1.00 [95% CI, 0.87–1.14], $p = 0.971$),

and this result was similar with the IVW estimate including both *cis*- and *trans*-variants (OR: 1.03 [95% CI, 0.83–1.27], $p = 0.812$) (Table 6-4 and Figure 6-2C). In addition, the intercept for the MR–Egger analysis was not significantly different from zero ($P_{Egger\text{-}intercept} = 0.712$), indicating little evidence of horizontal pleiotropic effects in the selected genetic instruments. Sensitivity analyses mirrored the IVW estimate, with the leave-one-out analysis affirming rs4988235's substantial impact (Figure 6-4C).

6.3.4 Meta-Analysis Combining FinnGen, PLCO, and Pan-UK Biobank Results

Meta-analysis combining the *cis*-MR estimates from FinnGen, PLCO, and Pan-UK Biobank showed an inverse association between genetically predicted elevated LPH and CRC risk (OR: 0.92 [95% CI, 0.89–0.95], $p < 0.001$), with no discernible heterogeneity in the effect across the three datasets ($I^2 = 0\%$, $P_{cochran\text{-}Q} = 0.470$) (Figure 6-5). Similarly, the combined IVW estimate for MR studies utilizing all four genetic variants showed a sightly attenuated association (OR: 0.93 [95% CI, 0.89–0.96], $p < 0.001$). We did not find strong evidence indicating heterogeneity across the three datasets ($I^2 = 0\%$, $P_{cochran\text{-}Q} = 0.554$) (Figure 6-5).

6.3.5 CRC Subtype-Specific MR Analyses

CRC subtype-specific MR analyses are reported in Tables 6-5 – 6-7, and sensitivity analysis results are presented in Figures 6-6 – 6-11. The *cis*-MR analysis of the FinnGen dataset revealed an inverse association between genetically predicted elevated LPH levels and colon cancer risk (OR: 0.92 [95% CI, 0.87–0.97], $p = 0.001$, Table 6-5). Although similar Wald ratio estimates were observed in the PLCO and Pan-UK Biobank datasets, statistical significance was

not reached, potentially due to small sample sizes (PLCO: OR 0.93 [95% CI, 0.85–1.02], p = 0.144; Pan-UK Biobank: OR 0.95 [95% CI, 0.86–1.05], p = 0.285, Table 6-5). Combining the *cis*-MR results from the three datasets, the meta-analyzed estimate (Table 6-7 and Figure 6-12) suggested a significant association between genetically predicted higher LPH levels and decreased risk of colon cancer (OR: 0.92 [95% CI, 0.89–0.96], $p < 0.001$). Results from the MR analyses utilizing all four genetic instruments further confirmed the association with similar estimates but wider confidence intervals (meta-analyzed OR: 0.93 [95% CI, 0.89–0.97], $p < 0.001$) (Tables 6-5 – 6-7 and Figure 6-12).

With respect to rectal cancer, the FinnGen dataset indicated an inverse association between genetically predicted elevated LPH levels and rectal cancer susceptibility when using the single *cis*-variant rs4988235 (OR: 0.91 [95% CI, 0.85–0.97], p = 0.005, Table 6-6). The PLCO dataset suggested a negative but non-significant estimate (OR: 0.86 [95% CI, 0.70–1.06], p = 0.172, Table 6-6). Results from the Pan-UK Biobank dataset, however, demonstrated an incongruous positive, albeit non-significant, estimate (OR: 1.13 [95% CI, 0.91–1.40], p = 0.267, Table 6-6). The subsequent meta-analysis (Table 6-7 and Figure 6-13) suggested an inverse association between elevated LPH levels and rectal cancer risk (OR: 0.92 [95% CI, 0.87, 0.98], p = 0.0083), and moderate heterogeneity was observed across the datasets (I^2 = 50%, $P_{cochran\text{-}Q}$ = 0.136). Further MR analyses including both *cis*- and *trans*-variants showed consistent results (Table 6-6, 6-7 and Figure 6-13), indicating the robustness of our *cis*-MR estimates.

6.4 Discussion

In this study, we leveraged summary-level statistics from three large-scale GWAS of European ancestry and employed a two-sample MR framework to investigate the potential causal

relationship between LPH levels and CRC risk using both *cis*-variants and all genetic instruments (*cis*- + *trans*-). The results from the *cis*-MR analysis provided genetic evidence suggesting an inverse causal association between elevated LPH levels and CRC risk. This finding was consistent and validated by MR analyses using both *cis*- and *trans*-variants. Further MR analyses by CRC subtypes indicated that this causal relationship seemed applicable to both colon cancer and rectal cancer.

While the FinnGen dataset showed a significant inverse association between genetically predicted elevated LPH levels and CRC risk, the findings from the PLCO and Pan-UK Biobank datasets were not statistically significant, likely due to insufficient statistical power attributed to smaller sample sizes and lower case-to-control ratios. We confirmed this hypothesis through power calculations, revealing 85% power in the FinnGen dataset to detect a 6% change in the odds of CRC, compared with just 39% and 15% power in the PLCO and Pan-UK Biobank datasets, respectively. Therefore, to bolster statistical power, we conducted a meta-analysis of the separate MR analyses within each of the three cohorts. Subgroup analyses for colon and rectal cancer revealed similar trends. With a relatively small number of rectal cancer cases in both the PLCO (320 cases) and Pan-UK Biobank (301 cases) datasets, these analyses were likely hindered by limited statistical power.

It is worth noting that the Pan-UK Biobank dataset showed a higher number of colon cancer cases compared to overall CRC cases. This discrepancy might be explained by the case identification method in the Pan-UK Biobank, which is reliant on self-reported cancer diagnoses and therefore subject to potential measurement error. Although more accurate cancer case ascertainment methods might be employed in individual-level UK Biobank datasets, such information was not available in the publicly accessible summary statistic data that we utilized.

The potential underlying biological mechanisms linking elevated LPH levels with reduced CRC risk warrant further exploration. Evidence suggests that lactase persistent individuals typically consume more milk than their lactase-non-persistent counterparts [241, 242]. Given the known impact of milk consumption on CRC risk [60, 66, 201, 202, 243-246], it is plausible that the protective effect of LPH on CRC risk is partially mediated through catalyzed products of milk [169, 247, 248] and key milk components, namely calcium [188, 189, 249] and vitamin D [121, 123]. Other milk-derived compounds, such as butyric acid, conjugated linoleic acid, sphingolipids, and lactoferrin [61, 126, 250], also contribute to the protective effect of LPH. Moreover, the effect of LPH on gut microbiota diversity could also play a role in modifying CRC risk [187].

Calcium and vitamin D, abundant components of milk, have been recognized for their multifaceted roles in CRC prevention. Calcium's protective effects can be attributed to its capacity to bind secondary bile acids and ionized fatty acids, thereby reducing their toxicity on colonocytes and inhibiting mucosal proliferation [133]. In addition, it may activate certain signaling pathways via the calcium-sensing receptor (CaSR), including E-cadherin expression promotion, beta-catenin/T cell factor activation suppression, and p38 mitogen-activated protein kinase cascade activation [249]. There is also evidence linking calcium to a lower risk of mutations in the *KRAS* gene, a significant determinant in the carcinogenesis of CRC [133]. Vitamin D modulates molecular pathways relevant to CRC development, including the downregulation of the *COX-2* gene and the upregulation of 15-hydroxyprostaglandin dehydrogenase (15-PDGH), leading to a reduction in local prostaglandin levels and hence inhibiting cancer cell survival [121]. Moreover, it interferes with β-catenin-mediated gene transcription, primarily by promoting Vitamin D receptor (VDR) binding to β-catenin, emphasizing its suppressive role on tumor growth [123].

Other milk compounds, such as butyric acid, conjugated linoleic acid, and lactoferrin, may also contribute to CRC prevention [61, 126, 250]. These components have shown various anti-carcinogenic effects in in vitro and animal studies, ranging from suppressing proliferation to enhancing immune function [61, 126, 250-253]. Additionally, LPH levels might impact CRC risk by modifying the gut microbiota. For instance, studies have linked increased LPH levels to a greater abundance of *Bifidobacterium* [187], which is known for augmenting antitumor immunity and facilitating the efficacy of immunotherapy [254]. In this context, our MR findings provide genetic support for this biological rationale, underscoring the relevance of LPH metabolism in CRC prevention.

While no study has directly investigated the effects of LPH on CRC risk, our findings are comparable to prior epidemiologic studies investigating CRC risk associated with LNP status or genetic instruments for milk consumption. Two studies conducted in Finnish and Hungarian populations observed a statistically significant increased risk of CRC risk among LNP individuals, with ORs reported at 1.40 and 4.04, respectively [197, 199]. Although other studies conducted in British, Spanish, and Italian populations observed no association between LNP and CRC, these had limited statistical power due to small sample sizes (44-283 CRC cases) [197, 200]. Furthermore, two other studies using rs4988235 as a genetic instrument for milk consumption found that genetically predicted milk intake was associated with a reduced risk of CRC (reported ORs of 0.89 and 0.95) [201, 202]. This is similar to the effect size observed in our current analysis for genetically predicted LPH levels and CRC risk (OR 0.92) using the same *cis*-variant (rs4988235).

Our findings on the protective effect of LPH against CRC development highlight its potential role in CRC prevention and treatment. Specifically, LNP individuals identified through

screening methods, such as lactose breath tests or genetic testing of the rs4988235 polymorphism, could benefit from specific dietary recommendations (e.g., calcium or vitamin D supplements) to mitigate CRC risk. Such targeted interventions could not only enhance individual health outcomes, but also contribute to more personalized and potentially cost-effective approaches to CRC risk management. Furthermore, LPH can perhaps serve as a novel therapeutic target for CRC, providing potential avenues for CRC treatment strategies.

Our study has several notable strengths. We implemented a *cis*-MR approach as our primary analysis, which not only mitigates biases such as residual confounding and reverse causation that typically complicate observational studies, but also minimizes potential horizontal pleiotropy. The use of the *cis*-variant (rs4988235), located within the *MCM6* gene and in close proximity of the LPH-encoded gene *LCT*, ensures that the observed effects on CRC can be attributed solely to variations in LPH expression, given the regulatory role of rs4988235 [133, 151, 152]. This study's findings suggest the potential therapeutic role of LPH for CRC, underscoring its clinical significance. Furthermore, the utilization of all genetic variants (*cis*- + *trans*-) served as a validation of the *cis*-MR approach and allowed for a series of sensitivity analyses. These included various MR methods, such as weighted median, mode-based estimation, and MR–Egger, which helped to examine the potential effects of horizontal pleiotropy from selected genetic instruments. Previous studies may have also been subject to several limitations, such as binary definitions of lactase persistence status and potential violations of the relevance assumption of MR [196-202]. Our study addressed these issues by using genetically predicted continuous LPH levels as the exposure and selecting genetic instruments directly associated with LPH levels from large-scale GWAS datasets. By our calculations, the SNPs selected in our study explained 36.43% of the variance in LPH levels, with rs4988235 displaying a strong association with LPH levels

(variance explained: 33.28%). In addition, by using distinct GWAS datasets for LPH levels (exposure) and CRC (outcome) in our two sample MR analyses, we also reduced the potential inflation of bias associated with weak instrument variables [255]. Furthermore, we accounted for heterogeneity introduced by specific SNPs with outlier causal estimates by employing penalized IVW and MR–Lasso estimations. The application of leave-one-out analyses also helped us verify the consistency of estimates across genetic instruments and determine whether specific SNPs substantially influenced our causal estimates. We further integrated three large-scale, independent GWAS datasets into our MR analyses and meta-analyses, ensuring sufficient sample sizes for the outcome. Lastly, by conducting MR analyses across different CRC subtypes, we offered a comprehensive view of LPH's potential biological role in various tumor locations.

However, our study has some limitations. We acknowledge that the limited number of CRC cases in the Pan-UK Biobank, and especially the smaller number of rectal cancer cases across all three cohorts, could have constrained our study's statistical power. To mitigate this limitation, we employed meta-analysis techniques, maximizing data utilization to yield more robust results and inferences. A further limitation of our study lies in our exclusive inclusion of individuals of European descent. It is worth noting that the prevalence of lactase non-persistence significantly varies across populations; it is highest in East Asians (for example, 85% in Chinese and 100% in South Koreans) and lowest in individuals of Northern European descent (for instance, 8% in Finns and 7.8% in Swedes) [152]. Consequently, these variations in LPH levels among different populations restrict the generalizability of our findings to individuals of European ancestry. Future research should include other populations and delve into sex-specific causal estimates for a more nuanced understanding of LPH and CRC.

This study, to our knowledge, is the first to explore the causal relationships between LPH levels and the risk of CRC using MR analyses with large-scale GWAS datasets. The findings underscore the importance of LPH and its downstream effects in influencing CRC risk. Moreover, it may provide new insights into preventive strategies and a potential drug target for interventions aimed at reducing the burden of CRC. Further studies are necessary to better delineate these mechanisms and validate the potential of LPH as a biomarker for CRC risk.

Our study suggests that there is an inverse causal relationship between LPH levels and CRC risk. These findings, consistent across cohorts for both colon and rectal cancers, highlight a potential causal role for LPH as a preventative biomarker. Further study is needed to clarify the mechanisms and extend these findings to other populations.

CHAPTER 7. SPECIFIC AIM 3 – PROTEOME-WIDE MENDELIAN RANDOMIZATION STUDY

7.1 Introduction

Despite notable advancements in medical technology and oncological research, the intricate nature of CRC's molecular characteristics and pathogenesis creates challenges in the development of tailored prevention and treatment options. Recent years have seen the development of several drugs targeting specific molecules or signaling pathways associated with cancer cells [256-260]. However, most of the biomarkers or pathways currently targeted address cancer types uniformly, neglecting the heterogeneity among cancer types. This limitation, coupled with emerging drug resistance [261, 262], poses challenges in CRC control or eradication.

Circulating plasma proteins play a pivotal role in numerous physiological processes (e.g., signaling, transport, cellular growth, DNA repair, and immune defense) and are particularly noted for their involvement in cancer progression and treatment [263]. These proteins not only serve as valuable biomarkers for cancer detection and prognosis, [264], but are also actively involved in tumor cell growth, invasion, and microenvironment formation [265]. Therefore, circulating proteins can potentially serve as molecular targets for the prevention and treatment of CRC.

Several observational studies have sought to identify plasma proteins that increase the risk of specific types of cancer [266-269]. However, as many of these studies measured protein levels at cancer diagnosis, they are inherently limited by potential reverse causality and residual confounding. Furthermore, they have typically analyzed only a small set of proteins in limited sample sizes. To address these challenges, we used Mendelian Randomization (MR) analysis with

protein quantitative trait loci (pQTLs) as instrumental variables (IVs) to investigate the causal impacts of genetically determined circulating protein levels on CRC [208]. This approach leverages the random assignment of genetic variants during meiosis to mitigate confounding bias and reverse causality issues typically found in observational studies [208-210].

For this analysis, we performed two-sample MR analysis utilizing over 20,000 pQTLs identified from three large-scale proteomic genome-wide association study (GWAS) of more than 4,000 plasma protein concentrations [270-272]. In conducting this comprehensive examination of genetically predicted plasma protein levels and CRC risk, we aimed to uncover potential drug targets or risk stratification biomarkers for preventive and therapeutic interventions in CRC.

7.2 Methods

7.2.1 Data sources for plasma proteins

Genetic instruments for the circulating plasma proteins were obtained from three proteomic GWAS: (1) the deCODE Health study [270]; (2) the Fenland study [271]; and (3) the UK Biobank Pharma Proteomics Project (UKB-PPP) [272]. Specifically, the deCODE Health study measured 4,907 plasma proteins with aptamers in 35,559 Icelanders using the SomaScan version 4 assay (SomaLogic) [270]. The Fenland study identified a total of 3,892 proteins genotyped in 10,708 participants of European ancestry using the SomaScan version 4 assay (SomaLogic) [271]. Meanwhile, the UKB-PPP performed proteomic profiling on 2,923 plasma proteins from 54,219 UK Biobank participants utilizing the antibody-based Olink Explore 3072 PEA platform [272].

We identified independent pQTLs as candidate genetic instruments for plasma proteins, adhering to guidelines from prior research [273]. We directly used the set of pQTLs from Zhang,

Y., et al.'s research for the deCODE study [273]. For the Fenland study and the UKB-PPP, we first selected single nucleotide polymorphisms (SNPs) linked to each protein based on the study-specific significance level ($1.004x10^{-11}$ for the Fenland study, and $1.7x10^{-11}$ for the UKB-PPP). Subsequently, SNPs within the human major histocompatibility complex (MHC) region (chr6 26Mb to chr6 34Mb) were excluded due to complicated linkage disequilibrium (LD) patterns in the MHC region. Further, correlated SNPs for each protein were removed by implementing LD clumping (upstream/downstream distance > 5000kb; the thresholds of LD: $r^2 < 0.01$). Lastly, SNPs associated with five or more proteins were excluded to minimize the risk of horizontal pleiotropy.

Selected pQTLs were further categorized into *cis*-pQTLs and *trans*-pQTLs based on their genomic position. *cis*-pQTLs were defined as SNPs within 500kb from the protein-encoding gene, whereases *trans*-pQTLs were SNPs on different chromosomes or residing beyond the 500kb region from the corresponding gene [215].

7.2.2 Data source for CRC

Summary-level statistics on the association of SNPs with colorectal cancer were derived from the FinnGen study R10 (https://www.finngen.fi/en/access_results, accessed on 20 February 2024) [274]. Detailed information is presented in Table 7-1. The FinnGen study consisted of 430,897 Finnish from multiple cohorts, and genotyping was conducted using Illumina (Illumina Inc., San Diego, USA) and Affymetrix chip arrays (Thermo Fisher Scientific, Santa Clara, CA, USA). All genetic association estimates between SNPs and cancers were conducted using logistic regression models comparing cases and controls, adjusting for age, sex, genotyping batch, and the first ten genetic principal components to account for population stratification. Ethical approval and

informed consent have been received by the original study [274]; thus no new ethics approval was required for this study.

7.2.3 Statistical analyses

Mendelian Randomization analysis

In this study, we investigated the potential causal relationships between genetically predicted plasma protein levels (exposures) and CRC risk (outcome). The MR analysis was performed separately using *cis*-pQTLs only and a combination of *cis*- and *trans*-pQTLs as genetic instruments for plasma proteins. The MR estimates were calculated using the Wald ratio method for proteins with single pQTLs, and the inverse variance weighted (IVW) method for proteins with ≥2 pQTLs. The IVW method assumes that all SNPs are valid instruments and that horizontal pleiotropic effects are absent or balanced, constraining the intercepts to zero [275]. To enhance the robustness of our study, additional analyses were used to account for horizontal pleiotropy and heterogeneity. Specifically, MR-Egger methods were utilized to test and correct horizontal pleiotropic effects [233]. Furthermore, if the Cochran's Q statistics suggested heterogeneity among genetic instruments, the weighted median methods were implemented to calculate the estimates [230].

The MR analyses were performed using the 'TwoSampleMR' package [276] in R software version 4.3.3 (The R Foundation for Statistical Computing) [131]. We adopted the false discovery rate (FDR) approach to adjust for multiple comparisons. All analyses were two-sided with the level of significance pre-determined at p-value < 0.05. A protein is defined as a MR prioritized protein if it is statistically significant associated with CRC risk at FDR corrected p-value < 0.05.

Colocalization analysis

For MR prioritized proteins using *cis*-pQTLs, we performed a Bayesian colocalization analysis utilizing summary statistics for pQTLs (both statistically significant and non-significant pQTLs) within the protein-encoding gene to evaluate whether a single variant in the corresponding gene affected both protein levels and cancer risk. Given that the summary statistics for non-significant pQTLs were not available in the Fenland Study, we performed colocalization analysis for pQTLs only from deCODE and UKBB-PPP.

The colocalization analysis assesses five distinct hypotheses: H_0, the gene is not associated with either trait; H_1, the gene is associated with trait 1 only; H_2, the gene is associated with trait 2 only; H_3, the gene is associated with both trait 1 and trait 2, but traits are associated with two distinct causal variants within the gene; H_4, one causal variant within the gene is associated with both trait 1 and trait 2 [277]. Colocalization calculates the posterior probability for each hypothesis, with a posterior probability for the shared causal variant hypothesis (PH_4)>0.8 indicating strong evidence of colocalization. The colocalization analysis was conducted using the 'coloc' package [277] in R.

Steiger filtering analysis

To evaluate potential reverse causality, we conducted the Steiger filtering analysis on all pQTLs that passed the multiple-testing threshold. This analysis evaluates the potential direction of effect by calculating the SNPs-exposure correlation ($r_{exposure}$) and the SNPs-outcome correlation ($r_{outcome}$), and performing hypothesis testing of $r_{exposure} = r_{outcome}$ [278]. A p-value less than 0.05 indicated that the direction of association might potentially originate from protein to CRC. The Steiger filtering analysis was conducted using the 'TwoSampleMR' package [276] in R.

Protein-altering variants (PAVs) annotation of cis-pQTLs

The proteomic profiling in the pQTL studies (deCODE, Fenland, and UKBB-PPP) utilized affinity-based techniques, which infer protein levels based on the binding affinity of epitopes rather than directly measuring protein concentrations. Consequently, the quantitative variations observed might be attributable to differences in aptamer binding efficiency across assays. Aptamer binding efficiency may be impacted by SNPs that induce changes in protein structure, known as protein-altering variants (PAVs). To identify potential PAVs, we used the Ensembl Variant Effect Predictor (https://useast.ensembl.org/Tools/VEP, accessed on 8 March 2024) to annotate the MR prioritized *cis*-pQTLs. Genetic variants were classified as PAVs if they are annotated or in LD (r^2 > 0.8) with SNPs annotated as "coding sequence variant", "frameshift variant", "in-frame deletion", "in-frame insertion", "missense variant", "PAV", "splice acceptor variant", "splice donor variant", "splice region variant", "start lost", "stop gained" or "stop lost to assess" [265].

Protein–protein interaction (PPI) and functional pathway enrichment analysis

We constructed PPI networks for the MR-evident proteins in each cancer type using the online tool 'Search Tool for the Retrieval of Interacting Genes' (STRING, version 12.0, https://string-db.org/, accessed on 8 March 2024) to evaluate the relationships between these proteins. A PPI enrichment test with p-value < 0.05 indicated statistically significantly more interactions among these proteins than expected, underscoring potential biological relevance.

Furthermore, to explore the underlying pathways enriched by MR-evident proteins for CRC, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were implemented using the 'ClusterProfiler' package [279, 280] in R. The FDR correction method was used to account for multiple hypotheses testing.

7.3 Results

The schematic overview of our study design is shown in Figure 7-1. We included a total of (1) 2,051 *cis*-pQTLs for 1,406 proteins and 3,376 *trans*-pQTLs for 1,461 proteins in the deCODE Health study; (2) 1,948 *cis*-pQTLs for 1,536 proteins and 1,547 *trans*-pQTLs for 1,263 proteins in the Fenland study; and (3) 7,617 *cis*-pQTLs for 1,823 proteins and 13,669 *trans*-pQTLs for 1,829 proteins within the UKB-PPP study. After removing missing SNPs in FinnGen, we analyzed (1) 1,349 proteins from *cis*-pQTLs and 2,079 proteins for all pQTLs in the deCODE Health study; (2) 1,363 proteins from *cis*-pQTLs and 2,060 proteins from all pQTLs in the Fenland study; and (3) 1,767 proteins from *cis*-pQTLs and 2,100 proteins from all pQTLs in UKB-PPP study.

Association between plasma proteins and CRC risk in cis-MR analyses

cis-MR analyses identified statistically significant (FDR-corrected $p<0.05$) associations with CRC risk for four unique proteins across the three pQTL datasets: greminlin-1 (GREM1), lactase/phlorizin hydrolase (LPH), cGMP-specific 3',5'-cyclic phosphodiesterase (PDE5A), and LIM domain and actin-binding protein 1 (LIMA1) (Figure 7-2). The strongest associations were observed for GREM1, which was positively associated with CRC risk using pQTLs from deCODE (OR = 1.16, FDR-corrected $p < 0.001$) and Fenland (OR = 1.12, FDR-corrected $p < 0.001$). PDE5A was also associated with increased CRC risk using pQTLs from Fenland (OR = 1.25, FDR-corrected $p = 0.017$) and UKBB-PPP (OR =1.63, FDR-corrected $p = 0.044$). Conversely, increasing levels of LPH was inversely associated with CRC risk in deCODE (OR = 0.93, FDR-corrected $p = 0.036$) and Fenland (OR = 0.92, FDR-corrected $p = 0.024$). Finally, LIMA1 was associated with increased CRC risk in Fenland only (OR = 1.49, FDR-corrected $p = 0.026$).

Association between plasma proteins and CRC risk in cis+trans-MR analyses

MR analysis using all pQTLs (*cis+trans*) identified statistically significant (FDR-corrected $p<0.05$) associations for eight proteins in deCODE, eight proteins in Fenland and six proteins in UKBB-PPP (19 unique proteins across the three datasets). The four proteins identified in the *cis*-MR analyses (GREM1, LPH, PDE5A, and LIMA1) had similar associations (direction and strength) in the *cis+trans* MR analysis.

In deCODE, significant associations with CRC risk were observed for GREM1, LPH, Tubulin-specific chaperone A (TBCA), Phosphatidylinositol transfer protein alpha isoform (PIPNA); D-3-phosphoglycerate dehydrogenase (SERA), ADP-ribosylation factor-like protein 2 (ARL2), Sialidase-1 (NEUR1), and Neuroligin-2 (NLGN2) (Table 7-2 and Figure 7-3A). Of these, LPH (OR = 0.93, 95% CI 0.90, 0.96) and NEUR1 (OR = 0.27, 95% CI 0.15, 0.49) exhibited inverse associations with CRC risk. The remaining six proteins were observed to be positively associated with CRC risk, with the highest OR observed for SERA (OR = 4.41, 95% CI 2.18, 8.92).

Utilizing all pQTLs from the Fenland Study (Table 7-2 and Figure 7-3B), we identified significant associations with CRC risk for GREM1, PDE5A, LPH, LIMA1, Peptide YY (PYY), Gremlin-2 (GREM2), Syntenin-1 (SDCB1), and PCNA-associated factor (PAF15). Odds ratios ranged from 0.36 (95% CI 0.26, 0.50) for GREM2 to 1.65 (95% CI 1.30, 1.39) for PAF15.

From UKBB-PPP, MR analyses with all pQTLs identified (Table 7-2 and Figure 7-3C) significant associations for PDE5A, septin-8 (SEPT8), tumor necrosis factor receptor superfamily member 6B (TNF6B), eotaxin (CCL11), mitogen-activated protein kinase kinase kinase kinase 5 (M4K5), and chordin-like protein 2 (CRDL2). Except for M4K5 (OR = 0.57, 95% CI 0.44, 0.75), all proteins were associated with increased risk of CRC.

Colocalization analysis

Of proteins identified in *cis*-MR, colocalization analyses were conducted for three proteins with full summary statistics available (GREM1 from deCODE, LPH from deCODE and PD35A from UKBB-PPP) (Table 7-3). We observed suggested evidence of colocalization for LPH (PH4 = 0.9185) and PDE5A (PH4 = 0.8531), but not for GREM1 (PH4 = 0.0098).

Steiger filtering analysis

According to the results from Steiger filtering analyses, all MR-identified associations were in the correct causal direction from genetically predicted plasma protein levels to CRC risk (Table 7-4).

PAV assessment

The four proteins identified in *cis*-MR analyses were linked to 6 unique *cis*-pQTLs across the three datasets: GREM1 with rs2293582 in deCODE and with rs58658771 in Fenland; LPH with rs4988235 in deCODE and Fenland; PDE5A with rs59867181 in Fenland and with rs58583086 in UKBB-PPP; and LIMA1 with rs10783342 in the Fenland study. Of these 6 unique *cis*-pQTLs, only rs4988235 was annotated to or was in LD with PAVs that potentially affect aptamer binding (r2=0.854). However, given the extensive research linking rs4988235 with LPH levels, the likelihood that this pQTL acts as a PAV is considered minimal [281, 282].

PPI network and Pathway enrichment analysis

In the PPI network analysis of *cis*-MR prioritized proteins, no interaction was observed among these 4 proteins (Figure 7-4), suggesting insufficient evidence for more interactions than would be expected by chance. Conversely, the network identified five interactions among the 19

MR-prioritized proteins with all pQTLs (p-value=0.0013), indicating a significant degree of interconnectivity (Figure 7-5).

GO enrichment analyses unveiled several pathways among these proteins (Figure 7-6 and 7-7). For *cis*-MR prioritized proteins, pathways such as leukocyte proliferation, carbohydrate derivative catabolic process, vascular endothelial growth factor (VEGF) receptor binding, cGMP binding, and cyclic-nucleotide phosphodiesterase activity, showed statistically significant evidence of enrichment. Furthermore, additional pathways, including cytokine receptor binding and G protein-coupled receptor binding, were enriched for *cis+trans* MR prioritized proteins.

KEGG enrichment analyses, in contrast, only identified two enriched pathways for proteins identified in *cis* but not *cis+trans* MR analysis, including galactose metabolism and carbohydrate digestion and absorption (Figure 7-8).

7.4 Discussion

We identified 4 unique proteins (GREM1, LPH, PDE5A, LIMA1) from the *cis*-MR analysis statistically significantly linked to CRC risk. MR analysis utilizing all (*cis+trans*) pQTLs found 15 additional proteins (TBCA, PIPNA, SERA, ARL2, NEUR1, NLGN2, PYY, GREM2, SDCB1, PAF15, SEPT8, TNF6B, CCL11, M4K5, CRDL2), resulting in a total of 19 proteins that were significantly associated with CRC risk. Subsequent validation efforts, including colocalization analysis, Steiger filtering analysis, and PAV assessment, reinforced these associations. Notably, two *cis*-prioritized proteins (LPH and PDE5A) were found to have strong evidence of colocalization with CRC risk. Steiger filtering analysis affirmed the directional accuracy from the plasma proteins to CRC among all identified associations. The pathway

enrichment analyses further revealed that these proteins' involvement in various pathways were related to protein bindings and enzyme activities, shedding light on their potential mechanistic roles in CRC pathogenesis.

Our findings were consistent with existing research on the underlying biological mechanisms of these MR prioritized proteins. For example, GREM1 has been involved in promoting metastasis and poor prognosis across multiple cancer types. Studies have demonstrated GREM1's role in enhancing CRC cell invasion through mechanisms involving the activation of ATF6, inhibition of ATF4 pathways, and induction of epithelial–mesenchymal transition [283, 284]. Ongoing clinical development of ginisortamab, an anti-gremlin-1 antibody, shows promising results in targeting GREM1 in various cancers, including CRC and prostate cancer [285].

LPH has garnered attention for its potential involvement in cancer development, particularly in colorectal and prostate cancers [196, 286, 287]. Although the relationship between LPH genotype and CRC remains contradictory, studies have suggested an association between lactose intolerance and decreased prostate cancer risk [287, 288]. The mechanistic basis for the association between LPH and CRC may involve modulating lactose-derived metabolites in the gut, influencing the risk of carcinogenesis through effects on the gut microbiome, inflammation, or other pathways. Further research is needed to elucidate the precise role of LPH in cancer development.

PDE5 has shown potential involvement in cancer development and progression, particularly in colorectal cancer. Inhibition of PDE5 induces apoptosis in colon tumor cells by sustaining cGMP levels, leading to the activation of protein kinase G2 and subsequent suppression of proliferation and promotion of differentiation in colon epithelial cells [289].

LIMA1 has emerged as a crucial regulator of cancer progression in various malignancies, including colorectal, breast, prostate, and lung cancers. In colorectal cancer cells, LIMA1, particularly its isoform EPLIN-β, is targeted by ornithine decarboxylase antizyme 1, leading to enhanced cellular migration [290, 291].

Our study has several strengths. By leveraging summary statistics from three extensive protein GWAS, we were able to investigate the association between a wide array of plasma proteins and CRC risk. Furthermore, we performed a series of subsequent analyses to comprehensively validate our findings. We conducted colocalization analysis to evaluate whether a single variant within the protein-encoding gene affected both protein levels and CRC risk. Further, we performed Steiger filtering analysis to assess the directionality of observed associations to address the possibility of reverse causality. Through PAV assessment, we examined whether the observed associations between *cis*-MR prioritized proteins and CRC could be influenced by aptamer binding artifacts, ensuring the robustness of our causal inferences. PPI and pathway enrichment analysis were performed to elucidate the complex interactions and biological functions of the MR prioritized proteins, thereby uncovering the underlying mechanisms that may contribute to CRC development.

Despite the strengths, this study is also subject to several limitations. The reliance on data from individuals of European descent restricts the generalizability of our findings to this particular population. Furthermore, colocalization analyses were confined to proteins within the deCODE Health study and UKB-PPP due to the unavailability of full summary data from the Fenland study, leaving uncertainties regarding the proteins identified in this dataset. Nevertheless, the overlap of proteins identified across the Fenland, deCODE, and UKB-PPP datasets leads us to expect comparable results from colocalization analyses. Lastly, another limitation stems from the

proteomic GWAS measuring circulating proteins, which are either secreted or passively leaked. As the concentrations of these circulating proteins may not reflect their levels within cells and tissues, our findings may not reflect CRC risk associated with cellular or tissue-specific protein abundances.

In conclusion, by implementing two-sample MR analyses, our study identified several circulating proteins causally associated with CRC risk. Our findings are further reinforced by a series of validation efforts, including colocalization analysis, Steiger filtering analysis, PAV assessment, as well as PPI network and pathway enrichment analysis. These results underscore the critical role of plasma proteins and their downstream effects in the pathogenesis of CRC, providing new insights into preventive strategies and potential therapeutic targets. Further research is warranted to fully elucidate the intricate biological mechanisms and the precise roles that these proteins play in the development of CRC.

CHAPTER 8. CONCLUSIONS AND PUBLIC HEALTH IMPLICATIONS

In this research, we explored the complex interplay between dietary factors, genetic predispositions, and circulating proteins in the context of CRC risk and progression. This work sought to unravel how CRC outcomes are influenced by dairy/milk intake, genetic polymorphisms related to dairy/milk digestion pathways, and circulating proteins levels.

Aim 1 delved into the association between dairy/milk intake with its key components (lactose, calcium, vitamin D) and CRC incidence and mortality. Findings from our cohort study highlighted an inverse relationship between dairy/milk consumption and CRC risk, affirming its protective role. Our survival analysis further investigated the effects of dairy/milk intake on CRC mortality, revealing varied impacts across different ethnic groups.

Aim 2 shifted focus to the genetic predispositions of CRC, examining the influence of candidate SNPs within genes related to dairy/milk digestion (*LCT, MCM6, CASR, VDR*) on CRC risk. Our candidate SNPs association study suggested significant associations between specific genetic variants and CRC risk and further underscored the genetic heterogeneity in CRC susceptibility. In addition, our Mendelian Randomization analysis indicated the causal relationship between genetically predicted levels of LPH and CRC incidence, providing compelling evidence for the protective effect of elevated LPH levels against CRC.

Aim 3 extended our inquiry into the realm of circulating proteins, leveraging the Mendelian Randomization framework to elucidate the causal links between circulating protein concentrations and CRC risk. Findings further enriched our understanding of CRC pathogenesis, providing insights on novel biomarkers and therapeutic targets.

Our findings significantly contribute to the broader CRC research literature, offering novel insights into the protective effects of dairy/milk intake, the intricate genetic landscape of CRC susceptibility, and the promising role of circulating proteins in CRC prevention and treatment. By unveiling these relationships, this work could bridge the critical gaps in our current understanding of CRC etiology and progression and further pave the way for more targeted and effective prevention, diagnostic, and treatment strategies.

The findings from this research illuminate several key areas with significant public health implications, including dietary recommendations, genetic profiling, and precision medicine. The protective roles of dairy/milk intake and their key components against CRC risk underscore the importance of incorporating these food items into dietary guidelines for CRC prevention, particularly considering individual genetic susceptibilities and ethnic-specific dietary patterns. In addition, the identification of specific SNPs associated with CRC risk highlights the potential for genetic screening to improve risk stratification and early detection, enabling targeted prevention strategies, especially in high-risk groups. The elucidation of causal relationships between circulating proteins and CRC risk offers novel insight on precision medicine, including the search of new biomarkers and the development of targeted therapies that modulate protein levels or activity.

In conclusion, this research provides valuable insights into our scientific understanding of the multifaceted etiology of CRC and lays the groundwork for innovative CRC prevention and treatment strategies that could significantly impact public health outcomes. Further research in larger, more diverse cohorts is essential to validate these findings and translate them into clinical practice and public health policies. Through targeted dietary recommendations, improved genetic screening, and the exploration of new therapeutic targets, the findings of this work hold the

promise of enhancing CRC prevention, early detection, and treatment, ultimately contributing to better health outcomes and quality of life for individuals worldwide.

TABLES

Table 1-1. Research gaps and corresponding dissertation aims

Topics	Research Gap	Aim Addressing the Gap	Details
Dairy/Milk's effect on CRC risk and survival	Elusive mechanisms underlying the protective roles of dairy/milk against CRC incidence and mortality	Aim 1a & 1b	Explores how key components of dairy and milk (lactose, calcium, vitamin D) influence CRC risk and survival, aiming to elucidate the biological mechanisms involved
Ethnic Heterogeneity in Dairy/Milk-CRC Associations	Lack of understanding regarding the heterogeneity of dairy and milk effects on CRC risk and survival across different races/ethnicities	Aim 1a & 1b	Examines the effects of dairy/milk consumption on CRC incidence and mortality within five racial and ethnic groups, aiming to provide insights into race-specific dietary guidelines
Interaction between Dairy/Milk consumption and digestion in CRC	Limited studies on the interactions between dairy/milk consumption and lactase persistence status on CRC risk	Aim 2a	Investigates the association between dairy/milk intake and CRC incidence, stratified by lactase persistence status, to uncover gene-diet interactions
Genetic polymorphisms and non-European population	Research on genetic polymorphisms linked to dairy/milk digestion primarily focused on European populations	Aim 2a	Examines SNPs within *LCT*, *MCM6*, *CASR*, and *VDR* genes across multiethnic groups, aiming to fill this knowledge gap
Lactase-phlorizin hydrolase (LPH)'s role in CRC risk	Limited understanding of the causal effect of genetically determined LPH levels on CRC risk	Aim 2b	Leverages Mendelian Randomization method to evaluate the casual effect of genetically determined LPH levels on CRC risk
Circulating Proteins and CRC Risk	Unclear causal links between circulating proteins and CRC risk	Aim 3	Utilizes MR framework to explore causal relationships between genetically determined levels of circulating proteins and CRC risk, aiming to identify potential biomarkers and therapeutic targets

Table 3-1. Baseline characteristics of participants according to intake of all dairy products, Multiethnic Cohort Study, 1993–2019

Characteristics	Total dairy products intake				
	Q1	Q2	Q3	Q4	Q5
No. of participants	38,529	38,529	38,528	38,529	38,529
Age at cohort entry in years					
Mean (SD)	59 (9)	59 (9)	61 (9)	61 (9)	62 (9)
Median (IQR)	59 (52, 67)	59 (52, 67)	61 (53, 68)	62 (54, 69)	63 (55, 69)
Sex, n (%)					
Men	19,959 (52%)	18,396 (48%)	17,763 (46%)	16,732 (43%)	13,887 (36%)
Ethnicity, n (%)					
White	5,318 (14%)	7,817 (20%)	9,244 (24%)	10,933 (28%)	13,685 (36%)
African American	5,980 (16%)	7,589 (20%)	7,163 (19%)	6,378 (17%)	5,655 (15%)
Native Hawaiian	3,311 (8.6%)	3,366 (8.7%)	2,725 (7.1%)	2,375 (6.2%)	1,967 (5.1%)
Japanese American	18,788 (49%)	11,574 (30%)	9,589 (25%)	8,187 (21%)	5,822 (15%)
Latino	5,132 (13%)	8,183 (21%)	9,807 (25%)	10,656 (28%)	11,400 (30%)
Family history of colorectal cancer, n (%)	3,290 (8.5%)	3,060 (7.9%)	3,028 (7.9%)	2,973 (7.7%)	2,942 (7.6%)
History of intestinal polyps, n (%)	2,406 (6.2%)	1,925 (5.0%)	2,090 (5.4%)	2,031 (5.3%)	2,047 (5.3%)
Education, n (%)					
≤12th grade	16,848 (44%)	16,896 (44%)	17,317 (45%)	17,019 (45%)	16,673 (44%)
Vocational/some college	11,588 (30%)	11,625 (31%)	11,075 (29%)	10,872 (29%)	10,579 (28%)
≥College graduate	9,695 (25%)	9,589 (25%)	9,677 (25%)	10,184 (27%)	10,787 (28%)
Body mass index, n (%)					
Underweight/Normal (<25 kg/m^2)	17,158 (45%)	14,787 (39%)	15,163 (40%)	15,695 (41%)	16,043 (42%)
Overweight (25-30 kg/m^2)	14,380 (38%)	14,998 (39%)	14,973 (39%)	14,797 (39%)	14,494 (38%)
Obese (≥30 kg/m^2)	6,628 (17%)	8,383 (22%)	7,957 (21%)	7,692 (20%)	7,547 (20%)
Diagnosis of diabetes, n (%)	3,874 (10%)	4,085 (11%)	4,486 (12%)	4,813 (12%)	5,370 (14%)
Physical activity (MET-hours/day), n (%)					
Q1 [0, 0.321)	7,097 (19%)	7,063 (19%)	6,976 (18%)	7,046 (19%)	7,667 (20%)
Q2 [0.321, 0.571)	7,781 (21%)	7,953 (21%)	7,649 (20%)	7,691 (20%)	7,555 (20%)
Q3 [0.571, 0.929)	7,648 (20%)	7,684 (20%)	7,752 (20%)	7,466 (20%)	7,500 (20%)
Q4 [0.929, 1.786)	7,607 (20%)	7,648 (20%)	7,926 (21%)	7,861 (21%)	7,426 (20%)
Q5 [1.786, 14.286)	7,621 (20%)	7,491 (20%)	7,545 (20%)	7,746 (20%)	7,474 (20%)
Smoking status and pack-years, n (%)					
Never	14,812 (40%)	15,700 (42%)	16,944 (46%)	17,636 (48%)	18,562 (50%)
Former, <20 pack-years	10,533 (28%)	10,869 (29%)	10,893 (29%)	10,920 (30%)	10,257 (28%)
Former, ≥20 pack-years	4,408 (12%)	3,703(10%)	3,527 (9.5%)	3,460 (9.4%)	3,412 (9.2%)
Current, <20 pack-years	3,729 (10%)	3,754 (10%)	3,237 (8.7%)	2,842 (7.7%)	2,559 (6.9%)
Current, ≥20 pack-years	3,816 (10%)	3,128 (8.4%)	2,484 (6.7%)	2,147 (5.8%)	2,141 (5.8%)
Alcohol consumption, n (%)					
Never	20,032 (52%)	18,984 (49%)	19,065 (49%)	19,666 (51%)	21,415 (56%)
<15 g/day	9,858 (26%)	12,645 (33%)	13,280 (34%)	13,354 (35%)	13,107 (34%)
15-30 g/day	3,065 (8.0%)	2,961 (7.7%)	2,949 (7.7%)	2,811 (7.3%)	2,208 (5.7%)

Characteristics	Total dairy products intake				
	Q1	Q2	Q3	Q4	Q5
≥30 g/day	5,574 (14%)	3,939 (10%)	3,234 (8.4%)	2,698 (7.0%)	1,799 (4.7%)
Multivitamin use, n (%)	17,405 (46%)	18,129 (48%)	19,328 (51%)	20,335 (54%)	21,059 (56%)
NSAIDs use, n (%)	16,810 (46%)	19,718 (53%)	20,426 (55%)	20,831 (56%)	21,248 (58%)
Use of hormone replacement therapy, n (%)					
Male, non-user	19,959 (52%)	18,396 (48%)	17,763 (46%)	16,732 (43%)	13,887 (36%)
Female, missing menotype	180 (0.5%)	214 (0.6%)	190 (0.5%)	196 (0.5%)	242 (0.6%)
Premenopausal female	2,868 (7.4%)	3,217 (8.3%)	2,879 (7.5%)	2,691 (7.0%)	2,570 (6.7%)
Postmenopausal female, ever use	8,072 (21%)	8,644 (22%)	9,406 (24%)	10,268 (27%)	12,029 (31%)
Postmenopausal female, never/unknown use	7,450 (19%)	8,058 (21%)	8,290 (22%)	8,642 (22%)	9,801 (25%)
Total energy intake (kcal/day)					
Mean (SD)	2,106 (1,016)	2,228 (1,108)	2,315 (1,154)	2,257 (1,072)	1,992 (948)
Median (IQR)	1,892 (1,397, 2,572)	1,997 (1,448, 2,725)	2,047 (1,513, 2,850)	2,051 (1,530, 2,708)	1,777 (1,352, 2,408)
Red meat (g/1,000 kcal/day), n (%)					
Q1 [0, 7.865)	7,443 (19%)	5,362 (14%)	6,380 (17%)	8,001 (21%)	11,343 (29%)
Q2 [7.865, 13.632)	6,887 (18%)	6,347 (16%)	7,373 (19%)	8,303 (22%)	9,619 (25%)
Q3 [13.632, 19.387)	7,447 (19%)	7,409 (19%)	7,999 (21%)	8,203 (21%)	7,470 (19%)
Q4 [19.387, 27.184)	7,673 (20%)	8,874 (23%)	8,288 (22%)	7,770 (20%)	5,924 (15%)
Q5 [27.184, 217.538)	9,079 (24%)	10,537 (27%)	8,488 (22%)	6,252 (16%)	4,173 (11%)
Processed meat (g/1,000 kcal/day), n (%)					
Q1 [0, 2.445)	7,004 (18%)	5,266 (14%)	6,611 (17%)	8,019 (21%)	11,629 (30%)
Q2 [2.445, 4.854)	6,522 (17%)	6,514 (17%)	7,423 (19%)	8,613 (22%)	9,457 (25%)
Q3 [4.854, 7.612)	7,172 (19%)	7,658 (20%)	8,086 (21%)	8,238 (21%)	7,374 (19%)
Q4 [7.612, 11.657)	8,092 (21%)	8,886 (23%)	8,287 (22%)	7,470 (19%)	5,794 (15%)
Q5 [11.657, 173.788)	9,739 (25%)	10,205 (26%)	8,121 (21%)	6,189 (16%)	4,275 (11%)
Dietary fiber (g/1,000 kcal/day), n (%)					
Q1 [0, 8.162)	13,128 (34%)	9,222 (24%)	6,125 (16%)	4,885 (13%)	5,169 (13%)
Q2 [8.162, 10.218)	7,287 (19%)	8,946 (23%)	8,260 (21%)	7,126 (18%)	6,910 (18%)
Q3 [10.218, 12.302)	5,925 (15%)	7,795 (20%)	8,293 (22%)	8,530 (22%)	7,985 (21%)
Q4 [12.302, 15.106)	5,310 (14%)	6,809 (18%)	8,328 (22%)	9,211 (24%)	8,871 (23%)
Q5 [15.106, 42.376)	6,879 (18%)	5,757 (15%)	7,522 (20%)	8,777 (23%)	9,594 (25%)
Folate (μg DFE/day), n (%)					
Q1 [48.382, 438.996)	11,502 (30%)	9,026 (23%)	6,351 (16%)	5,092 (13%)	6,558 (17%)
Q2 [438.996, 671.471)	8,760 (23%)	8,503 (22%)	7,482 (19%)	6,879 (18%)	6,905 (18%)
Q3 [671.471, 944.361)	7,484 (19%)	7,744 (20%)	7,789 (20%)	7,821 (20%)	7,690 (20%)
Q4 [944.361, 1359.429)	5,793 (15%)	6,995 (18%)	8,294 (22%)	9,040 (23%)	8,407 (22%)
Q5 [1359.429, 8163.964)	4,990 (13%)	6,261 (16%)	8,612 (22%)	9,697 (25%)	8,969 (23%)

The unit of the total dairy products intake is g/1000 kcal/day.
Abbreviations: SD, standard deviation; IQR, interquartile range; MET, metabolic equivalent of task; NSAIDs, non-steroidal anti-inflammatory drugs; DFE, dietary folate equivalents.

Table 3-2. Relative risk of colorectal cancer according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019

	White					African American					Native Hawaiian				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	No. of cases	RR (95% CI)[a]	p-value	RR (95% CI)	p-value	No. of cases	RR (95% CI)	p-value	RR (95% CI)	p-value	No. of cases	RR (95% CI)	p-value	RR (95% CI)	p-value
Total dairy products (g/1000 kcal/day)															
1	171	1.00 (ref)				238	1.00 (ref)				115	1.00 (ref)			
2	236	1.02 (0.83, 1.26)	0.839			290	0.91 (0.76, 1.10)	0.326			107	0.94 (0.71, 1.24)	0.653		
3	233	0.84 (0.68, 1.04)	0.103	0.96 (0.91, 1.01)	0.120	264	0.84 (0.69, 1.02)	0.080	0.97 (0.91, 1.04)	0.401	72	0.74 (0.54, 1.03)	0.072	0.91 (0.79, 1.05)	0.187
4	309	0.92 (0.75, 1.13)	0.439			259	0.94 (0.77, 1.15)	0.549			67	0.86 (0.62, 1.20)	0.373		
5	369	0.87 (0.71, 1.07)	0.185			213	0.84 (0.68, 1.03)	0.092			51	0.82 (0.57, 1.18)	0.285		
p for trend			0.114					0.185					0.173		
Milk (g/1000 kcal/day)															
1	221	1.00 (ref)				239	1.00 (ref)				85	1.00 (ref)			
2	226	0.93 (0.76, 1.13)	0.445			312	0.98 (0.81, 1.17)	0.804			116	1.23 (0.91, 1.65)	0.177		
3	248	0.85 (0.70, 1.03)	0.106	0.96 (0.91, 1.01)	0.108	273	0.90 (0.74, 1.09)	0.278	0.96 (0.90, 1.03)	0.246	87	1.06 (0.77, 1.47)	0.703	0.90 (0.78, 1.03)	0.138
4	291	0.91 (0.76, 1.10)	0.343			249	0.94 (0.77, 1.15)	0.541			69	0.88 (0.62, 1.25)	0.466		
5	332	0.84 (0.70, 1.01)	0.061			191	0.83 (0.67, 1.02)	0.079			55	0.93 (0.64, 1.35)	0.715		
p for trend			0.094					0.083					0.247		
Lactose (g/1000 kcal/day)															
1	190	1.00 (ref)				224	1.00 (ref)				102	1.00 (ref)			
2	227	0.96 (0.78, 1.17)	0.674			294	0.95 (0.79, 1.15)	0.600			108	1.09 (0.82, 1.45)	0.565		
3	239	0.84 (0.68, 1.02)	0.082	0.95 (0.90, 1.01)	0.091	268	0.87 (0.72, 1.06)	0.181	0.97 (0.91, 1.04)	0.363	79	0.86 (0.62, 1.19)	0.362	0.91 (0.80, 1.05)	0.192
4	308	0.90 (0.74, 1.10)	0.307			260	0.94 (0.77, 1.15)	0.566			65	0.84 (0.59, 1.18)	0.303		
5	354	0.83 (0.68, 1.01)	0.059			218	0.87 (0.70, 1.07)	0.182			58	0.94 (0.66, 1.34)	0.729		
p for trend			0.061					0.229					0.277		

	White					African American					Native Hawaiian				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	**No. of cases**	**RR (95% CI)[a]**	**p-value**	**RR (95% CI)**	**p-value**	**No. of cases**	**RR (95% CI)**	**p-value**	**RR (95% CI)**	**p-value**	**No. of cases**	**RR (95% CI)**	**p-value**	**RR (95% CI)**	**p-value**
Total calcium (mg/1000 kcal/day)															
1	202	1.00 (ref)				296	1.00 (ref)				169	1.00 (ref)			
2	217	0.78 (0.63, 0.96)	0.017			305	1.00 (0.83, 1.20)	0.991			94	1.04 (0.78, 1.38)	0.806		
3	286	0.91 (0.74, 1.12)	0.381	0.97 (0.91, 1.04)	0.407	270	1.01 (0.83, 1.24)	0.899	1.01 (0.92, 1.10)	0.862	63	0.99 (0.70, 1.40)	0.954	0.97 (0.79, 1.18)	0.725
4	283	0.81 (0.65, 1.01)	0.066			212	0.98 (0.78, 1.23)	0.866			48	0.86 (0.57, 1.29)	0.462		
5	330	0.80 (0.63, 1.02)	0.077			181	1.00 (0.77, 1.30)	0.997			38	0.86 (0.53, 1.42)	0.561		
p for trend			0.227					0.948					0.446		
Total vitamin D (IU/1000 kcal/day)															
1	206	1.00 (ref)				262	1.00 (ref)				92	1.00 (ref)			
2	226	0.83 (0.68, 1.01)	0.066			276	0.92 (0.76, 1.12)	0.405			114	1.03 (0.76, 1.38)	0.863		
3	287	0.83 (0.68, 1.01)	0.066	0.98 (0.89, 1.08)	0.692	282	0.98 (0.81, 1.19)	0.845	0.96 (0.88, 1.06)	0.442	98	0.88 (0.63, 1.21)	0.430	0.92 (0.74, 1.15)	0.457
4	295	0.85 (0.67, 1.07)	0.168			220	0.91 (0.71, 1.15)	0.419			59	0.80 (0.52, 1.23)	0.315		
5	304	0.79 (0.59, 1.06)	0.114			224	0.80 (0.59, 1.08)	0.150			49	0.98 (0.56, 1.73)	0.952		
p for trend			0.145					0.303					0.399		

[a] Adjusted for sex and 10-year age group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, history of intestinal polyps, education, BMI, smoking status and pack-years, alcohol consumption, physical activity, diabetes, use of NSAIDs, regular use of multivitamins, total energy intake, red meat intake, processed meat intake, folate intake, dietary fiber intake, and hormone use.

Abbreviations: CRC, colorectal cancer; RR, relative risk; CI, confidence interval; SD, standard deviation; BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 3-2 (*continue*). Relative risk of colorectal cancer according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019

	Japanese American					Latino					Test of heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of cases	RR (95% CI)[a]	p-value	RR (95% CI)	p-value	No. of cases	RR (95% CI)	p-value	RR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)											
1	827	1.00 (ref)				165	1.00 (ref)				
2	506	1.05 (0.94, 1.18)	0.372			261	0.96 (0.78, 1.18)	0.676			
3	383	0.95 (0.83, 1.08)	0.434	0.96 (0.91, 1.02)	0.214	282	0.90 (0.73, 1.11)	0.322	0.90 (0.84, 0.96)	0.002	0.958
4	324	1.00 (0.87, 1.15)	0.963			333	0.94 (0.77, 1.16)	0.584			
5	211	0.93 (0.79, 1.10)	0.393			289	0.77 (0.62, 0.95)	0.014			
p for trend			0.368					0.018			
Milk (g/1000 kcal/day)											
1	782	1.00 (ref)				178	1.00 (ref)				
2	436	0.92 (0.81, 1.04)	0.170			248	0.97 (0.79, 1.19)	0.754			
3	394	0.92 (0.81, 1.05)	0.241	0.96 (0.90, 1.02)	0.167	282	0.88 (0.72, 1.07)	0.202	0.89 (0.84, 0.95)	0.001	0.675
4	377	0.96 (0.84, 1.10)	0.579			343	0.98 (0.80, 1.19)	0.820			
5	262	0.91 (0.78, 1.06)	0.214			279	0.72 (0.59, 0.89)	0.002			
p for trend			0.277					0.005			
Lactose (g/1000 kcal/day)											
1	772	1.00 (ref)				186	1.00 (ref)				
2	512	1.01 (0.89, 1.13)	0.915			265	1.09 (0.89, 1.33)	0.405			
3	391	0.91 (0.80, 1.04)	0.169	0.96 (0.90, 1.02)	0.172	277	0.97 (0.80, 1.19)	0.797	0.91 (0.85, 0.97)	0.003	0.971
4	344	0.99 (0.87, 1.14)	0.924			320	1.04 (0.86, 1.27)	0.668			
5	232	0.92 (0.78, 1.08)	0.296			282	0.84 (0.68, 1.03)	0.096			
p for trend			0.297					0.062			

	Japanese American					Latino					Test of heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	**No. of cases**	**RR (95% CI)[a]**	**p-value**	**RR (95% CI)**	**p-value**	**No. of cases**	**RR (95% CI)**	**p-value**	**RR (95% CI)**	**p-value**	
Total calcium (mg/1000 kcal/day)											
1	818	1.00 (ref)				140	1.00 (ref)				
2	424	0.88 (0.78, 1.01)	0.068			294	0.92 (0.74, 1.15)	0.461			
3	336	0.92 (0.79, 1.07)	0.290	0.96 (0.91, 1.02)	0.196	366	0.87 (0.70, 1.09)	0.219	0.87 (0.79, 0.97)	0.015	0.509
4	318	0.91 (0.77, 1.08)	0.282			312	0.71 (0.56, 0.91)	0.006			
5	355	0.82 (0.68, 0.98)	0.025			218	0.71 (0.54, 0.94)	0.018			
p for trend			0.066					0.002			
Total vitamin D (IU/1000 kcal/day)											
1	586	1.00 (ref)				319	1.00 (ref)				
2	459	0.94 (0.83, 1.08)	0.387			311	1.05 (0.88, 1.25)	0.579			
3	352	0.89 (0.77, 1.03)	0.123	1.00 (0.92, 1.09)	0.970	295	0.95 (0.79, 1.14)	0.604	1.01 (0.91, 1.12)	0.814	0.628
4	428	0.91 (0.75, 1.11)	0.372			212	0.72 (0.57, 0.91)	0.005			
5	426	0.99 (0.78, 1.25)	0.909			193	0.83 (0.62, 1.11)	0.204			
p for trend			0.413					0.030			

[a] Adjusted for sex and 10-year age group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, history of intestinal polyps, education, BMI, smoking status and pack-years, alcohol consumption, physical activity, diabetes, use of NSAIDs, regular use of multivitamins, total energy intake, red meat intake, processed meat intake, folate intake, dietary fiber intake, and hormone use.

Abbreviations: CRC, colorectal cancer; RR, relative risk; CI, confidence interval; SD, standard deviation; BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 3-3. Relative risk of colorectal cancer according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different age groups, Multiethnic Cohort Study, 1993-2019

	<50 years old					50-60 years old				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	No. of cases	RR (95% CI)[a]	p-value	RR (95% CI)	p-value	No. of cases	RR (95% CI)	p-value	RR (95% CI)	p-value
Total dairy products (g/1000 kcal/day)										
1	177	1.00 (ref)				445	1.00 (ref)			
2	168	1.11 (0.89, 1.40)	0.347			381	0.96 (0.83, 1.11)	0.609		
3	93	0.81 (0.62, 1.07)	0.137	0.94 (0.84, 1.05)	0.298	302	0.90 (0.77, 1.05)	0.182	0.94 (0.88, 1.00)	0.049
4	96	1.02 (0.77, 1.34)	0.897			323	0.99 (0.84, 1.16)	0.897		
5	71	0.94 (0.69, 1.28)	0.697			240	0.82 (0.68, 0.98)	0.029		
p for trend			0.477					0.085		
Milk (g/1000 kcal/day)										
1	179	1.00 (ref)				447	1.00 (ref)			
2	159	0.89 (0.71, 1.11)	0.310			395	0.96 (0.83, 1.11)	0.586		
3	106	0.84 (0.65, 1.08)	0.176	0.94 (0.84, 1.05)	0.292	309	0.87 (0.74, 1.02)	0.079	0.93 (0.88, 0.99)	0.031
4	100	1.01 (0.78, 1.32)	0.92			301	0.92 (0.78, 1.08)	0.327		
5	61	0.73 (0.53, 0.99)	0.046			239	0.82 (0.69, 0.97)	0.025		
p for trend			0.173					0.026		
Lactose (g/1000 kcal/day)										
1	183	1.00 (ref)				434	1.00 (ref)			
2	156	1.02 (0.81, 1.27)	0.883			396	1.01 (0.87, 1.16)	0.923		
3	102	0.85 (0.65, 1.10)	0.219	0.94 (0.84, 1.05)	0.291	311	0.93 (0.80, 1.09)	0.387	0.94 (0.88, 1.00)	0.040
4	98	1.02 (0.78, 1.34)	0.885			307	0.97 (0.83, 1.14)	0.743		
5	66	0.84 (0.62, 1.15)	0.274			243	0.85 (0.71, 1.01)	0.065		
p for trend			0.339					0.085		

	<50 years old					50-60 years old				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	**No. of cases**	**RR (95% CI)[a]**	**p-value**	**RR (95% CI)**	**p-value**	**No. of cases**	**RR (95% CI)**	**p-value**	**RR (95% CI)**	**p-value**
					Total calcium (mg/1000 kcal/day)					
1	234	1.00 (ref)				468	1.00 (ref)			
2	130	0.82 (0.64, 1.04)	0.107			362	1.02 (0.87, 1.18)	0.841		
3	116	0.97 (0.74, 1.27)	0.814	0.80 (0.67, 0.96)	0.019	323	1.03 (0.87, 1.22)	0.744	0.99 (0.92, 1.06)	0.701
4	69	0.66 (0.47, 0.92)	0.016			289	0.98 (0.81, 1.19)	0.837		
5	56	0.70 (0.46, 1.06)	0.091			249	0.99 (0.79, 1.23)	0.898		
p for trend			0.050					0.825		
					Total Vitamin D (mg/1000 kcal/day)					
1	202	1.00 (ref)				451	1.00 (ref)			
2	133	0.92 (0.73, 1.16)	0.472			366	0.95 (0.82, 1.10)	0.470		
3	106	0.96 (0.74, 1.25)	0.751	0.89 (0.74, 1.06)	0.199	300	0.90 (0.76, 1.06)	0.188	1.09 (1.00, 1.19)	0.057
4	99	0.85 (0.61, 1.19)	0.339			295	0.89 (0.72, 1.09)	0.245		
5	65	0.64 (0.40, 1.02)	0.058			279	1.02 (0.79, 1.32)	0.879		
p for trend			0.186					0.449		

[a] Adjusted for sex and 10-year age group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, history of intestinal polyps, education, BMI, smoking status and pack-years, alcohol consumption, physical activity, diabetes, use of NSAIDs, regular use of multivitamins, total energy intake, red meat intake, processed meat intake, folate intake, dietary fiber intake, and hormone use.

Abbreviations: CRC, colorectal cancer; RR, relative risk; CI, confidence interval; SD, standard deviation; BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 3-3 (*continue*). Relative risk of colorectal cancer according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different age groups, Multiethnic Cohort Study, 1993-2019

	60-70 years old					≥70 years old					Test of heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of cases	RR (95% CI)[a]	p-value	RR (95% CI)	p-value	No. of cases	RR (95% CI)	p-value	RR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)											
1	616	1.00 (ref)				278	1.00 (ref)				
2	578	1.00 (0.89, 1.13)	0.943			273	0.97 (0.81, 1.17)	0.780			
3	525	0.86 (0.75, 0.97)	0.019	0.94 (0.89, 0.98)	0.004	314	0.95 (0.80, 1.14)	0.600	0.97 (0.91, 1.02)	0.247	0.758
4	563	0.94 (0.83, 1.07)	0.376			310	0.94 (0.78, 1.12)	0.469			
5	499	0.82 (0.72, 0.94)	0.004			323	0.94 (0.78, 1.14)	0.543			
p for trend			0.004					0.475			
Milk (g/1000 kcal/day)											
1	589	1.00 (ref)				290	1.00 (ref)				
2	562	1.06 (0.94, 1.20)	0.359			222	0.80 (0.66, 0.97)	0.027			
3	546	0.91 (0.80, 1.03)	0.126	0.93 (0.89, 0.97)	0.001	323	0.97 (0.81, 1.16)	0.725	0.96 (0.91, 1.02)	0.199	0.251
4	590	0.96 (0.85, 1.09)	0.51			338	0.95 (0.80, 1.13)	0.545			
5	494	0.82 (0.72, 0.94)	0.004			325	0.89 (0.74, 1.06)	0.201			
p for trend			0.002					0.633			
Lactose (g/1000 kcal/day)											
1	586	1.00 (ref)				271	1.00 (ref)				
2	585	1.04 (0.92, 1.17)	0.548			269	0.94 (0.78, 1.13)	0.509			
3	534	0.88 (0.77, 1.00)	0.047	0.94 (0.90, 0.98)	0.006	307	0.91 (0.76, 1.09)	0.291	0.96 (0.90, 1.02)	0.179	0.941
4	572	0.98 (0.86, 1.11)	0.737			320	0.91 (0.76, 1.09)	0.319			
5	504	0.85 (0.74, 0.97)	0.018			331	0.92 (0.76, 1.10)	0.349			
p for trend			0.015					0.354			

	60-70 years old					≥70 years old					Test of heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	**No. of cases**	**RR (95% CI)[a]**	**p-value**	**RR (95% CI)**	**p-value**	**No. of cases**	**RR (95% CI)**	**p-value**	**RR (95% CI)**	**p-value**	
Total calcium (mg/1000kcal/day)											
1	633	1.00 (ref)				290	1.00 (ref)				
2	564	0.88 (0.77, 1.00)	0.050			278	0.86 (0.71, 1.04)	0.110			
3	557	0.85 (0.74, 0.98)	0.023	0.94 (0.89, 0.99)	0.022	325	0.95 (0.78, 1.15)	0.605	0.99 (0.93, 1.06)	0.804	0.725
4	518	0.80 (0.69, 0.93)	0.004			297	0.86 (0.70, 1.06)	0.166			
5	509	0.77 (0.65, 0.91)	0.002			308	0.80 (0.63, 1.01)	0.058			
p for trend			0.002					0.114			
Total Vitamin D (IU/1000kcal/day)											
1	566	1.00 (ref)				246	1.00 (ref)				
2	583	0.94 (0.83, 1.07)	0.362			304	0.96 (0.80, 1.16)	0.697			
3	584	0.92 (0.80, 1.05)	0.196	0.95 (0.88, 1.02)	0.142	324	0.90 (0.75, 1.09)	0.298	0.99 (0.90, 1.08)	0.756	0.874
4	517	0.80 (0.68, 0.95)	0.009			303	0.95 (0.75, 1.19)	0.641			
5	531	0.85 (0.69, 1.04)	0.119			321	0.85 (0.64, 1.14)	0.285			
p for trend			0.036					0.296			

[a] Adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, history of intestinal polyps, education, BMI, smoking status and pack-years, alcohol consumption, physical activity, diabetes, use of NSAIDs, regular use of multivitamins, total energy intake, red meat intake, processed meat intake, folate intake, dietary fiber intake, and hormone use.

Abbreviations: CRC, colorectal cancer; RR, relative risk; CI, confidence interval; SD, standard deviation; BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-1. Baseline characteristics of CRC patients according to intake of total dairy products, Multiethnic Cohort Study, 1993–2019

Characteristic	Total dairy products intake				
	Q1	Q2	Q3	Q4	Q5
No. of participants	1,308	1,306	1,306	1,306	1,304
Age at CRC diagnosis in years					
Mean (SD)	72 (10)	72 (9)	74 (9)	74 (9)	76 (9)
Median (IQR)	72 (66, 79)	72 (66, 79)	74 (68, 80)	75 (69, 81)	76 (70, 82)
Sex, n (%)					
Men	763 (58%)	688 (53%)	687 (53%)	644 (49%)	542 (42%)
Ethnicity, n (%)					
White	148 (11%)	196 (15%)	251 (19%)	295 (23%)	420 (32%)
African American	209 (16%)	276 (21%)	267 (20%)	267 (20%)	231 (18%)
Native Hawaiian	106 (8.1%)	103 (7.9%)	77 (5.9%)	61 (4.7%)	62 (4.8%)
Japanese American	706 (54%)	497 (38%)	431 (33%)	352 (27%)	251 (19%)
Latino	139 (11%)	234 (18%)	280 (21%)	331 (25%)	340 (26%)
Family history of colorectal cancer, n (%)	143 (11%)	151 (12%)	135 (10%)	121 (9.3%)	129 (9.9%)
Education, n (%)					
≤12th grade	647 (50%)	611 (47%)	616 (48%)	609 (47%)	574 (45%)
Vocational/some college	362 (28%)	414 (32%)	361 (28%)	365 (28%)	399 (31%)
≥College graduate	287 (22%)	270 (21%)	313 (24%)	316 (24%)	310 (24%)
Body mass index, n (%)					
Underweight/Normal (<25 kg/m^2)	570 (44%)	505 (39%)	500 (39%)	481 (37%)	509 (39%)
Overweight (25-30 kg/m^2)	486 (37%)	489 (38%)	528 (41%)	501 (39%)	506 (39%)
Obese (≥30 kg/m^2)	244 (19%)	305 (23%)	269 (21%)	318 (24%)	283 (22%)
Physical activity (MET-hours/day), n (%)					
Q1 [0, 0.321)	269 (21%)	252 (20%)	235 (18%)	244 (19%)	244 (19%)
Q2 [0.321, 0.464)	206 (16%)	202 (16%)	184 (14%)	205 (16%)	185 (14%)
Q3 [0.464, 0.821)	273 (21%)	264 (21%)	286 (22%)	285 (22%)	283 (22%)
Q4 [0.821, 1.679)	275 (21%)	313 (24%)	297 (23%)	294 (23%)	312 (24%)
Q5 [1.679, 14.286)	259 (20%)	256 (20%)	280 (22%)	257 (20%)	256 (20%)
Smoking status and pack-years, n (%)					
Never	460 (36%)	489 (39%)	527 (42%)	533 (42%)	571 (45%)
Former, <20 pack-years	374 (29%)	394 (31%)	378 (30%)	432 (34%)	371 (30%)
Former, ≥20 pack-years	217 (17%)	180 (14%)	179 (14%)	158 (13%)	181 (14%)
Current, <20 pack-years	89 (7.0%)	88 (7.0%)	89 (7.1%)	71 (5.6%)	69 (5.5%)
Current, ≥20 pack-years	135 (11.0 %)	113 (8.9%)	86 (6.8%)	67 (5.3%)	64 (5.1%)
NSAIDs use, n (%)	519 (41%)	617 (49%)	651 (52%)	666 (53%)	690 (55%)
Total energy intake (kcal/day)					
Mean (SD)	2,082 (1,034)	2,175 (1,047)	2,276 (1,099)	2,309 (1,096)	2,022 (1,000)
Median (IQR)	1,860 (1,379, 2,545)	1,944 (1,423, 2,672)	2,042 (1,511, 2,760)	2,101 (1,549, 2,759)	1,805 (1,366, 2,386)
Comorbidities[a], n (%)					
None of diseases	700 (54%)	703 (54%)	723 (55%)	733 (56%)	747 (57%)

Characteristic	Total dairy products intake				
	Q1	Q2	Q3	Q4	Q5
1 of diseases	541 (41%)	521 (40%)	499 (38%)	483 (37%)	449 (34%)
≥2 diseases	67 (5.1%)	82 (6.3%)	84 (6.4%)	90 (6.9%)	108 (8.3%)
Radiation therapy, n (%)					
No	1,139 (87%)	1,123 (86%)	1,135 (87%)	1,154 (88%)	1,175 (90%)
Yes	100 (7.6%)	127 (9.7%)	108 (8.3%)	90 (6.9%)	76 (5.8%)
Unknown	69 (5.3%)	56 (4.3%)	63 (4.8%)	62 (4.7%)	53 (4.1%)
Chemotherapy, n (%)					
No	889 (68%)	900 (69%)	911 (70%)	946 (72%)	930 (71%)
Yes	378 (29%)	371 (28%)	352 (27%)	323 (25%)	332 (25%)
Unknown	41 (3.1%)	35 (2.7%)	43 (3.3%)	37 (2.8%)	42 (3.2%)
Tumor stage, n (%)					
In situ/non-invasive	89 (6.8%)	89 (6.8%)	73 (5.6%)	90 (6.9%)	80 (6.1%)
Localized	523 (40%)	507 (39%)	520 (40%)	509 (39%)	532 (41%)
Regional	445 (34%)	462 (35%)	463 (35%)	482 (37%)	420 (32%)
Distant	211 (16%)	200 (15%)	201 (15%)	174 (13%)	209 (16%)
Unknown	40 (3.1%)	48 (3.7%)	49 (3.8%)	51 (3.9%)	63 (4.8%)
Tumor grade, n (%)					
I	89 (6.8%)	88 (6.7%)	105 (8.0%)	83 (6.4%)	105 (8.1%)
II	763 (58%)	774 (59%)	767 (59%)	736 (56%)	741 (57%)
III	158 (12%)	159 (12%)	160 (12%)	166 (13%)	182 (14%)
IV	4 (0.3%)	8 (0.6%)	10 (0.8%)	15 (1.1%)	11 (0.8%)
Missing	294 (22%)	277 (21%)	264 (20%)	306 (23%)	265 (20%)
Tumor site, n (%)					
Colon	937 (72%)	963 (74%)	989 (76%)	992 (76%)	1,012 (78%)
Rectum	357 (27%)	325 (25%)	307 (24%)	311 (24%)	282 (22%)
Mixed	14 (1.1%)	18 (1.4%)	10 (0.8%)	3 (0.2%)	10 (0.8%)

The unit of the total dairy products intake is g/1000 kcal/day.

[a] Comorbidities consisted of heart disease, stroke, and hypertension.

Abbreviations: SD, standard deviation; IQR, interquartile range; MET, metabolic equivalent of task; NSAIDs, non-steroidal anti-inflammatory drugs.

Table 4-2. Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019

	White					African American					Native Hawaiian				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	No. of deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value
Total dairy products (g/1000 kcal/day)															
1	103	1.00 (ref)				150	1.00 (ref)				74	1.00 (ref)			
2	137	0.98 (0.75, 1.29)	0.894			204	1.06 (0.84, 1.34)	0.644			67	0.88 (0.61, 1.29)	0.524		
3	176	0.99 (0.76, 1.29)	0.961	0.97 (0.90, 1.04)	0.407	212	1.29 (1.01, 1.64)	0.038	1.10 (1.02, 1.18)	0.010	49	0.71 (0.47, 1.09)	0.119	1.04 (0.88, 1.23)	0.660
4	191	0.72 (0.56, 0.94)	0.015			201	1.10 (0.86, 1.40)	0.445			42	0.87 (0.57, 1.32)	0.507		
5	289	0.78 (0.61, 1.00)	0.050			186	1.24 (0.97, 1.59)	0.080			41	0.99 (0.63, 1.54)	0.959		
p for trend			0.002					0.093					0.721		
Milk (g/1000 kcal/day)															
1	127	1.00 (ref)				141	1.00 (ref)				45	1.00 (ref)			
2	145	1.05 (0.81, 1.35)	0.715			225	0.99 (0.79, 1.26)	0.958			81	1.31 (0.86, 2.01)	0.213		
3	183	1.00 (0.78, 1.28)	0.992	0.97 (0.91, 1.04)	0.446	229	1.27 (1.00, 1.61)	0.053	1.09 (1.02, 1.17)	0.014	56	0.99 (0.62, 1.59)	0.966	1.01 (0.84, 1.21)	0.932
4	185	0.85 (0.67, 1.09)	0.205			183	1.06 (0.83, 1.36)	0.648			51	0.80 (0.50, 1.27)	0.341		
5	256	0.79 (0.63, 0.99)	0.043			175	1.22 (0.95, 1.56)	0.122			40	1.25 (0.76, 2.06)	0.377		
p for trend			0.004					0.093					0.536		
Lactose (g/1000 kcal/day)															
1	114	1.00 (ref)				147	1.00 (ref)				67	1.00 (ref)			
2	143	0.93 (0.71, 1.21)	0.575			201	0.93 (0.73, 1.18)	0.549			65	0.77 (0.52, 1.13)	0.177		
3	162	0.96 (0.74, 1.25)	0.774	0.97 (0.90, 1.04)	0.398	216	1.27 (1.00, 1.61)	0.054	1.09 (1.02, 1.17)	0.014	54	0.65 (0.42, 0.99)	0.046	1.03 (0.87, 1.22)	0.759
4	207	0.73 (0.57, 0.93)	0.011			202	1.00 (0.79, 1.27)	0.990			44	0.81 (0.54, 1.23)	0.327		
5	270	0.75 (0.59, 0.95)	0.017			187	1.18 (0.93, 1.52)	0.178			43	0.97 (0.62, 1.51)	0.880		
p for trend			0.001					0.127					0.765		

	White					African American					Native Hawaiian				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	**No. of deaths**	**HR (95% CI)[a]**	**p-value**	**HR (95% CI)**	**p-value**	**No. of deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	**No. of deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
Total calcium (mg/1000 kcal/day)															
1	106	1.00 (ref)				188	1.00 (ref)				104	1.00 (ref)			
2	158	1.17 (0.90, 1.52)	0.241			233	0.90 (0.73, 1.12)	0.359			64	0.77 (0.54, 1.11)	0.161		
3	169	0.80 (0.61, 1.05)	0.111	1.04 (0.97, 1.12)	0.233	194	0.94 (0.75, 1.18)	0.579	1.09 (1.02, 1.17)	0.009	44	0.73 (0.49, 1.09)	0.123	0.96 (0.80, 1.16)	0.695
4	223	0.86 (0.66, 1.11)	0.243			181	1.02 (0.81, 1.29)	0.848			34	1.43 (0.91, 2.25)	0.119		
5	240	0.96 (0.73, 1.25)	0.736			157	1.42 (1.12, 1.81)	0.004			27	0.70 (0.43, 1.14)	0.156		
p for trend			0.242					0.005					0.611		
Total Vitamin D (IU/1000 kcal/day)															
1	129	1.00 (ref)				187	1.00 (ref)				58	1.00 (ref)			
2	142	0.73 (0.56, 0.95)	0.017			194	0.88 (0.70, 1.12)	0.300			77	0.68 (0.46, 1.02)	0.062		
3	193	0.67 (0.53, 0.85)	0.001	1.01 (0.95, 1.08)	0.666	225	0.90 (0.72, 1.12)	0.350	1.04 (0.98, 1.12)	0.202	71	0.67 (0.45, 1.01)	0.056	0.97 (0.80, 1.17)	0.732
4	205	0.72 (0.56, 0.91)	0.007			167	0.96 (0.76, 1.21)	0.724			36	0.57 (0.35, 0.92)	0.020		
5	227	0.84 (0.66, 1.07)	0.163			180	0.97 (0.77, 1.23)	0.816			31	0.79 (0.48, 1.31)	0.368		
p for trend			0.522					0.918					0.195		

[a] Adjusted for sex as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; BMI, body mass index; NASIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-2 (*continue*). Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019

	Japanese American					Latino					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of deaths	HR (95% CI) [a]	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)											
1	464	1.00 (ref)				90	1.00 (ref)				
2	292	0.92 (0.79, 1.07)	0.277			159	1.03 (0.78, 1.37)	0.812			
3	282	0.91 (0.78, 1.07)	0.262	0.97 (0.90, 1.04)	0.375	180	0.88 (0.67, 1.16)	0.374	1.00 (0.93, 1.07)	0.940	0.029
4	222	1.01 (0.85, 1.19)	0.937			236	1.10 (0.85, 1.43)	0.475			
5	162	0.95 (0.78, 1.15)	0.572			228	1.00 (0.76, 1.31)	0.997			
p for trend			0.743					0.702			
Milk (g/1000 kcal/day)											
1	457	1.00 (ref)				102	1.00 (ref)				
2	245	0.82 (0.69, 0.96)	0.014			147	1.05 (0.80, 1.37)	0.742			
3	259	0.85 (0.72, 0.99)	0.042	0.96 (0.90, 1.03)	0.224	188	0.94 (0.72, 1.22)	0.628	1.00 (0.94, 1.07)	0.954	0.022
4	259	0.89 (0.75, 1.04)	0.148			219	1.04 (0.81, 1.34)	0.779			
5	202	0.87 (0.73, 1.04)	0.130			237	1.04 (0.81, 1.34)	0.764			
p for trend			0.142					0.725			
Lactose (g/1000 kcal/day)											
1	447	1.00 (ref)				107	1.00 (ref)				
2	287	0.88 (0.75, 1.03)	0.100			167	1.18 (0.91, 1.53)	0.218			
3	285	0.86 (0.74, 1.01)	0.063	0.96 (0.90, 1.03)	0.275	177	0.95 (0.73, 1.24)	0.715	1.00 (0.93, 1.07)	0.989	0.014
4	228	0.92 (0.77, 1.09)	0.311			219	1.18 (0.92, 1.51)	0.201			
5	175	0.90 (0.75, 1.09)	0.291			223	1.04 (0.81, 1.34)	0.749			
p for trend			0.239					0.861			

	Japanese American					Latino					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	**No. of deaths**	**HR (95% CI)** [a]	**p-value**	**HR (95% CI)**	**p-value**	**No. of deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	
Total calcium (mg/1000 kcal/day)											
1	450	1.00 (ref)				66	1.00 (ref)				
2	268	0.88 (0.75, 1.03)	0.103			189	1.31 (0.97, 1.77)	0.078			
3	233	0.88 (0.75, 1.05)	0.155	0.94 (0.89, 0.99)	0.018	247	1.23 (0.92, 1.65)	0.166	1.01 (0.91, 1.11)	0.913	<0.001
4	223	0.95 (0.79, 1.13)	0.551			220	1.22 (0.90, 1.64)	0.199			
5	248	0.84 (0.71, 1.01)	0.060			171	1.21 (0.88, 1.66)	0.233			
p for trend			0.138					0.725			
Total vitamin D (IU/1000 kcal/day)											
1	325	1.00 (ref)				196	1.00 (ref)				
2	304	1.02 (0.86, 1.20)	0.858			195	1.04 (0.84, 1.28)	0.747			
3	213	0.90 (0.75, 1.08)	0.272	1.02 (0.96, 1.08)	0.610	190	0.92 (0.74, 1.15)	0.489	1.01 (0.93, 1.08)	0.895	0.566
4	278	0.96 (0.81, 1.14)	0.626			172	1.11 (0.88, 1.39)	0.388			
5	302	1.07 (0.90, 1.26)	0.440			140	1.00 (0.78, 1.28)	0.977			
p for trend			0.709					0.829			

[a] Adjusted for sex as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; BMI, body mass index; NASIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-3. Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019

	White					African American					Native Hawaiian				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	No. of CRC deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value
Total dairy products (g/1000 kcal/day)															
1	46	1.00 (ref)				73	1.00 (ref)				31	1.00 (ref)			
2	57	0.88 (0.57, 1.34)	0.548			92	1.02 (0.72, 1.43)	0.922			23	0.72 (0.39, 1.30)	0.274		
3	76	1.04 (0.69, 1.55)	0.866	0.98 (0.88, 1.09)	0.653	84	1.12 (0.79, 1.61)	0.525	1.18 (1.06, 1.30)	0.002	22	1.07 (0.57, 2.00)	0.825	1.22 (0.96, 1.57)	0.109
4	68	0.61 (0.41, 0.92)	0.019			82	1.00 (0.70, 1.44)	0.993			15	0.88 (0.44, 1.72)	0.701		
5	118	0.71 (0.49, 1.05)	0.084			80	1.31 (0.91, 1.88)	0.144			23	1.52 (0.80, 2.90)	0.204		
p for trend			0.013					0.199					0.235		
Milk (g/1000 kcal/day)															
1	57	1.00 (ref)				59	1.00 (ref)				16	1.00 (ref)			
2	63	0.97 (0.66, 1.43)	0.891			114	1.20 (0.84, 1.70)	0.315			37	1.37 (0.69, 2.75)	0.368		
3	65	0.81 (0.55, 1.20)	0.290	0.99 (0.89, 1.10)	0.881	93	1.29 (0.89, 1.88)	0.174	1.18 (1.06, 1.31)	0.002	21	0.88 (0.41, 1.89)	0.739	1.16 (0.89, 1.49)	0.268
4	76	0.81 (0.56, 1.18)	0.280			69	1.16 (0.78, 1.71)	0.458			19	1.03 (0.50, 2.13)	0.942		
5	104	0.70 (0.49, 0.99)	0.045			76	1.53 (1.05, 2.24)	0.028			21	1.58 (0.74, 3.38)	0.234		
p for trend			0.021					0.062					0.655		
Lactose (g/1000 kcal/day)															
1	53	1.00 (ref)				71	1.00 (ref)				26	1.00 (ref)			
2	57	0.76 (0.50, 1.15)	0.189			91	0.87 (0.61, 1.24)	0.452			25	0.72 (0.39, 1.32)	0.288		
3	72	0.96 (0.65, 1.42)	0.843	0.98 (0.88, 1.09)	0.701	85	1.17 (0.82, 1.66)	0.398	1.18 (1.06, 1.31)	0.002	23	0.93 (0.48, 1.79)	0.830	1.22 (0.96, 1.55)	0.105
4	72	0.56 (0.38, 0.83)	0.004			83	0.97 (0.67, 1.39)	0.856			16	0.90 (0.46, 1.78)	0.768		
5	111	0.65 (0.45, 0.93)	0.020			81	1.26 (0.88, 1.82)	0.213			24	1.54 (0.80, 2.97)	0.198		

	White					African American					Native Hawaiian				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	**No. of CRC deaths**	**HR (95% CI)[a]**	**p-value**	**HR (95% CI)**	**p-value**	**No. of CRC deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	**No. of CRC deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
p for trend			0.006					0.144					0.190		
Total calcium (mg/1000 kcal/day)															
1	42	1.00 (ref)				77	1.00 (ref)				44	1.00 (ref)			
2	66	1.29 (0.84, 1.98)	0.243			99	0.99 (0.71, 1.38)	0.955			25	0.91 (0.52, 1.59)	0.737		
3	63	0.93 (0.60, 1.43)	0.733	1.07 (0.95, 1.19)	0.266	92	1.19 (0.84, 1.69)	0.320	1.15 (1.05, 1.27)	0.003	19	0.74 (0.41, 1.35)	0.331	0.94 (0.69, 1.29)	0.712
4	94	1.07 (0.71, 1.61)	0.757			71	0.99 (0.69, 1.42)	0.951			18	1.98 (1.03, 3.80)	0.041		
5	100	1.09 (0.71, 1.66)	0.696			72	1.91 (1.33, 2.75)	<0.001			8	0.56 (0.24, 1.35)	0.198		
p for trend			0.958					0.004					0.834		
Total vitamin D (IU/1000 kcal/day)															
1	61	1.00 (ref)				84	1.00 (ref)				22	1.00 (ref)			
2	44	0.45 (0.30, 0.69)	<0.001			78	0.95 (0.67, 1.34)	0.763			31	0.69 (0.37, 1.28)	0.240		
3	80	0.60 (0.42, 0.86)	0.006	0.94 (0.84, 1.06)	0.320	102	0.98 (0.71, 1.35)	0.898	1.05 (0.95, 1.16)	0.319	29	0.85 (0.45, 1.61)	0.626	0.91 (0.67, 1.23)	0.536
4	91	0.67 (0.47, 0.96)	0.031			70	0.98 (0.69, 1.41)	0.932			20	0.89 (0.44, 1.80)	0.750		
5	89	0.60 (0.41, 0.88)	0.008			77	1.00 (0.70, 1.41)	0.986			12	0.79 (0.35, 1.76)	0.566		
p for trend			0.291					0.939					0.940		

[a] Adjusted for sex as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; BMI, body mass index; NASIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-3 (*continue*). Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different race/ethnicity, Multiethnic Cohort Study, 1993-2019

	Japanese American					Latino					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of CRC deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)											
1	198	1.00 (ref)				39	1.00 (ref)				
2	113	0.73 (0.57, 0.93)	0.010			70	0.92 (0.60, 1.41)	0.690			
3	98	0.74 (0.57, 0.95)	0.019	0.87 (0.77, 0.98)	0.027	75	0.90 (0.59, 1.36)	0.608	0.92 (0.82, 1.04)	0.194	0.108
4	87	0.84 (0.64, 1.09)	0.193			98	1.10 (0.74, 1.65)	0.642			
5	58	0.78 (0.57, 1.08)	0.136			88	0.84 (0.55, 1.28)	0.410			
p for trend			0.108					0.785			
Milk (g/1000 kcal/day)											
1	206	1.00 (ref)				40	1.00 (ref)				
2	90	0.69 (0.53, 0.88)	0.004			66	1.02 (0.67, 1.57)	0.909			
3	93	0.81 (0.51, 0.86)	0.002	0.87 (0.77, 0.98)	0.020	88	1.09 (0.73, 1.63)	0.689	0.93 (0.83, 1.04)	0.230	0.051
4	94	0.81 (0.57, 0.95)	0.020			90	1.15 (0.77, 1.71)	0.498			
5	71	0.70 (0.54, 0.97)	0.029			86	0.89 (0.59, 1.34)	0.585			
p for trend			0.010					0.661			
Lactose (g/1000 kcal/day)											
1	196	1.00 (ref)				42	1.00 (ref)				
2	110	0.74 (0.58, 0.94)	0.014			76	1.19 (0.79, 1.80)	0.414			
3	99	0.96 (0.52, 0.87)	0.002	0.87 (0.77, 0.98)	0.027	78	1.15 (0.77, 1.73)	0.501	0.93 (0.83, 1.01)	0.217	0.029
4	87	0.56 (0.60, 1.01)	0.062			92	1.32 (0.89, 1.96)	0.171			
5	62	0.65 (0.55, 1.01)	0.061			82	0.92 (0.61, 1.39)	0.686			
p for trend			0.022					0.694			

	Japanese American					Latino					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of CRC deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total calcium (mg/1000 kcal/day)											
1	179	1.00 (ref)				25	1.00 (ref)				
2	106	0.85 (0.66, 1.10)	0.221			90	1.69 (1.04, 2.76)	0.034			
3	84	0.88 (0.67, 1.16)	0.362	0.94 (0.86, 1.03)	0.169	102	1.54 (0.95, 2.49)	0.078	0.93 (0.79, 1.11)	0.424	0.001
4	96	0.98 (0.75, 1.29)	0.909			86	1.33 (0.80, 2.19)	0.267			
5	89	0.80 (0.60, 1.07)	0.132			67	1.28 (0.76, 2.15)	0.357			
p for trend			0.304					0.626			
Total vitamin D (IU/1000 kcal/day)											
1	130	1.00 (ref)				85	1.00 (ref)				
2	126	1.11 (0.86, 1.43)	0.443			80	1.07 (0.77, 1.49)	0.682			
3	75	0.89 (0.66, 1.20)	0.443	1.00 (0.91, 1.10)	0.957	74	0.95 (0.67, 1.34)	0.759	1.03 (0.92, 1.16)	0.582	0.209
4	114	0.99 (0.75, 1.29)	0.915			77	1.06 (0.75, 1.50)	0.754			
5	109	1.06 (0.81, 1.39)	0.671			54	0.99 (0.67, 1.45)	0.939			
p for trend			0.998					0.943			

[a] Adjusted for sex as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; BMI, body mass index; NASIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-4. Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different age groups, Multiethnic Cohort Study, 1993-2019

	<60 years old					60-70 years old					≥70 years old					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)																
1	82	1.00 (ref)				237	1.00 (ref)				562	1.00 (ref)				
2	65	0.91 (0.64, 1.31)	0.626			229	0.80 (0.66, 0.97)	0.026			565	1.04 (0.92, 1.19)	0.498			
3	50	0.95 (0.63, 1.44)	0.799	1.03 (0.88, 1.20)	0.743	189	0.86 (0.70, 1.07)	0.176	0.93 (0.82, 1.01)	0.072	660	1.03 (0.91, 1.16)	0.686	1.03 (0.99, 1.07)	0.195	0.420
4	42	0.84 (0.55, 1.28)	0.422			172	0.80 (0.65, 1.00)	0.049			678	1.04 (0.92, 1.18)	0.486			
5	30	1.12 (0.68, 1.85)	0.650			164	0.77 (0.62, 0.97)	0.026			712	1.04 (0.92, 1.18)	0.495			
p for trend			0.931					0.044					0.554			
Milk (g/1000 kcal/day)																
1	78	1.00 (ref)				211	1.00 (ref)				583	1.00 (ref)				
2	66	1.08 (0.75, 1.56)	0.687			233	0.94 (0.77, 1.15)	0.522			544	0.91 (0.80, 1.03)	0.137			
3	53	1.06 (0.72, 1.57)	0.769	1.05 (0.90, 1.23)	0.520	217	0.88 (0.72, 1.09)	0.243	0.94 (0.86, 1.02)	0.116	645	0.97 (0.86, 1.10)	0.661	1.01 (0.98, 1.05)	0.497	0.829
4	42	0.99 (0.66, 1.49)	0.959			168	0.81 (0.65, 1.01)	0.057			687	0.94 (0.84, 1.06)	0.333			
5	30	1.16 (0.71, 1.89)	0.564			162	0.83 (0.66, 1.05)	0.114			718	0.95 (0.84, 1.07)	0.363			
p for trend			0.769					0.040					0.593			
Lactose (g/1000 kcal/day)																
1	89	1.00 (ref)				230	1.00 (ref)				563	1.00 (ref)				
2	60	0.84 (0.59, 1.21)	0.349			233	0.87 (0.71, 1.06)	0.160			570	0.95 (0.84, 1.08)	0.440			
3	51	0.92 (0.62, 1.38)	0.703	1.03 (0.88. 1.21)	0.685	202	0.93 (0.76, 1.15)	0.506	0.93 (0.86, 1.01)	0.095	641	0.96 (0.85, 1.08)	0.468	1.02 (0.98, 1.06)	0.307	0.729
4	38	0.77 (0.50, 1.18)	0.230			171	0.83 (0.67, 1.04)	0.100			691	0.95 (0.84, 1.07)	0.405			
5	31	1.17 (0.72, 1.89)	0.533			155	0.78 (0.62, 0.98)	0.035			712	0.98 (0.86, 1.11)	0.722			
p for trend			0.891					0.043					0.786			
Total calcium (mg/1000kcal/day)																
1	98	1.00 (ref)		0.99 (0.82, 1.20)	0.954	248	1.00 (ref)		0.94 (0.85, 1.03)	0.192	568	1.00 (ref)		1.00 (0.97, 1.04)	0.966	0.011
2	59	0.72	0.079			236	1.03	0.786			617	0.95	0.454			

	<60 years old					60-70 years old					≥70 years old					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
		(0.50, 1.04)					(0.85, 1.25)					(0.84, 1.08)				
3	45	0.92 (0.61, 1.38)	0.685			187	0.93 (0.76, 1.15)	0.528			655	0.90 (0.79, 1.02)	0.098			
4	37	0.94 (0.59, 1.50)	0.788			186	0.91 (0.74, 1.13)	0.416			658	0.99 (0.88, 1.13)	0.911			
5	30	1.26 (0.74, 2.16)	0.392			134	0.72 (0.57, 0.91)	0.007			679	1.03 (0.91, 1.17)	0.613			
p for trend			0.536					0.008					0.393			
Total Vitamin D (IU/1000kcal/day)																
1	87	1.00 (ref)				244	1.00 (ref)				564	1.00 (ref)				
2	59	0.93 (0.65, 1.34)	0.703			218	0.97 (0.79, 1.18)	0.754			635	0.92 (0.81, 1.04)	0.184			
3	47	0.90 (0.61, 1.33)	0.600	0.96 (0.82, 1.12)	0.604	176	0.93 (0.75, 1.15)	0.515	0.98 (0.91, 1.05)	0.593	669	0.85 (0.75, 0.96)	0.009	1.03 (0.99, 1.07)	0.107	0.600
4	41	0.86 (0.56, 1.31)	0.475			180	0.83 (0.67, 1.02)	0.077			637	0.96 (0.84, 1.08)	0.479			
5	35	0.86 (0.54, 1.36)	0.518			173	0.99 (0.80, 1.22)	0.913			672	0.99 (0.87, 1.12)	0.870			
p for trend			0.405					0.372					0.756			

[a] Adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; BMI, body mass index; NASIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-5. Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different age groups, Multiethnic Cohort Study, 1993-2019

	<60 years old					60-70 years old					≥70 years old					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of CRC deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)																
1	43	1.00 (ref)				110	1.00 (ref)				234	1.00 (ref)				
2	37	0.94 (0.57, 1.55)	0.811			100	0.69 (0.52, 0.93)	0.013			218	0.92 (0.76, 1.13)	0.434			
3	28	1.04 (0.59, 1.84)	0.879	0.95 (0.73, 1.23)	0.694	87	0.80 (0.59, 1.09)	0.157	0.89 (0.79, 1.01)	0.064	240	0.93 (0.77, 1.13)	0.467	1.01 (0.95, 1.08)	0.716	0.518
4	24	0.79 (0.44, 1.43)	0.439			71	0.61 (0.44, 0.85)	0.003			255	0.95 (0.78, 1.15)	0.571			
5	13	0.96 (0.46, 2.04)	0.925			77	0.68 (0.49, 0.95)	0.023			277	0.94 (0.77, 1.14)	0.532			
p for trend			0.650					0.017					0.669			
Milk (g/1000 kcal/day)																
1	40	1.00 (ref)				100	1.00 (ref)				238	1.00 (ref)				
2	38	1.10 (0.65, 1.84)	0.729			116	0.88 (0.66, 1.18)	0.395			216	0.89 (0.73, 1.08)	0.241			
3	32	1.13 (0.67, 1.93)	0.643	0.96 (0.75, 1.24)	0.772	88	0.72 (0.53, 0.98)	0.038	0.88 (0.77, 1.00)	0.045	240	0.89 (0.73, 1.09)	0.255	1.01 (0.95, 1.07)	0.734	0.400
4	21	0.90 (0.50, 1.63)	0.725			68	0.61 (0.44, 0.85)	0.004			259	0.95 (0.78, 1.14)	0.560			
5	14	0.95 (0.47, 1.92)	0.883			73	0.65 (0.46, 0.91)	0.013			271	0.89 (0.74, 1.08)	0.240			
p for trend			0.754					0.001					0.459			
Lactose (g/1000 kcal/day)																
1	49	1.00 (ref)				108	1.00 (ref)				231	1.00 (ref)				
2	32	0.77 (0.46, 1.26)	0.298			106	0.80 (0.60, 1.06)	0.119			221	0.86 (0.70, 1.05)	0.128			
3	27	0.95 (0.55, 1.63)	0.843	0.96 (0.74, 1.24)	0.741	90	0.87 (0.64, 1.17)	0.358	0.89 (0.79, 1.01)	0.079	240	0.89 (0.74, 1.09)	0.264	1.01 (0.95, 1.08)	0.714	0.575
4	22	0.68 (0.38, 1.21)	0.187			68	0.60 (0.43, 0.83)	0.002			260	0.89 (0.73, 1.08)	0.235			
5	15	1.05 (0.53, 2.10)	0.886			73	0.68 (0.48, 0.95)	0.023			272	0.88 (0.73, 1.07)	0.210			
p for trend			0.620					0.005					0.376			
Total calcium (mg/1000kcal/day)																
1	49	1.00 (ref)		1.04	0.791	110	1.00 (ref)		0.92	0.267	208	1.00		1.02	0.582	0.360

	<60 years old					60-70 years old					≥70 years old					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of CRC deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
2	35	0.84 (0.51, 1.38)	0.487	(0.79, 1.36)		106	0.93 (0.70, 1.25)	0.646	(0.79, 1.07)		245	1.04 (0.85, 1.27)	0.708	(0.96, 1.08)		
3	27	1.17 (0.67, 2.04)	0.573			80	0.91 (0.66, 1.24)	0.540			253	1.00 (0.82, 1.22)	1.000			
4	21	1.20 (0.64, 2.24)	0.573			89	0.83 (0.60, 1.14)	0.252			255	1.07 (0.87, 1.31)	0.524			
5	13	1.17 (0.54, 2.53)	0.698			60	0.64 (0.45, 0.92)	0.015			263	1.10 (0.89, 1.35)	0.381			
p for trend			0.417					0.018					0.363			
Total Vitamin D (IU/1000kcal/day)																
1	48	1.00 (ref)				109	1.00 (ref)				225	1.00 (ref)				
2	34	0.98 (0.60, 1.60)	0.924			101	1.00 (0.75, 1.34)	0.985			224	0.85 (0.70, 1.04)	0.114			
3	22	0.72 (0.41, 1.27)	0.260	0.95 (0.76, 1.19)	0.660	84	0.95 (0.70, 1.30)	0.762	0.91 (0.81, 1.02)	0.106	254	0.87 (0.72, 1.06)	0.171	1.03 (0.97, 1.09)	0.400	0.138
4	21	0.80 (0.44, 1.46)	0.472			81	0.70 (0.51, 0.96)	0.029			270	1.03 (0.85, 1.26)	0.731			
5	20	1.02 (0.55, 1.87)	0.961			70	0.80 (0.57, 1.11)	0.173			251	0.94 (0.77, 1.14)	0.513			
p for trend			0.624					0.025					0.709			

[a] Adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; BMI, body mass index; NASIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-6. Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor stage groups, Multiethnic Cohort Study, 1993-2019

	In situ/non-invasive					Localized				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	No. of deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value
Total dairy products (g/1000 kcal/day)										
1	46	1.00 (ref)				296	1.00 (ref)			
2	48	0.67 (0.41, 1.10)	0.113			265	0.96 (0.80, 1.15)	0.678		
3	43	1.49 (0.90, 2.48)	0.121	1.05 (0.89, 1.22)	0.580	299	0.89 (0.75, 1.07)	0.212	1.04 (0.98, 1.10)	0.207
4	59	1.49 (0.92, 2.40)	0.102			286	0.94 (0.78, 1.13)	0.503		
5	43	0.99 (0.59, 1.63)	0.955			330	1.02 (0.85, 1.22)	0.836		
p for trend			0.199					0.883		
Milk (g/1000 kcal/day)										
1	41	1.00 (ref)				284	1.00 (ref)			
2	48	0.77 (0.46, 1.29)	0.328			251	0.95 (0.79, 1.14)	0.603		
3	42	1.12 (0.67, 1.87)	0.674	1.04 (0.90, 1.21)	0.585	324	1.02 (0.86, 1.21)	0.825	1.04 (0.99. 1.10)	0.152
4	59	1.49 (0.93, 2.38)	0.100			273	0.87 (0.72, 1.04)	0.129		
5	49	1.02 (0.61, 1.68)	0.950			344	1.09 (0.92, 1.30)	0.320		
p for trend			0.237					0.584		
Lactose (g/1000 kcal/day)										
1	42	1.00 (ref)				290	1.00 (ref)			
2	47	0.64 (0.39, 1.05)	0.077			280	0.98 (0.82, 1.17)	0.785		
3	49	1.37 (0.83, 2.28)	0.223	1.06 (0.91, 1.23)	0.475	281	0.85 (0.71, 1.02)	0.078	1.03 (0.98, 1.09)	0.282
4	59	1.50 (0.93, 2.41)	0.099			303	0.90 (0.76, 1.08)	0.275		
5	42	1.01 (0.60, 1.69)	0.981			322	1.00 (0.84, 1.20)	0.971		
p for trend			0.145					0.807		

	In situ/non-invasive					Localized				
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)	
Quintile	**No. of deaths**	**HR (95% CI)[a]**	**p-value**	**HR (95% CI)**	**p-value**	**No. of deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
Total calcium (mg/1000kcal/day)										
1	58	1.00 (ref)				318	1.00 (ref)			
2	49	1.11 (0.71, 1.75)	0.644			277	0.83 (0.69, 0.99)	0.036		
3	50	1.00 (0.63, 1.59)	0.994	0.98 (0.82, 1.16)	0.788	313	0.79 (0.67, 0.95)	0.010	0.94 (0.88, 0.99)	0.029
4	47	1.09 (0.67, 1.75)	0.739			282	0.83 (0.69, 1.00)	0.053		
5	35	1.16 (0.68, 1.98)	0.580			286	0.82 (0.68, 0.98)	0.033		
p for trend			0.653					0.071		
Total Vitamin D (IU/1000kcal/day)										
1	50	1.00 (ref)				286	1.00 (ref)			
2	50	0.79 (0.49, 1.29)	0.352			308	1.00 (0.84, 1.19)	0.972		
3	52	0.77 (0.47, 1.26)	0.307	0.96 (0.82, 1.13)	0.654	308	0.88 (0.73, 1.05)	0.143	1.00 (0.95, 1.06)	0.997
4	49	0.96 (0.60, 1.55)	0.878			273	0.84 (0.70, 1.01)	0.065		
5	38	0.75 (0.44, 1.29)	0.305			301	0.99 (0.83, 1.18)	0.903		
p for trend			0.633					0.340		

[a] Adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; BMI, body mass index; NASIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-6 (*continue*). Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor stage groups, Multiethnic Cohort Study, 1993-2019

	Regional					Distant					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)											
1	306	1.00 (ref)				201	1.00 (ref)				
2	322	1.00 (0.84, 1.19)	0.994			183	1.03 (0.80, 1.34)	0.812			
3	330	1.00 (0.84, 1.19)	0.999	0.99 (0.93, 1.05)	0.647	184	1.26 (0.97, 1.65)	0.086	0.98 (0.90, 1.06)	0.594	0.032
4	335	0.96 (0.81, 1.14)	0.646			166	1.33 (1.02, 1.73)	0.032			
5	280	0.93 (0.77, 1.12)	0.426			198	0.98 (0.75, 1.27)	0.856			
p for trend			0.374					0.525			
Milk (g/1000 kcal/day)											
1	319	1.00 (ref)				195	1.00 (ref)				
2	303	0.93 (0.79, 1.11)	0.421			194	0.95 (0.74, 1.21)	0.665			
3	332	1.00 (0.84, 1.18)	0.981	0.97 (0.91, 1.03)	0.339	175	0.90 (0.69, 1.18)	0.453	0.96 (0.89, 1.05)	0.376	0.060
4	347	0.89 (0.75, 1.05)	0.170			177	1.07 (0.83, 1.39)	0.597			
5	272	0.83 (0.69, 0.99)	0.038			191	0.89 (0.69, 1.15)	0.373			
p for trend			0.038					0.704			
Lactose (g/1000 kcal/day)											
1	314	1.00 (ref)				202	1.00 (ref)				
2	315	0.95 (0.80, 1.13)	0.559			181	1.00 (0.78, 1.30)	0.975			
3	333	0.98 (0.83, 1.16)	0.833	0.98 (0.92, 1.04)	0.541	189	1.00 (0.77, 1.30)	0.993	0.97 (0.89, 1.05)	0.403	0.022
4	324	0.89 (0.75, 1.06)	0.201			165	1.17 (0.90, 1.52)	0.244			
5	287	0.89 (0.74, 1.06)	0.194			195	0.92 (0.70, 1.20)	0.540			
p for trend			0.145					0.976			

	Regional					Distant					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total calcium (mg/1000kcal/day)											
1	316	1.00 (ref)				191	1.00 (ref)				
2	335	0.87 (0.74, 1.04)	0.121			208	1.32 (1.03, 1.70)	0.029			
3	315	0.92 (0.78, 1.10)	0.381	1.02 (0.96, 1.08)	0.534	163	1.23 (0.92, 1.62)	0.158	1.10 (1.01, 1.20)	0.028	0.259
4	303	0.92 (0.77, 1.10)	0.359			198	1.08 (0.83, 1.40)	0.552			
5	304	0.96 (0.80, 1.15)	0.664			172	1.41 (1.06, 1.88)	0.018			
p for trend			0.911					0.197			
Total Vitamin D (IU/1000kcal/day)											
1	322	1.00 (ref)				209	1.00 (ref)				
2	337	0.95 (0.81, 1.13)	0.584			179	0.92 (0.71, 1.18)	0.512			
3	304	0.91 (0.76, 1.09)	0.298	1.02 (0.97, 1.08)	0.416	173	0.83 (0.64, 1.08)	0.167	1.10 (1.00, 1.22)	0.048	0.031
4	297	0.97 (0.82, 1.16)	0.767			185	1.00 (0.77, 1.30)	0.979			
5	313	1.03 (0.87, 1.23)	0.741			186	1.04 (0.79, 1.36)	0.791			
p for trend			0.700					0.666			

[a] Adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; BMI, body mass index; NASIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-7. Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor stage groups, Multiethnic Cohort Study, 1993-2019

	Localized					Regional					Distant					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of CRC deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)																
1	52	1.00 (ref)				148	1.00 (ref)				166	1.00 (ref)				
2	45	0.77 (0.50, 1.19)	0.244			138	0.84 (0.65, 1.08)	0.178			151	1.06 (0.79, 1.41)	0.713			
3	58	1.02 (0.67, 1.54)	0.936	1.09 (0.95, 1.25)	0.202	136	0.92 (0.71, 1.18)	0.505	0.98 (0.90, 1.08)	0.745	136	1.16 (0.86, 1.58)	0.328	1.00 (0.92, 1.10)	0.938	0.606
4	41	0.68 (0.43, 1.07)	0.093			154	0.95 (0.74, 1.22)	0.689			127	1.30 (0.97, 1.75)	0.078			
5	60	1.06 (0.69, 1.62)	0.788			127	0.88 (0.67, 1.16)	0.377			151	1.00 (0.74, 1.34)	0.981			
p for trend			0.941					0.709					0.550			
Milk (g/1000 kcal/day)																
1	42	1.00 (ref)				153	1.00 (ref)				160	1.00 (ref)				
2	51	1.09 (0.70, 1.69)	0.716			138	0.86 (0.67, 1.10)	0.236			161	0.91 (0.69, 1.21)	0.532			
3	63	1.24 (0.81, 1.90)	0.324	1.11 (0.97, 1.26)	0.123	141	0.90 (0.70, 1.16)	0.411	0.96 (0.88, 1.06)	0.449	129	0.81 (0.60, 1.09)	0.164	0.99 (0.91, 1.09)	0.886	0.079
4	39	0.79 (0.49, 1.27)	0.330			157	0.91 (0.72, 1.16)	0.456			128	0.98 (0.73, 1.30)	0.866			
5	61	1.34 (0.87, 2.06)	0.179			114	0.74 (0.57, 0.97)	0.032			153	0.92 (0.69, 1.23)	0.571			
p for trend			0.486					0.097								
Lactose (g/1000 kcal/day)																
1	53	1.00 (ref)				150	1.00 (ref)				165	1.00 (ref)				
2	47	0.82 (0.54, 1.26)	0.366			141	0.87 (0.68, 1.12)	0.277			149	1.00 (0.75, 1.33)	0.982			
3	52	0.83 (0.54, 1.27)	0.397	1.10 (0.88, 1.25)	0.181	134	0.92 (0.72, 1.19)	0.538	0.98 (0.89, 1.07)	0.648	146	0.99 (0.73, 1.34)	0.950	0.99 (0.91, 1.08)	0.901	0.418
4	45	0.68 (0.44, 1.05)	0.082			155	0.95 (0.74, 1.22)	0.695			119	1.11 (0.83, 1.51)	0.479			
5	59	1.06 (0.70, 1.60)	0.792			123	0.82 (0.62, 1.07)	0.150			152	0.97 (0.72, 1.31)	0.832			
p for trend			0.942					0.351					0.876			
Total calcium (mg/1000kcal/day)																
1	54	1.00 (ref)		1.01	0.941	139	1.00 (ref)		0.95	0.739	156	1.00 (ref)		1.12	0.023	0.287

	Localized					Regional					Distant					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of CRC deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
2	55	1.00 (0.66, 1.52)	1.000	(0.88, 1.15)		147	0.96 (0.75, 1.24)	0.766	(0.90, 1.08)		162	1.25 (0.94, 1.66)	0.119	(1.02, 1.24)		
3	50	0.77 (0.50, 1.18)	0.228			157	1.12 (0.87, 1.44)	0.382			128	1.21 (0.89, 1.66)	0.227			
4	45	0.84 (0.54, 1.32)	0.456			131	0.92 (0.70, 1.20)	0.530			156	1.07 (0.80, 1.44)	0.628			
5	52	1.00 (0.64, 1.57)	0.990			129	0.96 (0.73, 1.27)	0.791			129	1.40 (1.02, 1.92)	0.040			
p for trend			0.742					0.696					0.211			
Total Vitamin D (IU/1000kcal/day)																
1	43	1.00 (ref)				154	1.00 (ref)				169	1.00 (ref)				
2	65	1.43 (0.94, 2.18)	0.095			139	0.92 (0.72, 1.18)	0.524			133	0.84 (0.63, 1.13)	0.253			
3	52	1.13 (0.73, 1.75)	0.573	0.98 (0.85, 1.13)	0.754	148	1.00 (0.78, 1.28)	0.993	0.98 (0.91, 1.07)	0.710	130	0.80 (0.59, 1.07)	0.126	1.12 (1.00, 1.24)	0.051	0.087
4	48	1.01 (0.64, 1.60)	0.962			132	0.90 (0.70, 1.15)	0.398			160	1.12 (0.84, 1.50)	0.430			
5	48	1.12 (0.72, 1.77)	0.610			130	0.90 (0.69, 1.16)	0.416			139	0.98 (0.73, 1.32)	0.896			
p for trend			0.737					0.403					0.509			

Unable to conduct analysis on in situ/non-invasive tumor stage due to insufficient sample size.

[a] Adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; BMI, body mass index; NASIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Table 4-8. Hazard of overall survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor grade groups, Multiethnic Cohort Study, 1993-2019

	Grade I					Grade II					Grade III					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)																
1	53	1.00 (ref)				546	1.00 (ref)				126	1.00 (ref)				
2	51	0.69 (0.44, 1.09)	0.113			521	0.95 (0.84, 1.08)	0.461			128	0.90 (0.67, 1.21)	0.491			
3	60	0.62 (0.39, 0.96)	0.033	1.13 (0.98, 1.31)	0.100	532	0.92 (0.81, 1.05)	0.208	0.98 (0.94, 1.03)	0.506	130	0.92 (0.69, 1.24)	0.601	0.99 (0.90, 1.09)	0.835	0.02
4	55	0.93 (0.60, 1.43)	0.739			506	0.89 (0.78, 1.02)	0.099			127	0.91 (0.68, 1.22)	0.543			
5	73	0.95 (0.61, 1.48)	0.816			511	0.90 (0.78, 1.03)	0.129			146	0.93 (0.70, 1.25)	0.635			
p for trend			0.673					0.080					0.708			
Milk (g/1000 kcal/day)																
1	53	1.00 (ref)				528	1.00 (ref)				139	1.00 (ref)				
2	53	0.86 (0.54, 1.36)	0.525			498	0.90 (0.79, 1.03)	0.126			116	0.91 (0.68, 1.22)	0.523			
3	53	0.71 (0.46, 1.10)	0.127	1.07 (0.93, 1.23)	0.338	565	1.00 (0.88, 1.14)	0.993	1.00 (0.95, 1.04)	0.853	135	0.79 (0.60, 1.05)	0.105	0.98 (0.89, 1.08)	0.729	0.07
4	67	1.00 (0.65, 1.55)	0.986			502	0.88 (0.77, 1.01)	0.060			126	0.83 (0.63, 1.10)	0.190			
5	66	0.97 (0.63, 1.50)	0.885			523	0.91 (0.79, 1.04)	0.158			141	0.91 (0.69, 1.19)	0.481			
p for trend			0.824					0.164					0.354			
Lactose (g/1000 kcal/day)																
1	53	1.00 (ref)				550	1.00 (ref)				128	1.00 (ref)				
2	50	0.71 (0.45, 1.12)	0.143			509	0.86 (0.75, 0.97)	0.019			132	1.04 (0.78, 1.39)	0.793			
3	55	0.63 (0.40, 0.97)	0.037	1.12 (0.97, 1.29)	0.110	536	0.88 (0.77, 1.00)	0.048	0.98 (0.94, 1.03)	0.458	127	1.06 (0.78, 1.43)	0.712	0.99 (0.90, 1.10)	0.916	0.001
4	65	1.03 (0.67, 1.56)	0.907			511	0.82 (0.71, 0.93)	0.003			128	1.04 (0.78, 1.38)	0.804			
5	69	1.01 (0.65, 1.57)	0.975			510	0.84 (0.74, 0.96)	0.013			142	0.97 (0.73, 1.30)	0.856			

	Grade I					Grade II					Grade III					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	**No. of deaths**	**HR (95% CI)[a]**	**p-value**	**HR (95% CI)**	**p-value**	**No. of deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	**No. of deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	
p for trend			0.471					0.013					0.858			
							Total calcium (mg/1000kcal/day)									
1	56	1.00 (ref)				585	1.00 (ref)				108	1.00 (ref)				
2	57	1.08 (0.69, 1.67)	0.747			541	0.88 (0.78, 1.00)	0.051			138	1.01 (0.74, 1.37)	0.958			
3	67	1.22 (0.79, 1.89)	0.366	0.94 (0.82, 1.06)	0.313	496	0.78 (0.68, 0.89)	<0.001	0.98 (0.94, 1.03)	0.387	138	1.27 (0.94, 1.72)	0.116	1.08 (0.98, 1.20)	0.129	0.004
4	55	1.18 (0.75, 1.85)	0.472			517	0.87 (0.76, 1.00)	0.044			124	1.05 (0.75, 1.46)	0.771			
5	57	0.95 (0.60, 1.53)	0.846			477	0.82 (0.72, 0.95)	0.007			149	1.53 (1.11, 2.10)	0.009			
p for trend			0.998					0.014					0.009			
							Total Vitamin D (IU/1000kcal/day)									
1	55	1.00 (ref)				560	1.00 (ref)				127	1.00 (ref)				
2	58	1.14 (0.72, 1.81)	0.577			548	0.92 (0.81, 1.04)	0.173			134	0.91 (0.68, 1.21)	0.509			
3	59	0.97 (0.62, 1.51)	0.881	0.99 (0.87, 1.12)	0.830	529	0.86 (0.75, 0.98)	0.020	1.02 (0.98, 1.07)	0.322	106	1.02 (0.75, 1.39)	0.880	1.03 (0.94, 1.13)	0.547	0.383
4	69	1.07 (0.70, 1.64)	0.759			450	0.81 (0.71, 0.93)	0.003			160	1.06 (0.80, 1.41)	0.678			
5	51	1.31 (0.81, 2.12)	0.277			529	0.95 (0.83, 1.08)	0.438			130	1.10 (0.82, 1.48)	0.516			
p for trend			0.449					0.152					0.303			

[a] Adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; NASIDs, non-steroidal anti-inflammatory drugs.

Table 4-9. Hazard of CRC-specific survival among CRC patients according to quintile of dairy, milk, lactose, calcium, and vitamin D intake in different tumor grade groups, Multiethnic Cohort Study, 1993-2019

	Grade I					Grade II					Grade III					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	No. of CRC deaths	HR (95% CI)[a]	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	No. of CRC deaths	HR (95% CI)	p-value	HR (95% CI)	p-value	
Total dairy products (g/1000 kcal/day)																
1	13	1.00 (ref)				244	1.00 (ref)				85	1.00 (ref)				
2	18	0.93 (0.36, 2.40)	0.874			204	0.80 (0.65, 0.97)	0.026			81	0.78 (0.54, 1.14)	0.202			
3	15	0.48 (0.17, 1.36)	0.168	0.90 (0.65, 1.25)	0.538	214	0.91 (0.74, 1.11)	0.336	0.98 (0.91, 1.05)	0.555	75	0.93 (0.64, 1.35)	0.699	0.98 (0.87, 1.11)	0.771	0.153
4	14	0.82 (0.32, 2.10)	0.674			198	0.81 (0.66, 1.00)	0.051			73	0.78 (0.53, 1.14)	0.193			
5	17	0.67 (0.24, 1.86)	0.442			196	0.83 (0.67, 1.02)	0.081			91	0.86 (0.59, 1.24)	0.411			
p for trend			0.443					0.139					0.441			
Milk (g/1000 kcal/day)																
1	17	1.00 (ref)				223	1.00 (ref)				90	1.00 (ref)				
2	16	1.14 (0.43, 3.02)	0.797			222	0.95 (0.78, 1.16)	0.596			76	0.83 (0.57, 1.21)	0.342			
3	13	0.53 (0.19, 1.49)	0.230	0.84 (0.61, 1.17)	0.307	225	0.97 (0.79, 1.18)	0.736	0.99 (0.92, 1.06)	0.735	78	0.76 (0.53, 1.10)	0.143	1.00 (0.88, 1.13)	0.976	0.669
4	17	0.85 (0.34, 2.13)	0.730			194	0.91 (0.74, 1.12)	0.384			73	0.78 (0.54, 1.12)	0.179			
5	14	0.70 (0.26, 1.84)	0.467			192	0.86 (0.70, 1.07)	0.175			88	0.86 (0.60, 1.22)	0.399			
p for trend			0.392					0.175					0.345			
Lactose (g/1000 kcal/day)																
1	18	1.00 (ref)				242	1.00 (ref)				84	1.00 (ref)				
2	15	0.45 (0.16, 1.22)	0.115			205	0.77 (0.63, 0.94)	0.010			83	0.97 (0.67, 1.40)	0.869			
3	12	0.29 (0.10, 0.85)	0.023	0.89 (0.65, 1.22)	0.460	219	0.90 (0.74, 1.09)	0.278	0.98 (0.91, 1.05)	0.577	73	1.10 (0.75, 1.62)	0.630	1.00 (0.89, 1.13)	0.974	0.077
4	16	0.60 (0.25, 1.43)	0.248			196	0.76 (0.62, 0.94)	0.011			80	1.07 (0.74, 1.56)	0.710			
5	16	0.55 (0.21, 1.44)	0.223			194	0.79 (0.64, 0.98)	0.032			85	0.87 (0.60, 1.27)	0.466			
p for trend			0.308					0.054					0.665			

	Grade I					Grade II					Grade III					p-heterogeneity
	Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		Food intake as a categorical variable			Food intake as a continuous variable (per SD increase)		
Quintile	**No. of CRC deaths**	**HR (95% CI)[a]**	**p-value**	**HR (95% CI)**	**p-value**	**No. of CRC deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	**No. of CRC deaths**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	
Total calcium (mg/1000kcal/day)																
1	15	1.00 (ref)				242	1.00 (ref)				64	1.00 (ref)				
2	16	0.94 (0.33, 2.69)	0.910			235	0.94 (0.77, 1.14)	0.500			83	1.19 (0.80, 1.79)	0.394			
3	17	0.91 (0.33, 2.49)	0.852	0.54 (0.33, 0.90)	0.018	193	0.80 (0.65, 0.99)	0.037	0.98 (0.91, 1.05)	0.515	92	1.64 (1.11, 2.43)	0.014	1.08 (0.94, 1.23)	0.287	0.001
4	20	1.13 (0.43, 2.94)	0.808			207	0.91 (0.74, 1.12)	0.392			77	1.08 (0.70, 1.67)	0.721			
5	9	0.22 (0.06, 0.84)	0.026			179	0.80 (0.64, 1.00)	0.048			89	1.72 (1.13, 2.62)	0.011			
p for trend			0.099					0.063					0.029			
Total Vitamin D (IU/1000kcal/day)																
1	19	1.00 (ref)				238	1.00 (ref)				80	1.00 (ref)				
2	15	0.69 (0.26, 1.85)	0.464			212	0.90 (0.74, 1.10)	0.311			82	0.89 (0.61, 1.28)	0.523			
3	14	0.48 (0.18, 1.25)	0.134	0.76 (0.50, 1.17)	0.212	220	0.92 (0.75, 1.12)	0.418	0.99 (0.93, 1.07)	0.854	69	1.17 (0.80, 1.72)	0.420	1.00 (0.89, 1.14)	0.944	0.259
4	23	0.74 (0.30, 1.83)	0.521			191	0.89 (0.72, 1.09)	0.252			95	0.96 (0.66, 1.39)	0.831			
5	6	0.34 (0.10, 1.18)	0.090			195	0.88 (0.71, 1.08)	0.216			79	1.00 (0.69, 1.46)	0.999			
p for trend			0.191					0.233					0.877			

[a] Adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; NASIDs, non-steroidal anti-inflammatory drugs.

Table 5-1. Baseline characteristics of participants according to genotype of rs4988235, Multiethnic Cohort Study

Characteristics	rs4988235 Genotype		
	GG	GA	AA
No. of participants	45,967	16,709	5,750
Age at cohort entry in years			
Mean (SD)	67.6 (8.42)	67.2 (8.28)	66.5 (8.40)
Median (IQR)	67 (61, 74)	67 (61, 74)	65 (60, 73)
Sex, n (%)			
Men	21,068 (45.8%)	7,776 (46.5%)	2,753 (47.9%)
Ethnicity, n (%)			
White	2,835 (6.17%)	6,652 (39.8%)	4,390 (76.4%)
African American	7,052 (15.3%)	2,802 (16.8%)	328 (5.7%)
Native Hawaiian	3,713 (8.1%)	1,330 (8.0%)	201 (3.5%)
Japanese American	22,496 (48.9%)	314 (1.9%)	3 (0.1%)
Latino	9,871 (21.5%)	5,611 (33.6%)	828 (14.4%)
Family history of colorectal cancer, n (%)	4,129 (9.0%)	1,295 (7.8%)	463 (8.1%)
History of intestinal polyps, n (%)	2,957 (6.4%)	883 (5.3%)	357 (6.2%)
Education, n (%)			
≤12th grade	17,618 (38.3%)	6,115 (36.6%)	1,229 (21.4%)
Vocational/some college	13,993 (30.4%)	5,000 (29.9%)	1,715 (29.8%)
≥College graduate	13,945 (30.3%)	5,442 (32.6%)	2,760 (48.0%)
Missing	411 (0.9%)	152 (0.9%)	46 (0.8%)
Body mass index, n (%)			
Underweight/Normal (<25 kg/m^2)	19,744 (43.0%)	5,945 (35.6%)	2,485 (43.2%)
Overweight (25-30 kg/m^2)	17,895 (28.9%)	6,786 (40.6%)	2,206 (38.4%)
Obese (≥30 kg/m^2)	8,105 (17.6%)	3,897 (23.3%)	1,054 (18.3%)
Missing	223 (0.5%)	81 (0.5%)	5 (0.1%)
Diagnosis of diabetes, n (%)	4,671 (10.2%)	1,436 (8.6%)	285 (5.0%)
Physical activity (MET-hours/day), n (%)			
Q1 [0, 0.321)	7,440 (16.2%)	2,649 (15.8%)	609 (10.6%)

Characteristics	rs4988235 Genotype		
	GG	GA	AA
Q2 [0.321, 0.571)	9,130 (19.9%)	2,975 (17.8%)	876 (15.2%)
Q3 [0.571, 0.929)	9,730 (21.2%)	3,300 (19.8%)	1,134 (19.7%)
Q4 [0.929, 1.786)	9,915 (21.6%)	3,746 (22.4%)	1,545 (26.9%)
Q5 [1.786, 14.286)	9,118 (19.8%)	3,767 (22.5%)	1,548 (26.9%)
Missing	634 (1.4%)	272 (1.6%)	38 (0.7%)
Smoking status and pack-years, n (%)			
Never	21,999 (47.9%)	7,254 (43.4%)	2,443 (42.5%)
Former, <20 pack-years	13,272 (28.9%)	5,003 (29.9%)	1,789 (31.1%)
Former, ≥20 pack-years	3,658 (8.0%)	1,521 (9.1%)	642 (11.2%)
Current, <20 pack-years	3,293 (7.2%)	1,335 (8.0%)	317 (5.5%)
Current, ≥20 pack-years	2,316 (5.0%)	1,014 (6.1%)	425 (7.4%)
Missing	1,429 (3.1%)	582 (3.5%)	134 (2.3%)
Alcohol consumption, n (%)			
Never	24,790 (53.9%)	6,886 (41.2%)	1,861 (32.4%)
<15 g/day	15,036 (32.7%)	6,367 (38.1%)	2,285 (39.7%)
15-30 g/day	3,007 (6.5%)	1,571 (9.4%)	753 (13.1%)
≥30 g/day	3,134 (6.8%)	1,885 (11.3%)	851 (14.8%)
Multivitamin use, n (%)	22,870 (49.8%)	8,570 (51.3%)	3,060 (53.2%)
NSAID use, n (%)	20,865 (45.4%)	9,398 (56.2%)	3,139 (54.6%)
Total energy intake (kcal/day)			
Mean (SD)	2,180 (1027)	2,232 (1089)	2,152 (936)
Median (IQR)	1,968 (1473, 2643)	1,997 (1487, 2709)	1,972 (1513, 2580)
Red meat (g/1,000 kcal/day), n (%)			
Q1 [0, 7.87)	8,360 (18.2%)	3,437 (20.6%)	1,432 (24.9%)
Q2 [7.87, 13.60)	8,963 (19.5%)	3,310 (19.8%)	1,155 (20.1%)
Q3 [13.60, 19.40)	9,395 (20.4%)	3,229 (19.3%)	1,128 (19.6%)
Q4 [19.40, 27.20)	9,841 (21.4%)	3,316 (19.8%)	1,063 (18.5%)
Q5 [27.20, 190.00)	9,408 (20.5%)	3,417 (20.4%)	972 (16.9%)

Characteristics	rs4988235 Genotype		
	GG	GA	AA
Processed meat (g/1,000 kcal/day), n (%)			
Q1 [0, 2.45)	8,377 (18.2%)	3,744 (22.4%)	1,474 (25.6%)
Q2 [2.45, 4.85)	8,792 (19.1%)	3,635 (21.8%)	1,363 (23.7%)
Q3 [4.85, 7.61)	9,589 (20.9%)	3,447 (20.6%)	1,161 (20.2%)
Q4 [7.61, 11.70)	10,049 (21.9%)	3,002 (18.0%)	984 (17.1%)
Q5 [11.70, 95.90)	9,160 (19.9%)	2,881 (17.2%)	768 (13.4%)
Dietary fiber (g/1,000 kcal/day), n (%)			
Q1 [0.265, 8.160)	10,443 (22.7%)	2,681 (16.0%)	997 (17.3%)
Q2 [8.160, 10.200)	9,490 (20.6%)	3,194 (19.1%)	1,141 (19.8%)
Q3 [10.200, 12.300)	9,066 (19.7%)	3,556 (21.3%)	1,260 (21.9%)
Q4 [12.300, 15.100)	8,748 (19.0%)	3,641 (21.8%)	1,150 (20.0%)
Q5 [15.100, 40.400)	8,220 (17.9%)	3,637 (21.8%)	1,202 (20.9%)
Folate (μg DFE/day), n (%)			
Q1 [51.00, 439.00)	9,419 (20.5%)	2,811 (16.8%)	908 (15.8%)
Q2 [439.00, 671.00)	9,673 (21.0%)	3,264 (19.5%)	1,073 (18.7%)
Q3 [671.00, 944.00)	9,361 (20.4%)	3,362 (20.1%)	1,212 (21.1%)
Q4 [944.00, 1359.00)	9,038 (19.7%)	3,517 (21.0%)	1,273 (22.1%)
Q5 [1359, 6860.00)	8,476 (18.4%)	3,755 (22.5%)	1,284 (22.3%)

Abbreviations: SD, standard deviation; IQR, interquartile range; MET, metabolic equivalent of task; NSAIDs, non-steroidal anti-inflammatory drugs; DFE, dietary folate equivalents.

Table 5-2. Genotype distribution of rs4988235 according to race and ethnicity group

Population	No. of participants	Effect allele (A) frequency (%)	GG	GA	AA
Total	68,426	20.61%	45,967 (67.18%)	16,709 (24.42%)	5,750 (8.40%)
White	13,877	55.60%	2,835 (20.43%)	6,652 (47.94%)	4,390 (31.64%)
African American	10,182	16.98%	7,052 (69.26%)	2,802 (27.52%)	328 (3.22%)
Native Hawaiian	5,244	16.51%	3,713 (70.80%)	1,330 (25.36%)	201 (3.83%)
Japanese American	22,813	0.70%	22,496 (98.61%)	314 (1.38%)	3 (0.01%)
Latino	16,310	22.28%	9,871 (60.52%)	5,611 (34.40%)	828 (5.08%)

Table 5-3. Association between rs4988235 and the risk of CRC

Genetic model	Overall			White			African American			Latino			Test of heterogeneity
	RR[a]	95% CI	P value	RR	95% CI	P value	RR	95% CI	P value	RR	95% CI	P value	
Additive model	0.86	(0.77, 0.96)	0.007	0.925	(0.77, 1.11)	0.400	0.85	(0.68, 1.06)	0.148	0.80	(0.67, 0.97)	0.020	0.030
Dominant model AA vs. GG/GA (ref)	0.84	(0.68, 1.05)	0.128	0.818	(0.63, 1.07)	0.141	1.31	(0.76, 2.25)	0.328	0.69	(0.40, 1.18)	0.176	0.013
Recessive model AA/GA vs. GG (ref)	0.83	(0.71, 0.96)	0.011	1.07	(0.76, 1.49)	0.712	0.76	(0.58, 0.98)	0.037	0.79	(0.63, 0.98)	0.030	0.016

[a] Adjusted for sex as strata variable and PC1-10.

Table 5-4. Overall and race-specific association of dairy and milk intake on CRC risk by genotypes of rs4988235

Dietary intake		Dominant model						Recessive model				
		GG/GA			AA		Test of heterogeneity	GG		GA/AA		Test of heterogeneity
		RR	95% CI	P-value	RR[a]	P-value		RR	P-value	RR	P-value	
All dairy products (per SD increase)	Overall	1.00	(0.93, 1.08)	0.907	0.90	(0.75, 1.09)	0.282	0.389	1.01 (0.91, 1.13)	0.844	1.00 (0.91, 1.09)	0.918	0.063
	White	0.98	(0.95, 1.14)	0.799	0.87	(0.70, 1.09)	0.226	0.331	1.23 (0.91, 1.67)	0.184	0.91 (0.79, 1.04)	0.158	0.196
	African American	1.09	(0.96, 1.23)	0.167	NA[b]		NA	NA	1.10 (0.94, 1.28)	0.257	1.19 (0.99, 1.44)	0.070	0.064
	Latino	0.95	(0.84, 1.07)	0.402	0.57	(0.16, 2.00)	0.378	0.509	0.90 (0.76, 1.08)	0.255	1.01 (0.85, 1.21)	0.914	0.144
Milk (per SD increase)	Overall	1.02	(0.95, 1.09)	0.640	0.92	(0.77, 1.10)	0.379	0.381	1.02 (0.92, 1.14)	0.689	1.01 (0.92, 1.10)	0.822	0.057
	White	1.00	(0.87, 1.16)	0.948	0.84	(0.52, 1.36)	0.487	0.319	1.36 (1.05, 1.88)	0.020	0.92 (0.80, 1.04)	0.190	0.040
	African American	1.10	(0.98, 1.24)	0.107	NA		NA	NA	1.10 (0.94, 1.29)	0.249	1.20 (1.01, 1.43)	0.038	0.056
	Latino	0.95	(0.84, 1.07)	0.405	0.70	(0.23, 2.08)	0.520	0.489	0.89 (0.75, 1.06)	0.192	1.02 (0.86, 1.20)	0.841	0.115

[a] Adjusted for sex and 10-year age group at baseline as strata variables, PC1-PC10, family history of colorectal cancer, history of intestinal polyps, education, BMI, diabetes, smoking status, alcohol consumption, quintiles of physical activity, use of NSAIDs, regular use of multivitamins, log-transformed total energy intake, quintiles of red meat density intake, quintiles of processed meat density intake, quintiles of dietary fiber density intake, and quintiles of folate.

[b] Could not be estimated because of limited sample size in this group.

Abbreviations: CRC, colorectal cancer

Table 5-5. Significant associations (before FDR correction) between candidate SNP and risk of CRC among participants in the Multiethnic Cohort Study

Gene	Race and Ethnicity	SNP	Reference allele	Alternate allele	RR	95% CI	p-value	p-value (FDR)	Model
LCT	White	rs2322659	T	C	1.33	(1.04, 1.7)	0.022	0.088	Dominate (TT+TC vs. CC)
LCT	African American	rs3820790	A	T	1.39	(1.14, 1.69)	0.001	0.037	Additive (T allele)
LCT	African American	rs112742092	G	A	0.75	(0.60, 0.94)	0.014	0.407	Dominate (GG+GA vs. AA)
LCT	African American	rs3820790	A	T	0.52	(0.30, 0.91)	0.022	0.407	Dominate (AA+AT vs. TT)
LCT	African American	rs1807356	T	C	0.76	(0.59, 0.99)	0.044	0.543	Dominate (TT+TC vs. CC)
LCT	African American	rs892715	T	C	0.76	(0.60, 0.96)	0.023	0.426	Recessive (TT vs. CC+TC)
LCT	African American	rs3820790	A	T	1.41	(1.12, 1.77)	0.003	0.111	Recessive (AA vs. TT+AT)
LCT	Latino	rs72972158	G	A	1.38	(1.07, 1.77)	0.012	0.168	Additive (A allele)
LCT	Latino	rs72972158	G	A	1.380	(1.05, 1.81)	0.020	0.252	Recessive (GG vs. AA+GA)
LCT	Latino	rs575712683	C	G	1.503	(1.03, 2.20)	0.036	0.252	Recessive (CC vs. GG+CG)
MCM6	African American	rs72972196	G	A	1.40	(1.10, 1.80)	0.007	0.112	Additive (A allele)
MCM6	African American	rs309132	C	G	1.20	(1.03, 1.41)	0.023	0.184	Additive (G allele)
MCM6	African American	rs3213871	C	T	1.22	(1.01, 1.46)	0.036	0.192	Additive (T allele)
MCM6	African American	rs309132	C	G	0.64	(0.50, 0.83)	0.001	0.016	Dominate (CC+CG vs. GG)
MCM6	African American	rs3213871	C	T	0.63	(0.41, 0.99)	0.044	0.352	Dominate (CC+CT vs. TT)
MCM6	African American	rs72972196	G	A	1.44	(1.10, 1.88)	0.007	0.112	Recessive (GG vs. AA+GA)
MCM6	African American	rs680428	A	G	0.76	(0.60, 0.96)	0.022	0.176	Recessive (AA vs. GG+AG)
MCM6	African American	rs4988235	G	A	0.76	(0.58, 0.98)	0.037	0.197	Recessive (GG vs. AA+GA)
MCM6	Latino	rs4988235	G	A	0.80	(0.66, 0.97)	0.020	0.200	Additive (A allele)
MCM6	Latino	rs4988235	G	A	0.785	(0.63, 0.98)	0.030	0.238	Recessive (GG vs. AA+GA)
CASR	White	rs3792290	A	C	0.77	(0.59, 1.00)	0.049	0.886	Additive (C allele)
CASR	White	rs11711722	C	A	1.36	(1.05, 1.78)	0.022	0.783	Recessive (CC vs. AA+CA)
CASR	White	rs3792290	A	C	0.74	(0.55, 0.99)	0.045	0.783	Recessive (AA vs. CC+AC)
CASR	African American	rs7617898	C	G	0.81	(0.69, 0.95)	0.011	0.444	Additive (G allele)
CASR	African American	rs115334608	A	T	0.82	(0.70, 0.96)	0.015	0.444	Additive (T allele)
CASR	African American	rs2173961	T	G	0.84	(0.72, 0.98)	0.027	0.444	Additive (G allele)
CASR	African American	rs3845918	C	T	0.74	(0.56, 0.98)	0.034	0.444	Additive (T allele)
CASR	African American	rs13089000	G	C	0.81	(0.67, 0.99)	0.037	0.444	Additive (C allele)
CASR	African American	rs34346613	C	T	2.37	(1.17, 4.78)	0.016	0.300	Dominate (CC+CT vs. TT)

Gene	Race and Ethnicity	SNP	Reference allele	Alternate allele	RR	95% CI	p-value	p-value (FDR)	Model
CASR	African American	rs115334608	A	T	1.34	(1.05, 1.70)	0.017	0.300	Dominate (AA+AT vs. TT)
CASR	African American	rs148668166	G	A	2.18	(1.13, 4.24)	0.021	0.300	Dominate (GG+GA vs. AA)
CASR	African American	rs7617898	C	G	1.30	(1.04, 1.63)	0.022	0.300	Dominate (CC+CG vs. GG)
CASR	African American	rs3749205	T	C	1.44	(1.05, 1.99)	0.025	0.300	Dominate (TT+TC vs. CC)
CASR	African American	rs3749204	A	T	1.43	(1.03, 1.98)	0.030	0.300	Dominate (AA+AT vs. TT)
CASR	African American	rs2173961	T	G	1.28	(1.01, 1.62)	0.039	0.334	Dominate (TT+TG vs. GG)
CASR	African American	rs1604447	C	T	0.61	(0.44, 0.83)	0.002	0.120	Recessive (CC vs. TT+CT)
CASR	African American	rs13089000	G	C	0.53	(0.33, 0.87)	0.011	0.330	Recessive (GG vs. CC+GC)
CASR	African American	rs3845918	C	T	0.70	(0.52, 0.95)	0.022	0.440	Recessive (CC vs. TT+CT)
CASR	Japanese American	rs62271399	C	T	0.75	(0.62, 0.91)	0.004	0.068	Additive (T allele)
CASR	Japanese American	rs79920788	G	A	0.74	(0.58, 0.93)	0.011	0.094	Additive (A allele)
CASR	Japanese American	rs115319101	C	T	1.45	(1.04, 2.03)	0.027	0.153	Additive (T allele)
CASR	Japanese American	rs115319101	C	T	0.14	(0.05, 0.45)	0.001	0.017	Dominate (CC+CT vs. TT)
CASR	Japanese American	rs62271399	C	T	0.74	(0.60, 0.92)	0.006	0.102	Recessive (CC vs. TT+CT)
CASR	Japanese American	rs79920788	G	A	0.74	(0.58, 0.95)	0.016	0.136	Recessive (GG vs. AA+GA)
VDR	White	rs11574052	C	T	0.38	(0.20, 0.71)	0.002	0.142	Additive (T allele)
VDR	White	rs4760650	G	T	0.83	(0.70, 0.98)	0.030	0.961	Additive (T allele)
VDR	White	rs4760650	G	T	1.41	(1.05, 1.88)	0.021	0.753	Dominate (GG+GT vs. TT)
VDR	White	rs12721414	T	A	0.47	(0.24, 0.92)	0.029	0.753	Dominate (TT+TA vs. AA)
VDR	White	rs2238136	C	T	1.84	(1.01, 3.37)	0.047	0.753	Dominate (CC+CT vs. TT)
VDR	White	rs11574052	C	T	0.37	(0.20, 0.70)	0.002	0.142	Recessive (CC vs. TT+CT)
VDR	African American	rs10783218	G	A	1.38	(1.14, 1.68)	0.001	0.166	Additive (A allele)
VDR	African American	rs12302580	G	C	1.24	(1.03, 1.50)	0.022	0.909	Additive (C allele)
VDR	African American	rs11168275	T	C	1.20	(1.01, 1.42)	0.041	0.909	Additive (C allele)
VDR	African American	rs11574042	C	G	0.41	(0.23, 0.72)	0.002	0.332	Dominate (CC+CG vs. GG)
VDR	African American	rs111336890	T	C	0.38	(0.19, 0.77)	0.007	0.581	Dominate (TT+TC vs. CC)
VDR	African American	rs11168275	T	C	0.66	(0.46, 0.93)	0.019	0.697	Dominate (TT+TC vs. CC)
VDR	African American	rs11168266	C	T	1.34	(1.05, 1.73)	0.021	0.697	Dominate (CC+CT vs. TT)
VDR	African American	rs7965360	A	G	1.31	(1.04, 1.64)	0.021	0.697	Dominate (AA+AG vs. GG)
VDR	African American	rs78831519	A	T	0.23	(0.06, 0.94)	0.040	0.916	Dominate (AA+AT vs. TT)
VDR	African American	rs10783218	G	A	1.45	(1.15, 1.82)	0.002	0.332	Recessive (GG vs. AA+GA)

Gene	Race and Ethnicity	SNP	Reference allele	Alternate allele	RR	95% CI	p-value	p-value (FDR)	Model
VDR	African American	rs2283344	T	C	0.70	(0.54, 0.91)	0.007	0.581	Recessive (TT vs. CC+TC)
VDR	African American	rs4760650	G	T	0.73	(0.57, 0.93)	0.011	0.609	Recessive (GG vs. TT+GT)
VDR	African American	rs12302580	G	C	1.27	(1.01, 1.59)	0.039	0.941	Recessive (GG vs. CC+GC)
VDR	African American	rs11168268	G	A	0.75	(0.57, 0.99)	0.042	0.941	Recessive (GG vs. AA+GA)
VDR	Japanese American	rs2254210	G	A	1.20	(1.06, 1.36)	0.005	0.252	Additive (A allele)
VDR	Japanese American	rs4760650	G	T	1.19	(1.05, 1.36)	0.008	0.252	Additive (T allele)
VDR	Japanese American	rs143244503	G	A	0.65	(0.45, 0.92)	0.016	0.336	Additive (A allele)
VDR	Japanese American	rs7965266	T	A	0.88	(0.78, 0.99)	0.039	0.387	Additive (A allele)
VDR	Japanese American	rs2283342	A	G	0.89	(0.79, 1.00)	0.042	0.387	Additive (G allele)
VDR	Japanese American	rs2853566	G	A	0.88	(0.77, 1.00)	0.042	0.387	Additive (A allele)
VDR	Japanese American	rs142418811	T	C	0.75	(0.56, 0.99)	0.043	0.387	Additive (C allele)
VDR	Japanese American	rs2238137	C	T	0.49	(0.28, 0.87)	0.015	0.410	Dominate (CC+CT vs. TT)
VDR	Japanese American	rs4583039	G	A	0.69	(0.51, 0.93)	0.017	0.410	Dominate (GG+GA vs. AA)
VDR	Japanese American	rs7965266	T	A	1.21	(1.03, 1.43)	0.021	0.410	Dominate (TT+TA vs. AA)
VDR	Japanese American	rs2283342	A	G	1.25	(1.03, 1.52)	0.026	0.410	Dominate (AA+AG vs. GG)
VDR	Japanese American	rs4303288	A	C	1.21	(1.01, 1.44)	0.037	0.410	Dominate (AA+AC vs. CC)
VDR	Japanese American	rs2853566	G	A	1.19	(1.01, 1.39)	0.039	0.410	Dominate (GG+GA vs. AA)
VDR	Japanese American	rs4760650	G	T	1.27	(1.08, 1.49)	0.005	0.189	Recessive (GG vs. TT+GT)
VDR	Japanese American	rs2254210	G	A	1.25	(1.07, 1.47)	0.006	0.189	Recessive (GG vs. AA+GA)
VDR	Japanese American	rs143244503	G	A	0.63	(0.43, 0.91)	0.013	0.205	Recessive (GG vs. AA+GA)
VDR	Japanese American	rs2189480	G	T	0.75	(0.60, 0.94)	0.013	0.205	Recessive (GG vs. TT+GT)
VDR	Japanese American	rs10783218	G	A	0.71	(0.52, 0.99)	0.043	0.441	Recessive (GG vs. AA+GA)
VDR	Japanese American	rs1540339	C	T	0.76	(0.58, 0.99)	0.044	0.441	Recessive (CC vs. TT+CT)
VDR	Japanese American	rs142418811	T	C	0.74	(0.55, 1.00)	0.049	0.441	Recessive (TT vs. CC+TC)
VDR	Latino	rs11168275	T	C	0.68	(0.49, 0.94)	0.021	0.991	Dominate (TT+TC vs. CC)

Table 6-1. Summary of GWAS datasets used for LPH levels and CRC

LPH GWAS					CRC GWAS					Colon Cancer GWAS		Rectal Cancer GWAS	
Study	**First Author (Year)**	**Sample Size**	**Population**	**Sex**	**Study**	**N Cases**	**N Controls**	**Population**	**Sex**	**CRC Ascertainment**	**N Cases**	**N Controls**	**N Cases**	**N Controls**
Fenland Study	Pietzner (2021) [44]	10,708	100% European	53% Female	FinnGen Study	6509	287,137	100% European	42% Female	ICD10: C18-C20	3935	287,137	2361	287,137
					PLCO Atlas	2065	67,500	100% European	45% Female	ICD-O-2 Site: 180/182-189/199/209/212/218	1611	65,142	320	65,142
					Pan-UK Biobank	592	419,881	100% European	44% Female	Self-reported diagnosis of large bowel cancer / colorectal caner	1384	419,089	301	420,172

Abbreviations: GWAS: genome-wide association studies; LPH: lactase-phlorizin hydrolase; CRC: colorectal cancer.

Table 6-2. Characteristics of genetic instruments for elevated LPH levels from the GWAS identified in the GWAS Catalog

RSID	Position (GRCh38)	Effect Allele	Other Allele	EAF	R^2	F	Beta	SE	*p*-Value	Associated Gene	*cis*/*trans* Variant for LPH Levels
rs4988235	chr2:135851076	A	G	0.31	33.28%	5340.06	0.882	0.011	3×10^{-1451}	*MCM6*	*cis*
rs516246	chr19:48702915	C	T	0.49	0.81%	87.01	0.127	0.013	2×10^{-22}	*FUT2*	*trans*
rs532436	chr9:133274414	G	A	0.20	1.27%	137.41	0.199	0.016	3×10^{-35}	*ABO*	*trans*
rs641476	chr18:32225445	C	T	0.61	1.07%	115.85	0.150	0.013	5×10^{-30}	*GAREM1*	*trans*

Abbreviations: LPH, lactase-phlorizin hydrolase; GWAS: genome-wide association studies; EAF: effect allele frequency; SE: standard error.

Table 6-3. Summary of four genetic instruments and their proxies (where necessary) from the FinnGen, PLCO, and Pan-UK Biobank GWAS on CRC

SNP Selected	Effect Allele	Beta	SE	*p*-Value
FinnGen				
rs4988235	A	−0.081	0.020	0.001
rs516246	C	−0.024	0.020	0.291
rs635634 (proxy) [a]	G	0.006	0.025	0.386
rs641476	C	0.004	0.020	0.967
PLCO				
rs4988235	A	−0.072	0.039	0.063
rs516246	C	−0.034	0.032	0.292
rs532436	G	0.047	0.040	0.237
rs641476	C	0.020	0.033	0.537
Pan-UK Biobank				
rs4988235	A	−0.002	0.061	0.971
rs516246	C	−0.072	0.051	0.162
rs532436	G	0.144	0.066	0.030
rs641476	C	0.061	0.053	0.245

[a] rs635634 at chr9:133279427 (effect allele G) was used as a proxy for rs532436 ($r^2 = 0.99$) in the FinnGen study. Abbreviations: CRC: colorectal cancer; GWAS: Genome-wide association studies; SNP: single nucleotide polymorphism; SE: standard error.

Table 6-4. Associations of genetically predicted elevated LPH Levels and CRC in the FinnGen, PLCO, and Pan-UK Biobank datasets

Method	OR	Lower 95% CI	Upper 95% CI	P-value	MR-Egger Intercept	Heterogeneity: Q, P
FinnGen						
***cis*-MR**						
Wald ratio	0.91	0.88	0.95	1.3×10^{-5}		
MR using all genetic variants						
IVW	0.92	0.88	0.95	1.8×10^{-5}		
Penalized IVW	0.92	0.87	0.97	0.001		
Robust IVW	0.92	0.90	0.93	8.9×10^{-37}		
Penalized robust IVW	0.92	0.90	0.94	2.9×10^{-15}		
MR-Egger	0.90	0.85	0.96	0.001	0.009, 0.552	2.5, 0.482
Weighted median	0.92	0.88	0.95	3.3×10^{-5}		
Mode-based estimation	0.91	0.88	0.95	1.6×10^{-5}		
MR Lasso	0.92	0.88	0.95	1.8×10^{-5}		
PLCO						
***cis*-MR**						
Wald ratio	0.92	0.85	1.00	0.0631		
MR using all genetic variants						
IVW	0.94	0.85	1.03	0.170		
Penalized IVW	0.94	0.85	1.03	0.170		
Robust IVW	0.94	0.90	0.98	0.002		
Penalized robust IVW	0.94	0.90	0.98	0.002		
MR-Egger	0.91	0.78	1.05	0.193	0.020, 0.532	3.9, 0.273
Weighted median	0.93	0.86	1.02	0.109		
Mode-based estimation	0.92	0.85	1.01	0.070		
MR Lasso	0.94	0.85	1.03	0.170		
Pan-UK Biobank						
***cis*-MR**						
Wald ratio	1.00	0.87	1.14	0.9707		
MR using all genetic variants						
IVW	1.03	0.83	1.27	0.812		
Penalized IVW	1.07	0.77	1.49	0.681		
Robust IVW	1.02	0.94	1.11	0.569		
Penalized robust IVW	1.07	0.88	1.30	0.510		
MR-Egger	0.98	0.69	1.40	0.901	0.029, 0.712	7.9, 0.049
Weighted median	1.02	0.89	1.17	0.787		
Mode-based estimation	1.01	0.88	1.16	0.925		
MR-Lasso	1.03	0.83	1.27	0.812		

LPH, lactase-phlorizin hydrolase; CRC, colorectal cancer; OR: odds ratio; CI: confidence interval; MR: Mendelian Randomization; IVW: inverse variance weighted.

Table 6-5. Associations of genetically predicted elevated LPH levels and colon cancer in the FinnGen, PLCO, and Pan-UK Biobank datasets

Method	OR	Lower 95% CI	Upper 95% CI	P-value	MR-Egger Intercept	Heterogeneity: Q, P
FinnGen						
***cis*-MR**						
Wald ratio	0.92	0.87	0.97	0.0011		
MR using all genetic variants						
IVW	0.92	0.87	0.97	0.002		
Penalized IVW	0.92	0.87	0.97	0.002		
Robust IVW	0.92	0.91	0.93	5×10^{-31}		
Penalized robust IVW	0.92	0.91	0.93	5×10^{-31}		
MR-Egger	0.92	0.83	1.01	0.074	0.002, 0.641	3.3, 0.345
Weighted median	0.93	0.89	0.97	0.002		
Mode-based estimation	0.92	0.87	0.97	0.001		
MR Lasso	0.92	0.87	0.97	0.002		
PLCO						
***cis*-MR**						
Wald ratio	0.93	0.85	1.02	0.1435		
MR using all genetic variants						
IVW	0.95	0.85	1.06	0.361		
Penalized IVW	0.98	0.82	1.16	0.814		
Robust IVW	0.95	0.90	1.00	0.038		
Penalized robust IVW	0.98	0.87	1.10	0.708		
MR-Egger	0.91	0.76	1.09	0.311	0.024, 0.534	4.5, 0.211
Weighted median	0.95	0.86	1.04	0.253		
Mode-based estimation	0.94	0.85	1.03	0.170		
MR Lasso	0.95	0.85	1.06	0.361		
Pan-UK Biobank						
***cis*-MR**						
Wald ratio	0.95	0.86	1.05	0.2851		
MR using all genetic variants						
IVW	0.95	0.84	1.07	0.401		
Penalized IVW	0.95	0.84	1.07	0.401		
Robust IVW	0.95	0.91	0.98	0.006		
Penalized robust IVW	0.95	0.91	0.98	0.006		
MR-Egger	0.94	0.76	1.16	0.564	0.006, 0.099	4.8, 0.184
Weighted median	0.94	0.85	1.04	0.234		
Mode-based estimation	0.94	0.85	1.05	0.233		
MR Lasso	0.95	0.84	1.07	0.401		

LPH, lactase-phlorizin hydrolase; OR: odds ratio; CI: confidence interval; MR: Mendelian Randomization; IVW: inverse variance weighted.

Table 6-6. Associations of genetically predicted elevated LPH levels and rectal cancer in the FinnGen, PLCO, and Pan-UK Biobank datasets

Method	OR	Lower 95% CI	Upper 95% CI	P-value	MR-Egger Intercept	Heterogeneity: Q, P
FinnGen						
***cis*-MR**						
Wald ratio	0.91	0.85	0.97	0.0051		
MR using all genetic variants						
IVW	0.92	0.86	0.98	0.009		
Penalized IVW	0.92	0.85	0.99	0.035		
Robust IVW	0.92	0.89	0.94	1×10^{-11}		
Penalized robust IVW	0.92	0.89	0.95	3×10^{-6}		
MR-Egger	0.89	0.81	0.97	0.013	0.023, 0.328	1.6, 0.654
Weighted median	0.91	0.85	0.97	0.006		
Mode-based estimation	0.91	0.85	0.97	0.005		
MR Lasso	0.92	0.86	0.98	0.009		
PLCO						
***cis*-MR**						
Wald ratio	0.86	0.70	1.06	0.1720		
MR using all genetic variants						
IVW	0.86	0.70	1.05	0.86		
Penalized IVW	0.86	0.70	1.05	0.86		
Robust IVW	0.86	0.82	0.90	0.86		
Penalized robust IVW	0.86	0.82	0.90	0.86		
MR-Egger	0.88	0.66	1.18	0.88	-0.017, 0.780	4.5, 0.211
Weighted median	0.86	0.70	1.06	0.151		
Mode-based estimation	0.86	0.70	1.05	0.135		
MR Lasso	0.86	0.69	1.07	0.173		
Pan-UK Biobank						
***cis*-MR**						
Wald ratio	1.13	0.91	1.40	0.2670		
MR using all genetic variants						
IVW	1.10	0.90	1.35	0.360		
Penalized IVW	1.10	0.90	1.35	0.360		
Robust IVW	1.10	1.03	1.18	0.006		
Penalized robust IVW	1.10	1.03	1.18	0.006		
MR-Egger	1.22	0.91	1.64	0.178	-0.064, 0.323	2.1, 0.559
Weighted median	0.91	0.85	0.97	0.315		
Mode-based estimation	1.13	0.91	1.41	0.270		
MR Lasso	1.10	0.90	1.35	0.360		

LPH, lactase-phlorizin hydrolase; OR: odds ratio; CI: confidence interval; MR: Mendelian Randomization; IVW: inverse variance weighted.

Table 6-7. Meta-analysis[a] results for the association between elevated LPH levels and CRC, colon cancer, and rectal cancer

	OR	Lower 95% CI	Upper 95% CI	P-value	I^2	Q-statistics, P
CRC						
***cis*-MR**						
Fixed effects model	0.92	0.89	0.95	4.66×10^{-6}	0%	1.51, 0.4695
Random effects model	0.92	0.89	0.95	4.66×10^{-6}		
MR using all variants						
Fixed effects model	0.92	0.89	0.96	1.20×10^{-5}	0%	1.18, 0.5536
Random effects model	0.92	0.89	0.96	1.20×10^{-5}		
Colon cancer						
***cis*-MR**						
Fixed effects model	0.92	0.89	0.96	0.0002	0%	0.34, 0.8455
Random effects model	0.92	0.89	0.96	0.0002		
MR using all variants						
Fixed effects model	0.93	0.89	0.97	0.0010	0%	0.43, 0.8054
Random effects model	0.93	0.89	0.97	0.0010		
Rectal cancer						
***cis*-MR**						
Fixed effects model	0.92	0.87	0.98	0.0083	49.8%	3.99, 0.1363
Random effects model	0.94	0.83	1.07	0.3601		
MR using all variants						
Fixed effects model	0.93	0.87	0.98	0.0108	40.6%	3.36, 0.1859
Random effects model	0.94	0.85	1.03	0.1590		

[a.] The meta-analysis was performed to combine the effect estimates across all three cohorts.
LPH, lactase-phlorizin hydrolase; CRC: colorectal cancer; OR: odds ratio; CI: confidence interval; MR: Mendelian Randomization.

Table 7-1. Summary of GWAS datasets for plasma proteins and CRC

Study	Phenotype	Sample Size	Population	Adjustment
deCODE Health Study	Plasma protein	35,559	100% European	Age, sex and sample age.
Fenland Study	Plasma protein	10,708	100% European	Age, sex, the first ten genetic principal components and test site.
UKB-PPP	Plasma protein	54,219	100% European	Age, age^2, sex, age × sex, age^2 × sex, batch, UKB center, UKB genetic array, time between blood sampling and measurement and the first 20 genetic principal components.
FinnGen	CRC	412,181	100% European	Age, gender, batch, UKB center, UKB genetic array, time between blood sampling and measurement and the first 20 genetic principal components.

Abbreviations: GWAS: Genome-wide association studies; CRC, colorectal cancer; UKBB-PPP, The UK Biobank Pharma Proteomics Project.

Table 7-2. MR analysis results for the association between plasma proteins and CRC using *cis* and *cis+trans* pQTLs

pQTLs dataset	Uniprot	Protein name	No of SNP included	MR method	OR	95% CI	p-value	p-value (FDR)
			***cis* only**					
deCODE	O60565	Gremlin-1 (GREM1)	1	Wald ratio	1.16	(1.10, 1.21)	1.64E-09	2.21E-06
	P09848	Lactase/phlorizin hydrolase (LPH)	1	Wald ratio	0.93	(0.90, 0.96)	5.29E-05	3.56E-02
Fenland	O60565	Gremlin-1 (GREM1)	1	Wald ratio	1.12	(1.08, 1.16)	5.47E-11	7.46E-08
	O76074	cGMP-specific 3',5'-cyclic phosphodiesterase (PDE5A)	1	Wald ratio	1.25	(1.13, 1.39)	2.50E-05	1.70E-02
	P09848	Lactase/phlorizin hydrolase (LPH)	1	Wald ratio	0.92	(0.88, 0.96)	5.29E-05	2.40E-02
	Q9UHB6	LIM domain and actin-binding protein 1 (LIMA1)	1	Wald ratio	1.49	(1.22, 1.82)	7.63E-05	2.60E-02
UKBB	O76074	cGMP-specific 3',5'-cyclic phosphodiesterase (PDE5A)	1	Wald ratio	1.63	(1.30, 2.04)	2.50E-05	4.42E-02
			cis* + *trans					
deCODE	O60565	Gremlin-1 (GREM1)	1	Wald ratio	1.16	(1.10, 1.21)	1.64E-09	3.41E-06
	P09848	Lactase/phlorizin hydrolase (LPH)	2	Inverse variance weighted	0.93	(0.90, 0.96)	4.98E-05	1.48E-02
	O75347	Tubulin-specific chaperone A (TBCA)	1	Wald ratio	2.49	(1.60, 3.86)	4.83E-05	1.48E-02
	Q00169	Phosphatidylinositol transfer protein alpha isoform (PIPNA)	1	Wald ratio	4.31	(2.13, 8.71)	4.63E-05	1.48E-02
	O43175	D-3-phosphoglycerate dehydrogenase (SERA)	1	Wald ratio	4.41	(2.18, 8.92)	3.55E-05	1.48E-02
	P36404	ADP-ribosylation factor-like protein 2 (ARL2)	1	Wald ratio	3.37	(1.88, 6.06)	4.83E-05	1.48E-02
	Q99519	Sialidase-1 (NEUR1)	1	Wald ratio	0.27	(0.15, 0.49)	1.46E-05	1.48E-02
	Q8NFZ4	Neuroligin-2 (NLGN2)	1	Wald ratio	2.54	(1.59, 4.07)	1.02E-04	2.64E-02
Fenland	O60565	Gremlin-1 (GREM1)	4	Inverse variance weighted	1.12	(1.08, 1.16)	5.52E-11	1.14E-07
	P10082	Peptide YY (PYY)	1	Wald ratio	2.53	(1.90, 3.37)	2.62E-10	2.70E-07
	Q9H772	Gremlin-2 (GREM2)	1	Wald ratio	0.36	(0.26, 0.50)	1.64E-09	1.13E-06
	O00560	Syntenin-1 (SDCB1)	1	Wald ratio	2.12	(1.56, 2.87)	1.37E-06	7.06E-04
	Q15004	PCNA-associated factor (PAF15)	1	Wald ratio	1.65	(1.30, 2.08)	2.99E-05	1.03E-02
	O76074	cGMP-specific 3',5'-cyclic phosphodiesterase (PDE5A)	1	Wald ratio	1.25	(1.13, 1.39)	2.50E-05	1.03E-02
	P09848	Lactase/phlorizin hydrolase (LPH)	3	Inverse variance weighted	0.92	(0.89, 0.96)	4.35E-05	1.28E-02
	Q9UHB6	LIM domain and actin-binding protein 1 (LIMA1)	1	Wald ratio	1.49	(1.22, 1.82)	7.63E-05	1.96E-02
UKBB	Q92599	Septin-8 (SEPT8)	3	Inverse variance weighted	1.77	(1.42, 2.20)	4.10E-07	8.61E-04
	O95407	Tumor necrosis factor receptor superfamily member 6B (TNF6B)	11	Inverse variance weighted	1.21	(1.11, 1.33)	2.94E-05	2.06E-02

O76074	cGMP-specific 3',5'-cyclic phosphodiesterase (PDE5A)	1	Wald ratio	1.63	(1.30, 2.04)	2.50E-05	2.06E-02
P51671	Eotaxin (CCL11)	9	Inverse variance weighted	1.39	(1.18, 1.63)	6.78E-05	2.85E-02
Q9Y4K4	Mitogen-activated protein kinase kinase kinase kinase 5 (M4K5)	1	Wald ratio	0.57	(0.44, 0.75)	5.66E-05	2.85E-02
Q6WN34	Chordin-like protein 2 (CRDL2)	7	Inverse variance weighted	1.22	(1.10, 1.35)	1.16E-04	4.08E-02

Abbreviations: MR, Mendelian Randomization; CRC, colorectal cancer; pQTL, ; protein quantitative trait loci; OR, odds ratio; FDR, false discovery rate; UKBB-PPP, The UK Biobank Pharma Proteomics Project.

Table 7-3. Colocalization analysis of *cis*-pQTLs and CRC

pQTL dataset	Uniprot	Protein name	PH4 for colocalization analysis	Evidence of colocalization? (PH4>0.8)
deCODE	O60565	Gremlin-1 (GREM1)	0.0098	NO
deCODE	P09848	Lactase/phlorizin hydrolase (LPH)	0.9185	YES
UKB-PPP	O76074	cGMP-specific 3',5'-cyclic phosphodiesterase (PDE5A)	0.8531	YES

[a] UniProt is a high-quality and comprehensive database of protein sequence and functional information (https://www.uniprot.org/).
Abbreviations: pQTL, protein quantitative trait loci; CRC, colorectal cancer; UKBB-PPP, The UK Biobank Pharma Proteomics Project.

Table 7-4. Steiger filtering analysis of causal directionality test between plasma proteins and CRC

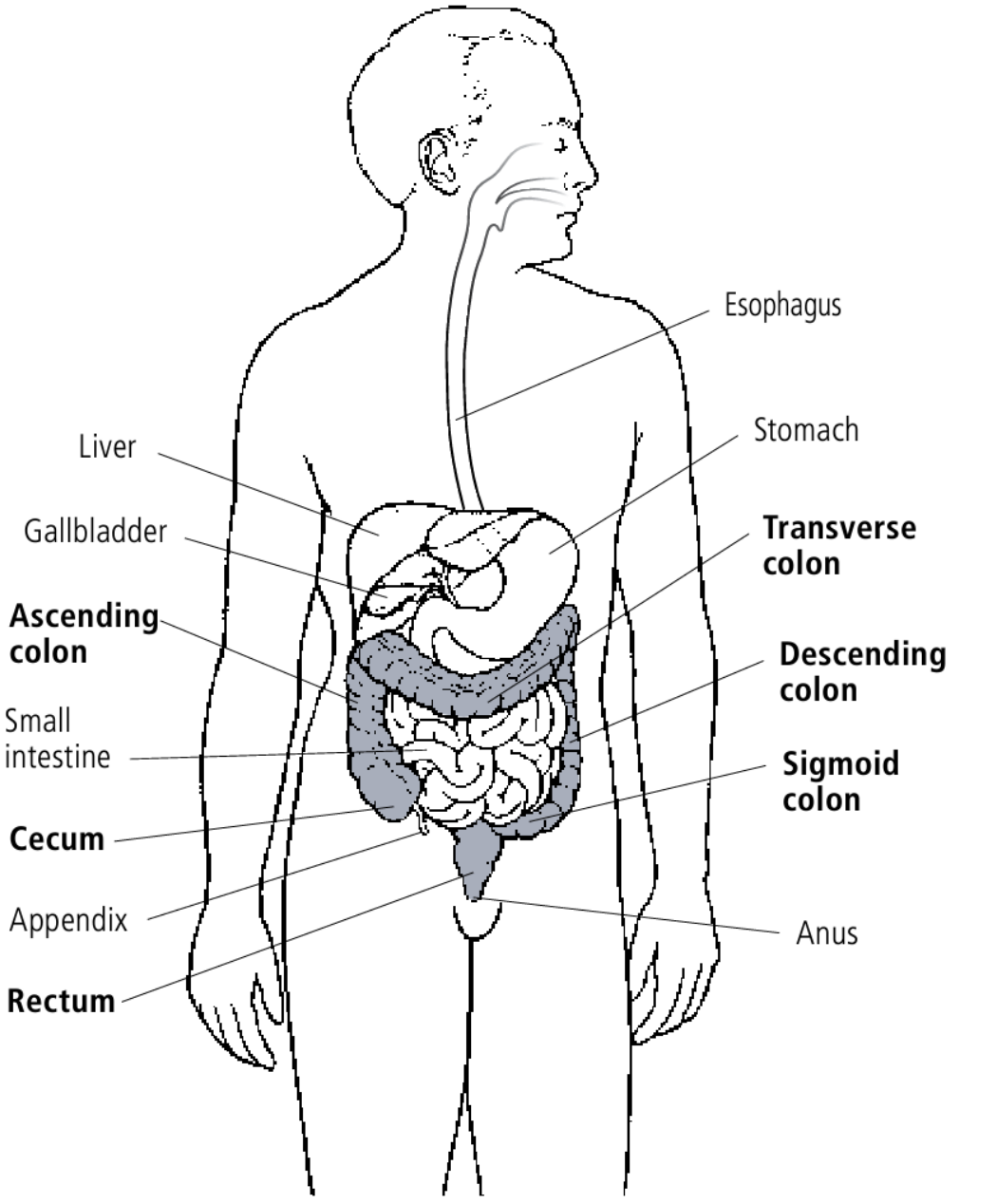
pQTLs dataset	Uniprot	Protein name	r^2: pQTL and exposure (r^2.exposure)	r^2: pQTL and outcome (r^2.outcome)	r^2.exposure > r^2.outcome	P value H0: r^2.exposure = r^2.outcome
			***cis* only**			
deCODE	O60565	Gremlin-1 (GREM1)	0.0254	0.0008	Yes	5.29E-123
	P09848	Lactase/phlorizin hydrolase (LPH)	0.0254	0.0004	Yes	7.00E-140
Fenland	O60565	Gremlin-1 (GREM1)	0.0818	0.0010	Yes	2.24E-157
	O76074	cGMP-specific 3',5'-cyclic phosphodiesterase (PDE5A)	0.0667	0.0004	Yes	1.45E-135
	P09848	Lactase/phlorizin hydrolase (LPH)	0.0818	0.0004	Yes	4.88E-171
	Q9UHB6	LIM domain and actin-binding protein 1 (LIMA1)	0.0165	0.0004	Yes	6.46E-29
UKB-PPP	O76074	cGMP-specific 3',5'-cyclic phosphodiesterase (PDE5A)	0.0075	0.0004	Yes	4.03E-46
			cis* + *trans			
deCODE	O60565	Gremlin-1 (GREM1)	0.0254	0.0008	Yes	5.29E-123
	P09848	Lactase/phlorizin hydrolase (LPH)	0.0302	0.0004	Yes	2.18E-170
	O75347	Tubulin-specific chaperone A (TBCA)	0.0021	0.0004	Yes	2.62E-06
	Q00169	Phosphatidylinositol transfer protein alpha isoform (PIPNA)	0.0011	0.0004	Yes	1.38E-02
	O43175	D-3-phosphoglycerate dehydrogenase (SERA)	0.0011	0.0004	Yes	2.60E-02
	P36404	ADP-ribosylation factor-like protein 2 (ARL2)	0.0012	0.0004	Yes	6.14E-03
	Q99519	Sialidase-1 (NEUR1)	0.0013	0.0004	Yes	5.53E-03
	Q8NFZ4	Neuroligin-2 (NLGN2)	0.0030	0.0004	Yes	2.26E-10
Fenland	O60565	Gremlin-1 (GREM1)	0.1064	0.0010	Yes	9.10E-214
	P10082	Peptide YY (PYY)	0.0090	0.0009	Yes	3.83E-11
	Q9H772	Gremlin-2 (GREM2)	0.0099	0.0008	Yes	5.95E-13
	O00560	Syntenin-1 (SDCB1)	0.0067	0.0005	Yes	1.78E-09
	Q15004	PCNA-associated factor (PAF15)	0.0108	0.0004	Yes	1.22E-17
	O76074	cGMP-specific 3',5'-cyclic phosphodiesterase (PDE5A)	0.0667	0.0004	Yes	5.81E-136
	P09848	Lactase/phlorizin hydrolase (LPH)	0.1027	0.0004	Yes	4.20E-220
	Q9UHB6	LIM domain and actin-binding protein 1 (LIMA1)	0.0165	0.0004	Yes	5.36E-29

pQTLs dataset	Uniprot	Protein name	r^2: pQTL and exposure (r^2.exposure)	r^2: pQTL and outcome (r^2.outcome)	r^2.exposure > r^2.outcome	P value H0: r^2.exposure = r^2.outcome
UKB-PPP	Q92599	Septin-8 (SEPT8)	0.0059	0.0006	Yes	1.57E-28
	O95407	Tumor necrosis factor receptor superfamily member 6B (TNF6B)	0.0348	0.0005	Yes	2.49E-275
	O76074	Gremlin-1 (GREM1)	0.0075	0.0004	Yes	4.03E-46
	P51671	Lactase/phlorizin hydrolase (LPH)	0.0175	0.0005	Yes	1.09E-123
	Q9Y4K4	Tubulin-specific chaperone A (TBCA)	0.0099	0.0003	Yes	3.83E-67
	Q6WN34	Phosphatidylinositol transfer protein alpha isoform (PIPNA)	0.0346	0.0005	Yes	9.58E-279

[a] UniProt is a high-quality and comprehensive database of protein sequence and functional information (https://www.uniprot.org/).
Abbreviations: CRC, colorectal cancer; pQTL, protein quantitative trait loci; UKBB-PPP, The UK Biobank Pharma Proteomics Project.

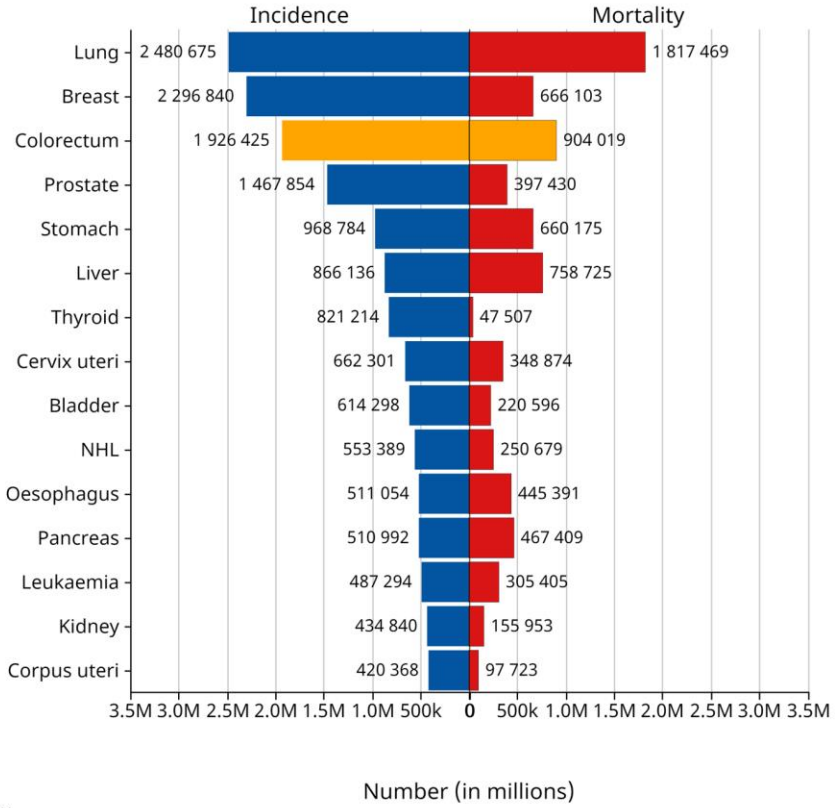
FIGURES

Figure 1-1. Anatomy of the Gastrointestinal System



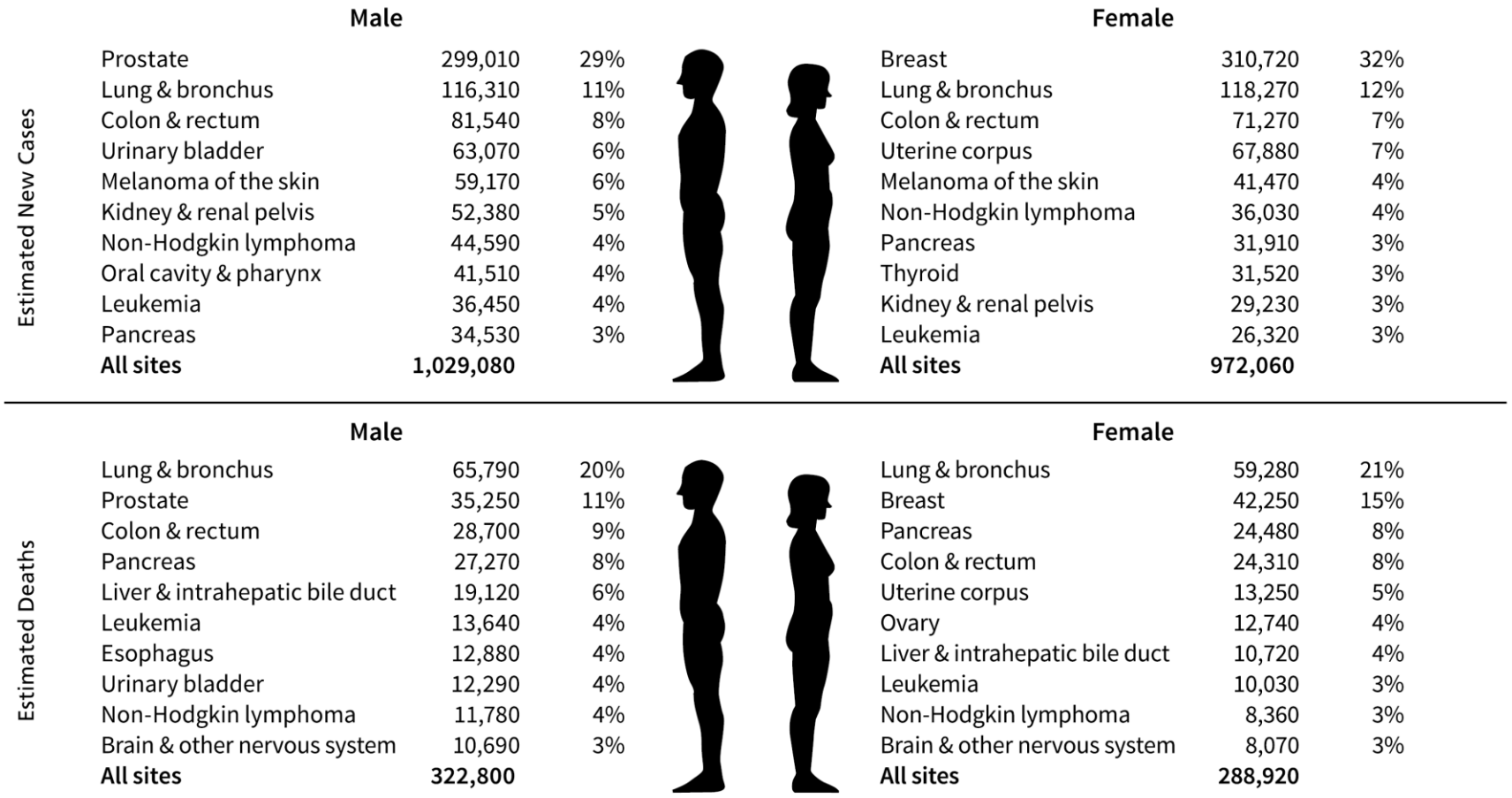
Source: American Cancer Society (ACS): Colorectal Cancer Facts & Figures 2023-2025 [292]

Figure 1-2. Global Incidence and Mortality of leading cancers in 2022



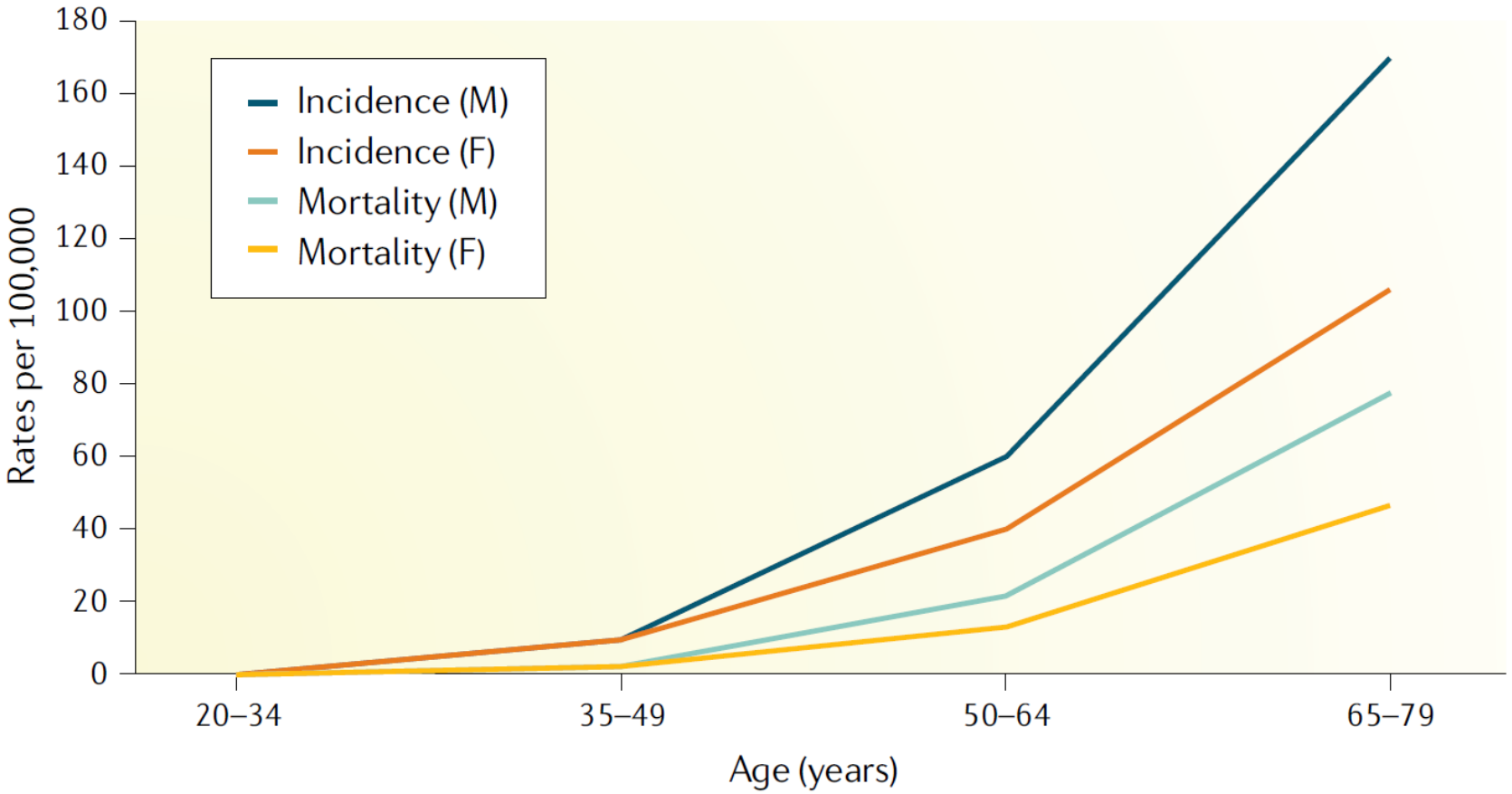
Source: Global Cancer Observatory: Cancer Today [2]

Figure 1-3. Ten leading cancer types for the estimated new cancer cases and deaths by sex, United States, 2024



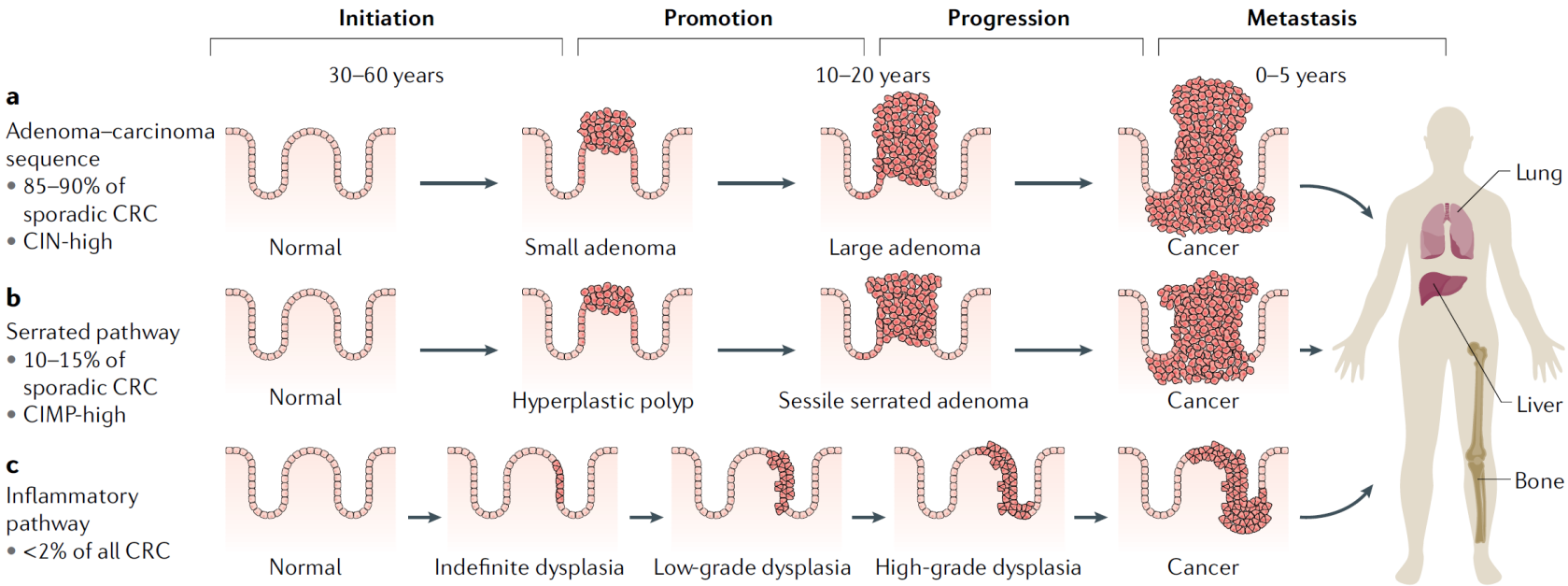
Source: American Cancer Society (ACS): Cancer statistics, 2024 [5]

Figure 1-4. Incidence and mortality of colorectal cancer by age and sex worldwide



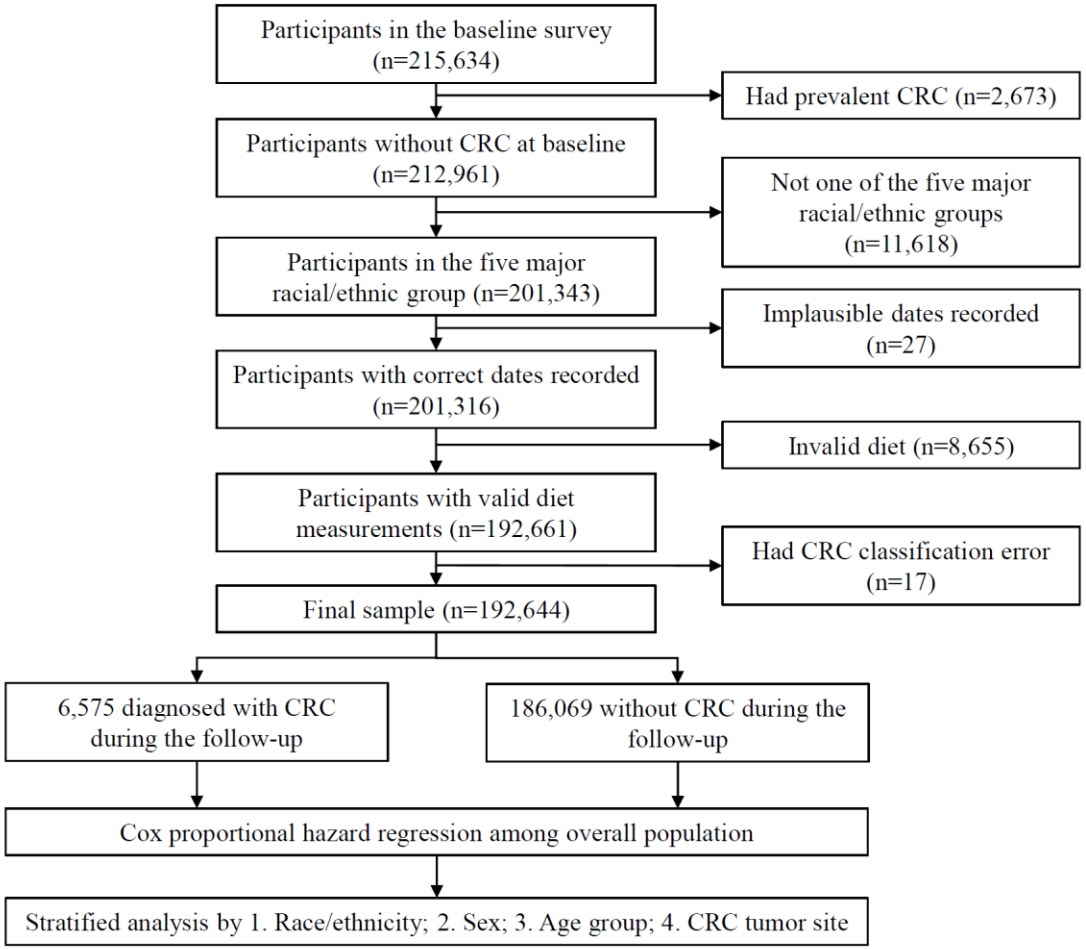
Source: Keum and Giovannucci (2019) [1]

Figure 1-5. Pathways of colorectal carcinogenesis



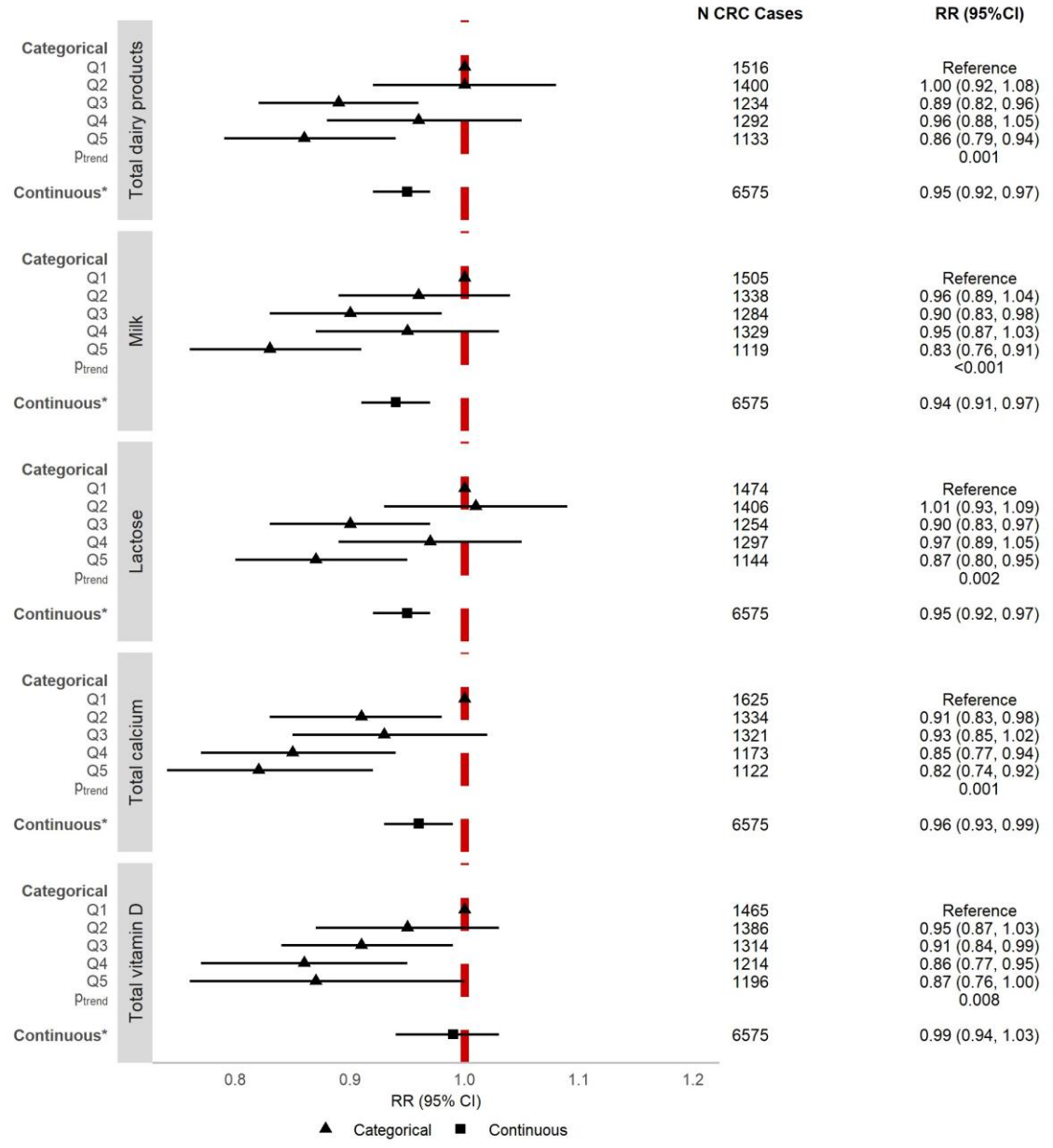
Source: Keum and Giovannucci (2019) [1]

Figure 3-1. Flowchart of the study



Abbreviations: CRC, colorectal cancer.

Figure 3-2. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of CRC among participants in the Multiethnic Cohort



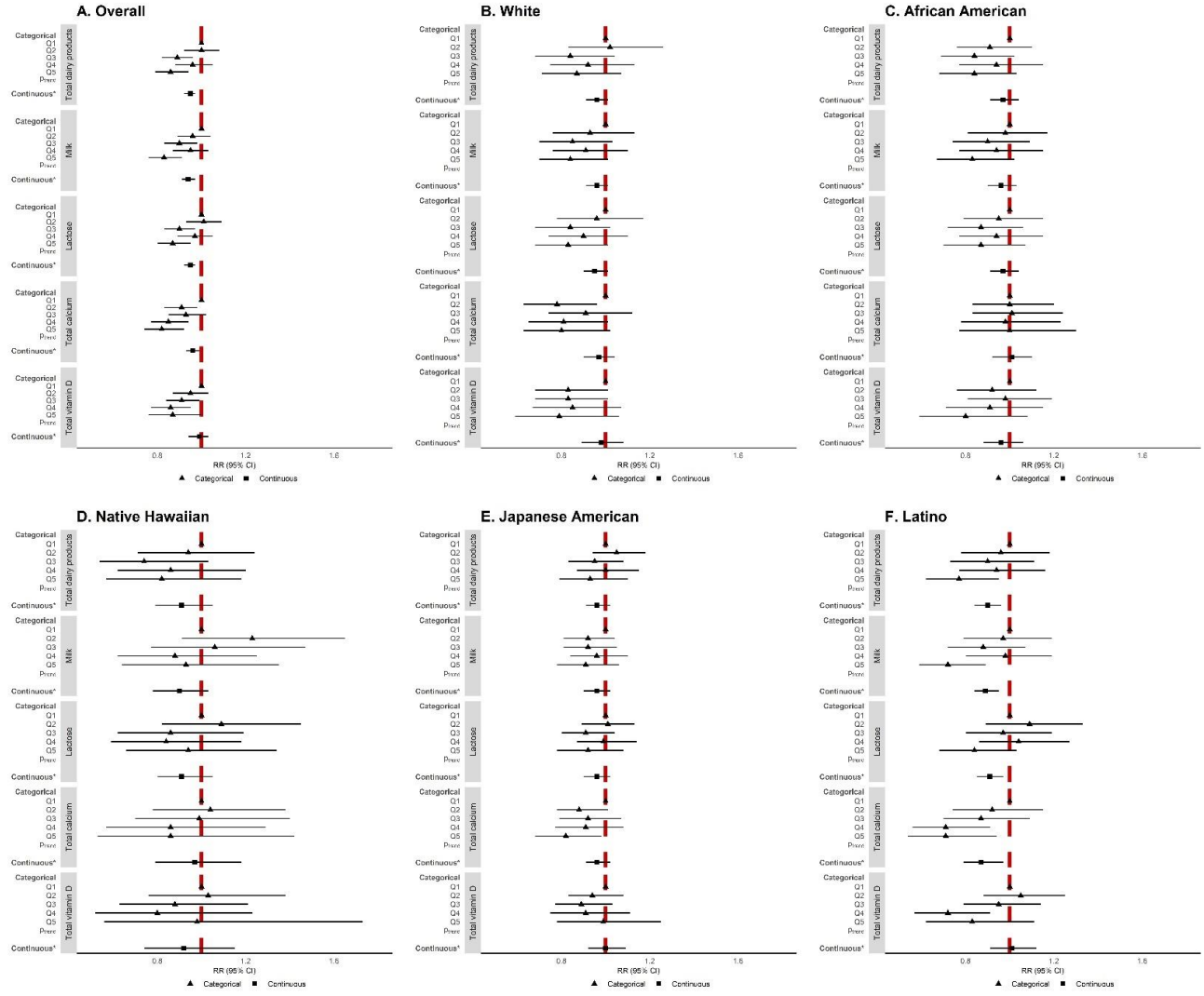
[a] The Cox proportional hazards models adjusted for sex, race and ethnicity group, 10-year age group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, history of intestinal polyps, education, BMI, smoking status and pack-years, alcohol consumption, physical activity, diabetes, use of NSAIDs, regular use of multivitamins, total energy intake, red meat intake, processed meat intake, folate intake, dietary fiber intake, and hormone use.

* Unit for the continuous intake is per SD increase.

Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.

Abbreviations: CRC, colorectal cancer; RR, relative risk; CI, confidence interval; SD, standard deviation; BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Figure 3-3. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of CRC across race and ethnicity groups in the Multiethnic Cohort



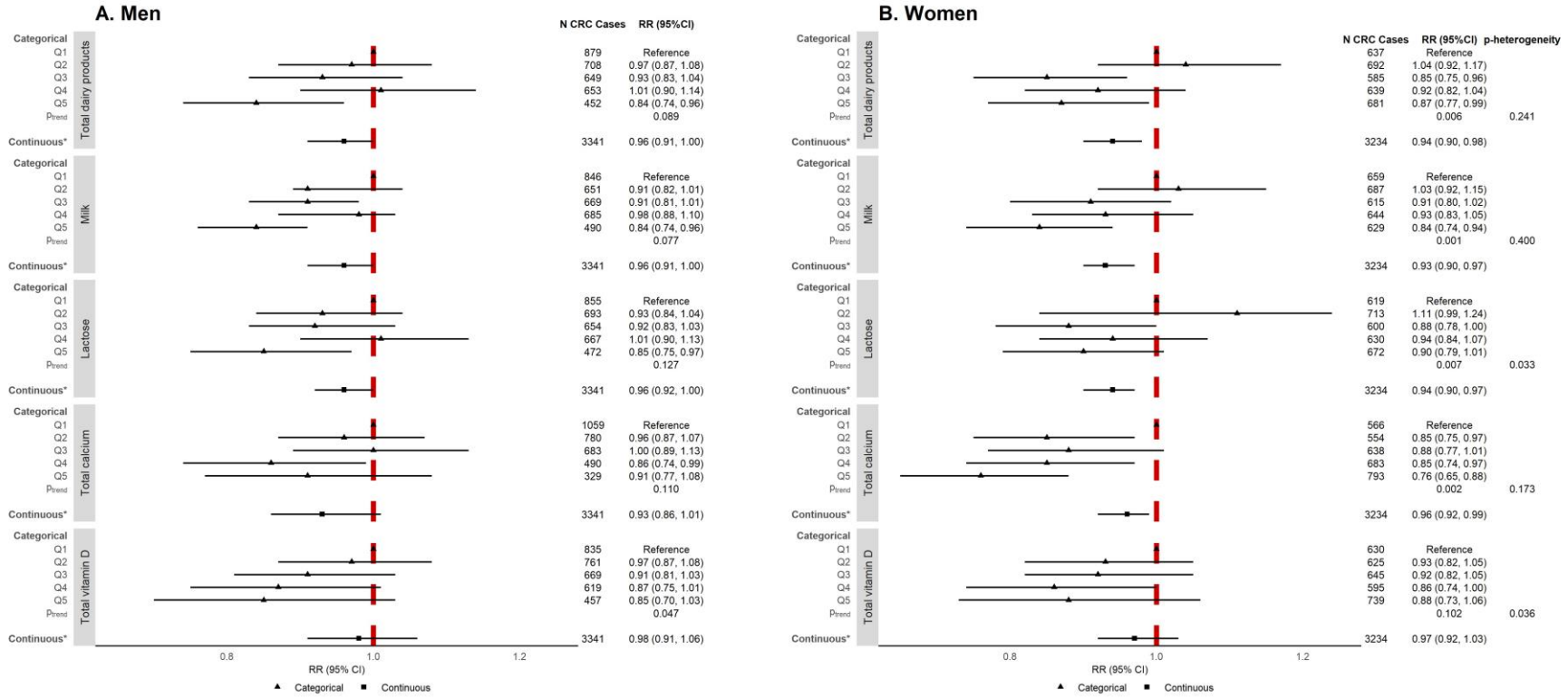
[a] The Cox proportional hazards models adjusted for sex and 10-year age group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, history of intestinal polyps, education, BMI, smoking status and pack-years, alcohol consumption, physical activity, diabetes, use of NSAIDs, regular use of multivitamins, total energy intake, red meat intake, processed meat intake, folate intake, dietary fiber intake, and hormone use.

* Unit for the continuous intake is per SD increase.

Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.

Abbreviations: CRC, colorectal cancer; RR, relative risk; CI, confidence interval; SD, standard deviation; BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Figure 3-4. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of CRC in men and women in the Multiethnic Cohort



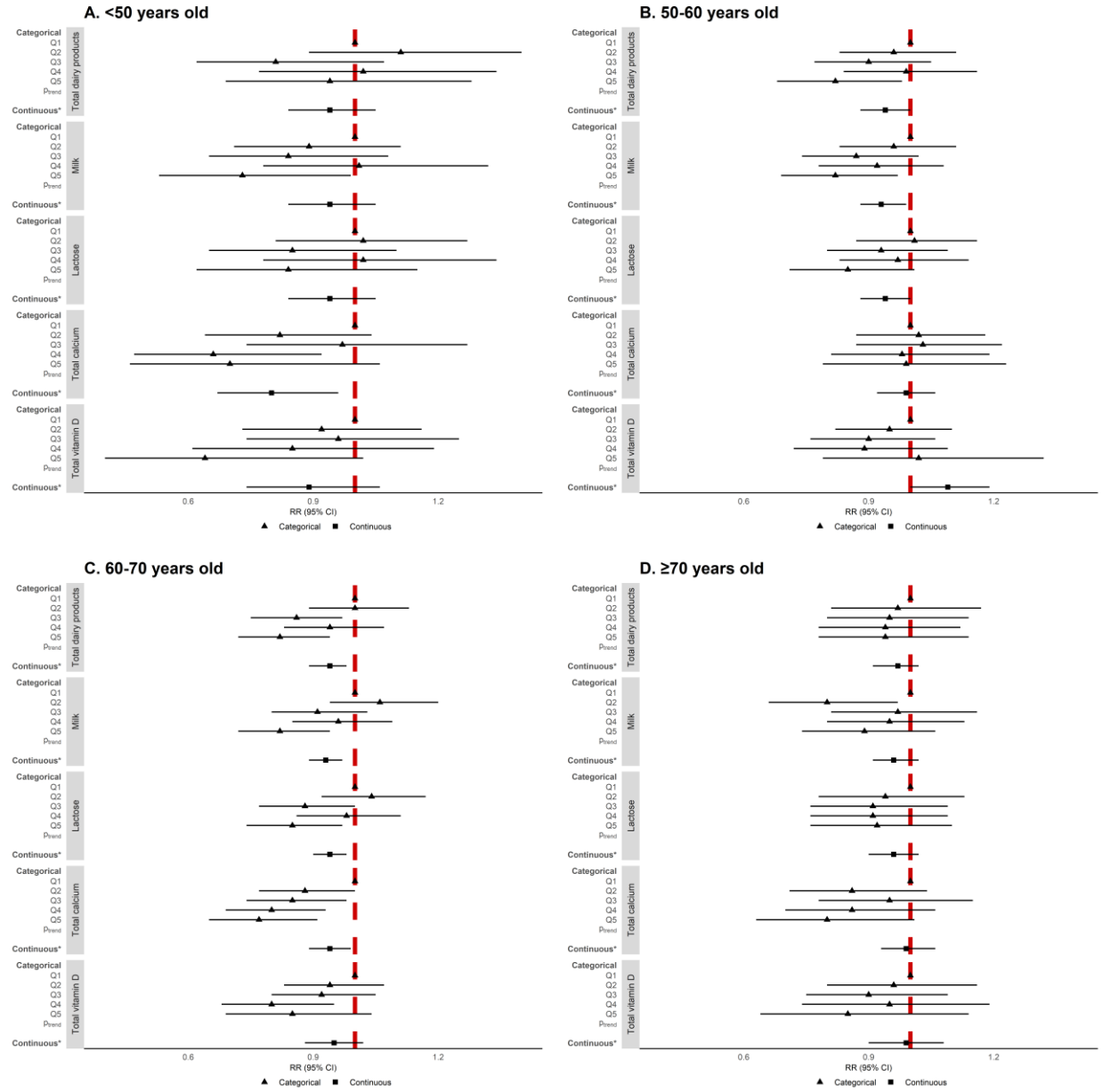
[a] The Cox proportional hazards models adjusted for race and ethnicity group and 10-year age group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, history of intestinal polyps, education, BMI, smoking status and pack-years, alcohol consumption, physical activity, diabetes, use of NSAIDs, regular use of multivitamins, total energy intake, red meat intake, processed meat intake, folate intake, dietary fiber intake, and hormone use.

* Unit for the continuous intake is per SD increase.

Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.

Abbreviations: CRC, colorectal cancer; RR, relative risk; CI, confidence interval; SD, standard deviation; BMI, body mass index; BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Figure 3-5. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of CRC across age groups in the Multiethnic Cohort



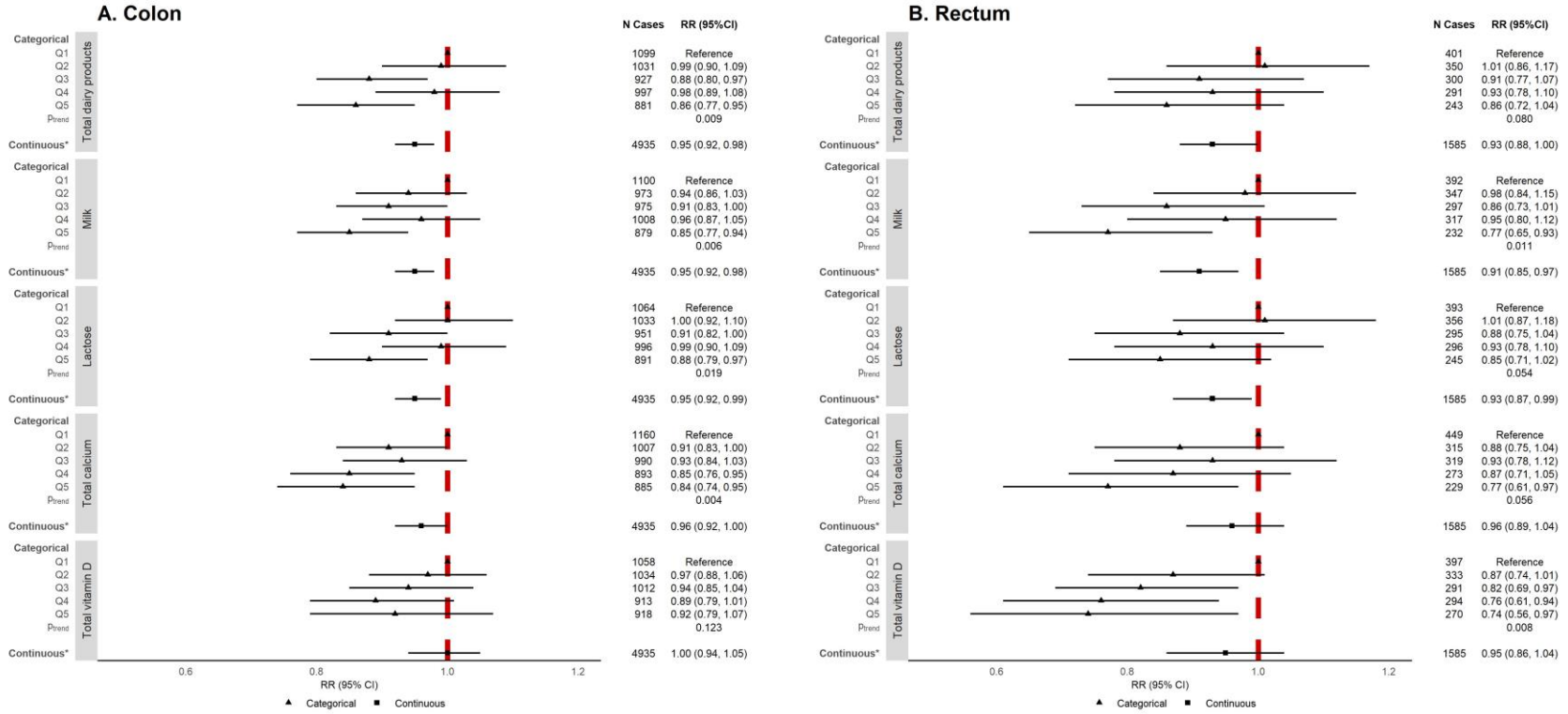
[a] The Cox proportional hazards models adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, history of intestinal polyps, education, BMI, smoking status and pack-years, alcohol consumption, physical activity, diabetes, use of NSAIDs, regular use of multivitamins, total energy intake, red meat intake, processed meat intake, folate intake, dietary fiber intake, and hormone use.

* Unit for the continuous intake is per SD increase.

Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.

Abbreviations: CRC, colorectal cancer; RR, relative risk; CI, confidence interval; SD, standard deviation; BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Figure 3-6. Forest plot for association between intakes of total dairy products, milk, lactose, calcium, vitamin D and the relative risk of colon and rectal cancer among participants in the Multiethnic Cohort



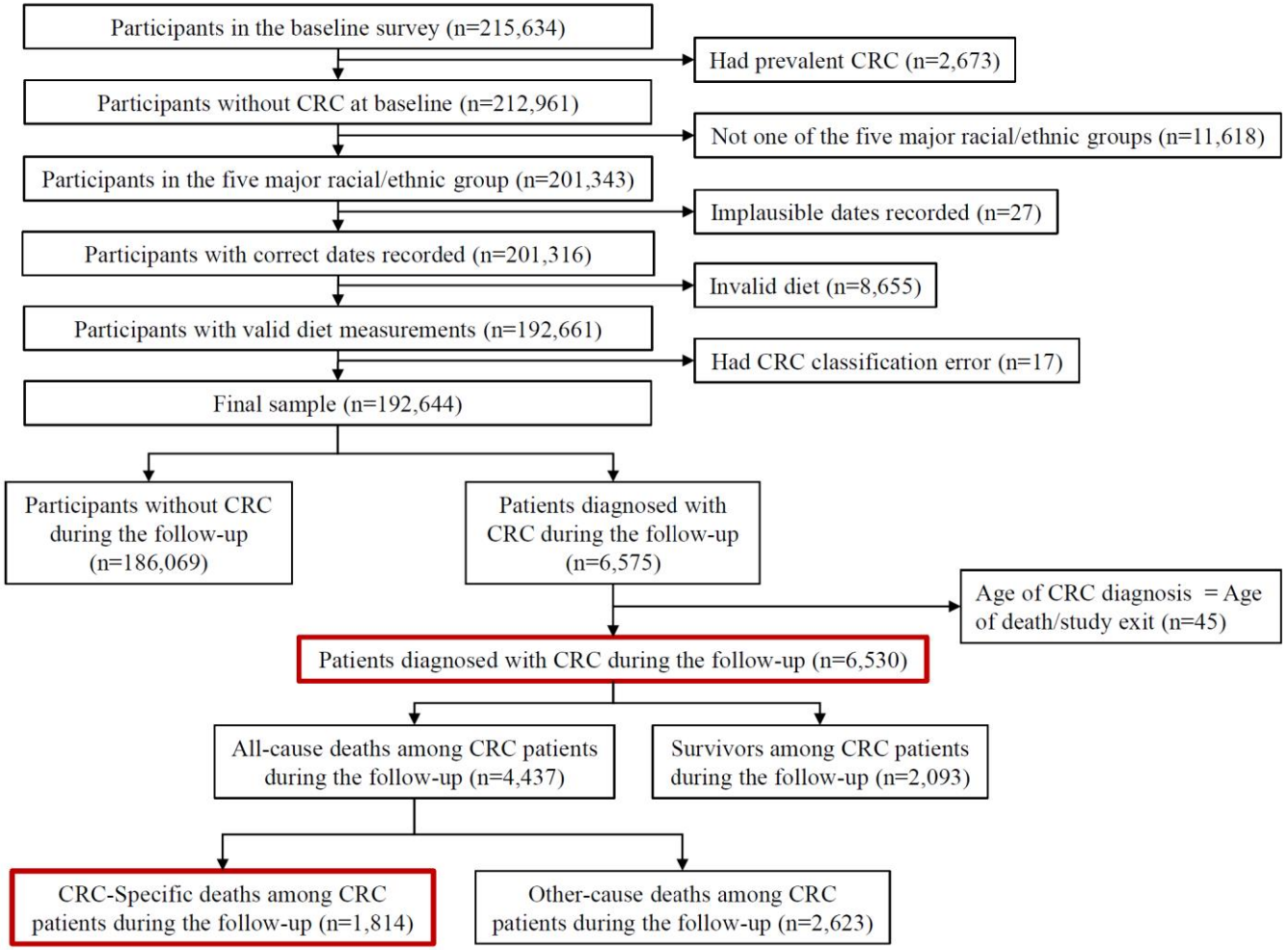
[a] The Cox proportional hazards models adjusted for sex, race and ethnicity group, 10-year age group as strata variables and the following variables as proportional hazards covariates: family history of colorectal cancer, history of intestinal polyps, education, BMI, smoking status and pack-years, alcohol consumption, physical activity, diabetes, use of NSAIDs, regular use of multivitamins, total energy intake, red meat intake, processed meat intake, folate intake, dietary fiber intake, and hormone use.

* Unit for the continuous intake is per SD increase.

Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.

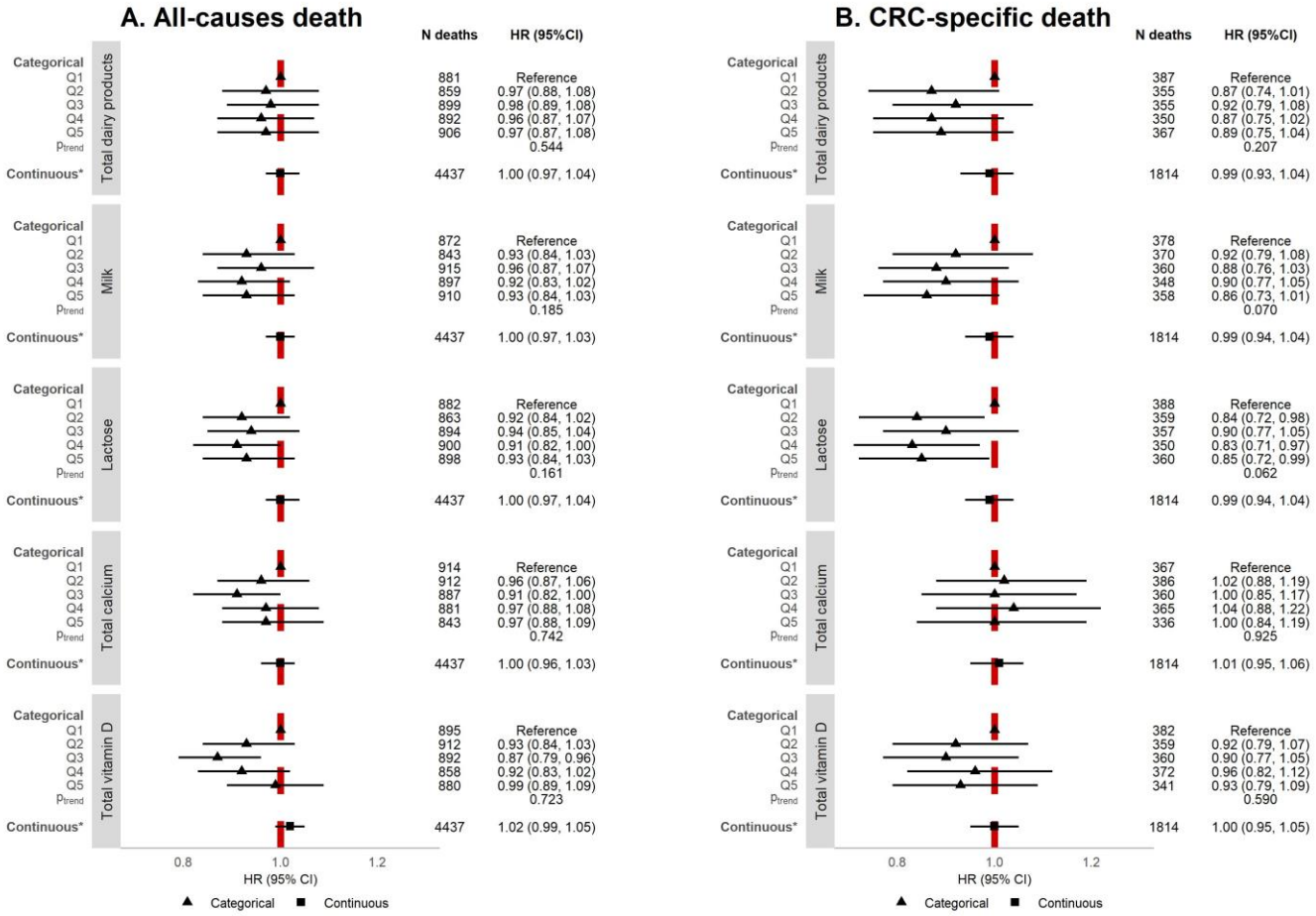
Abbreviations: CRC, colorectal cancer; RR, relative risk; CI, confidence interval; SD, standard deviation; BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs; IU, international units.

Figure 4-1. Flowchart of the study



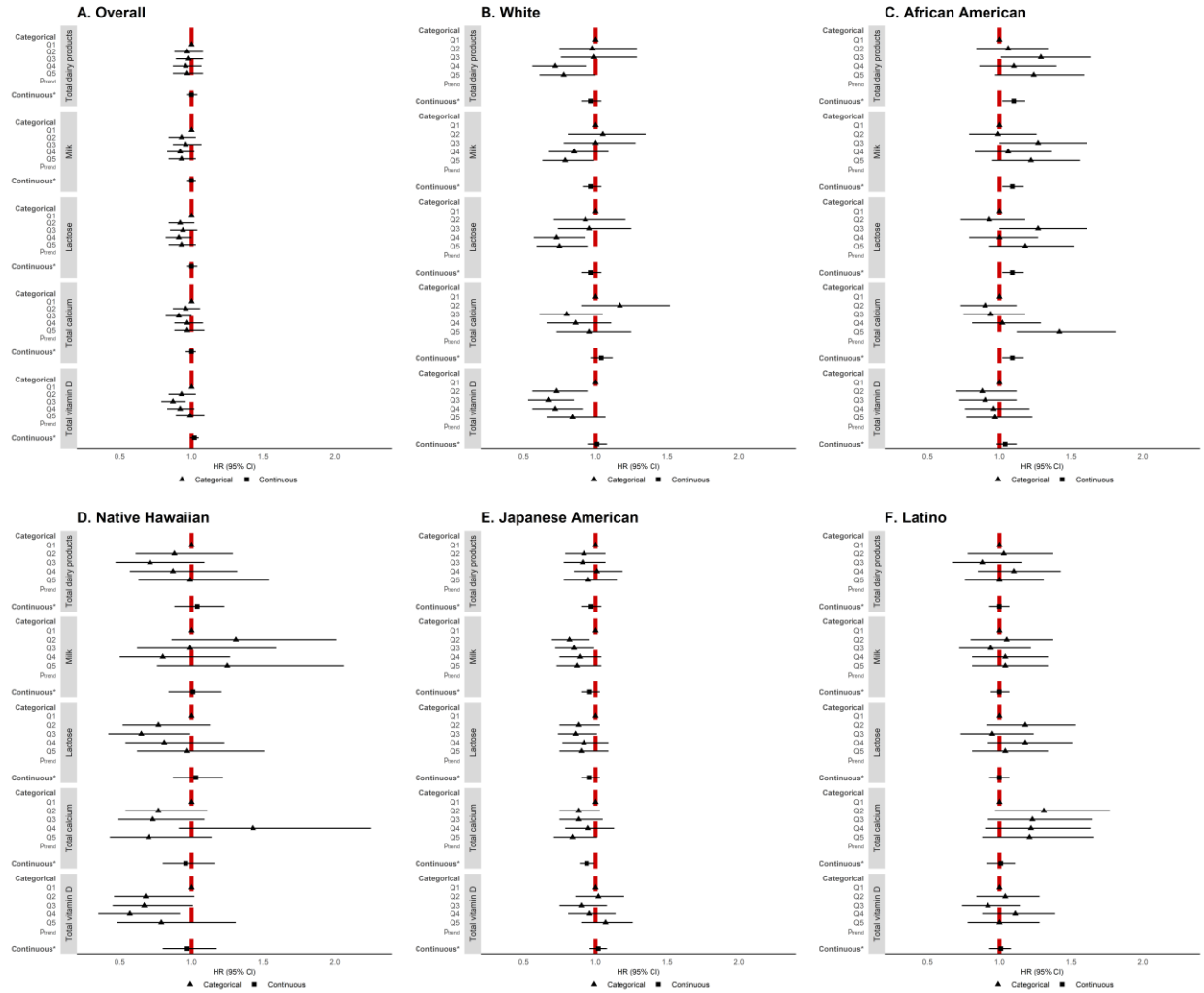
Abbreviations: CRC, colorectal cancer.

Figure 4-2. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and survival among CRC patients in the Multiethnic Cohort



[a] The Cox proportional hazards models adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.
* Unit for the continuous intake is per SD increase.
Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.
Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; NASIDs, non-steroidal anti-inflammatory drugs.

Figure 4-3. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and the overall survival among CRC patients across race and ethnicity groups in the Multiethnic Cohort



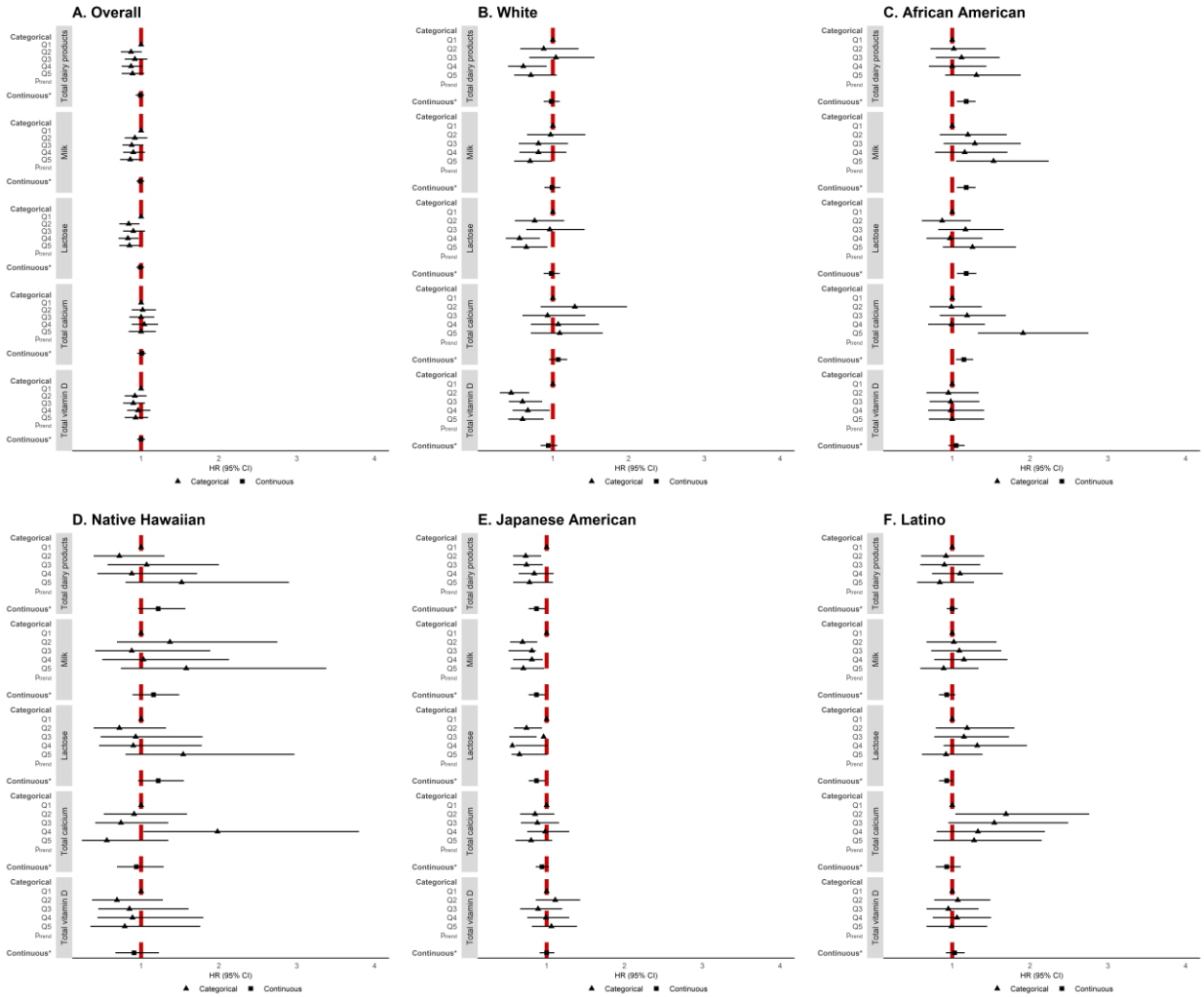
[a] The Cox proportional hazards models adjusted for sex as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

* Unit for the continuous intake is per SD increase.

Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.

Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; NASIDs, non-steroidal anti-inflammatory drugs.

Figure 4-4. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and the CRC-specific survival among CRC patients across race and ethnicity groups in the Multiethnic Cohort



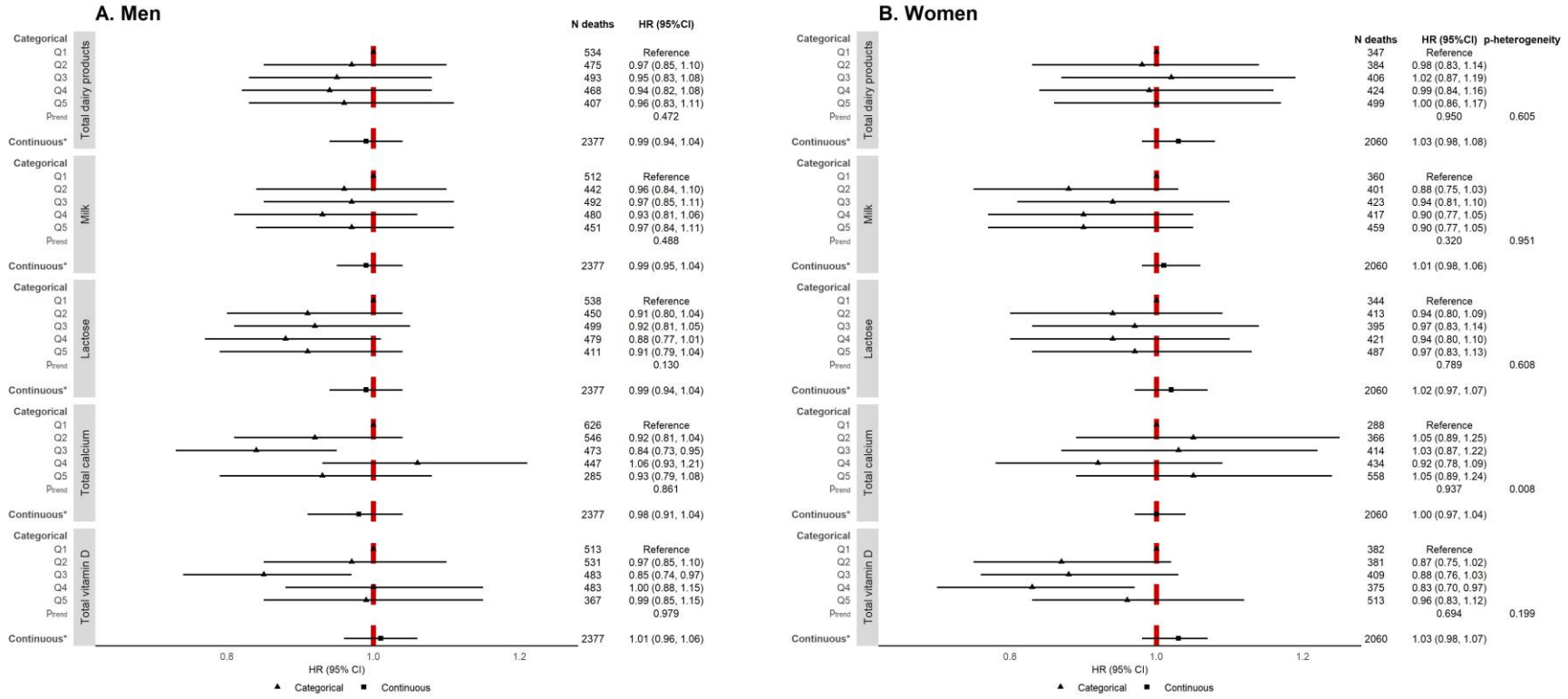
[a] The Cox proportional hazards models adjusted for sex as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.

* Unit for the continuous intake is per SD increase.

Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.

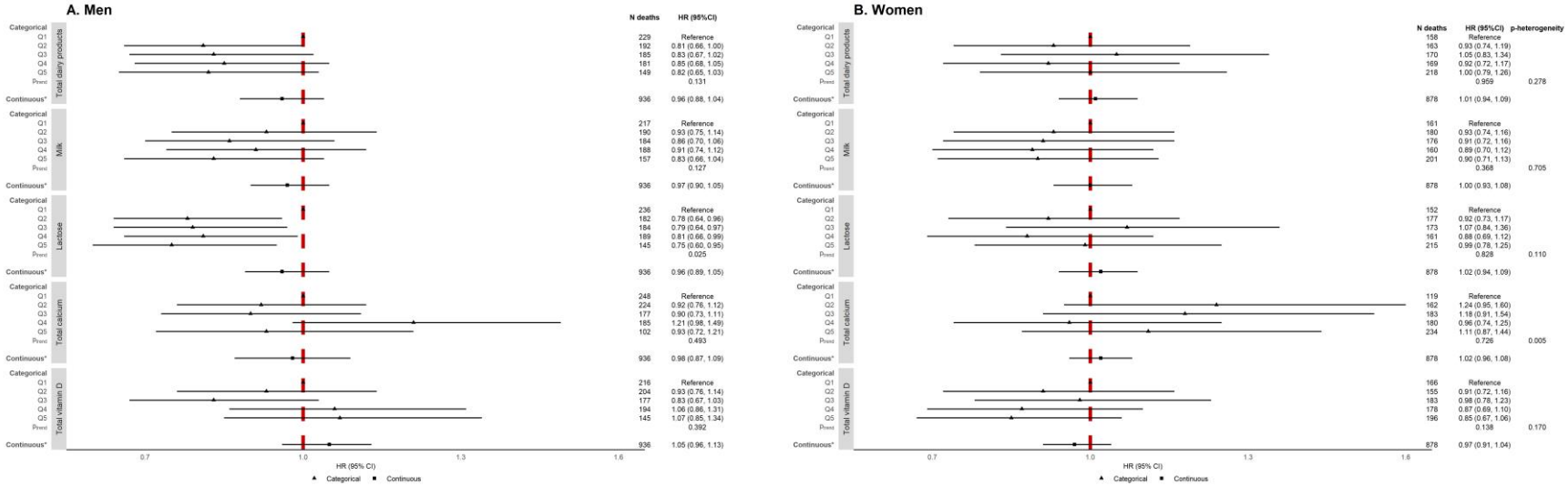
Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; NASIDs, non-steroidal anti-inflammatory drugs.

Figure 4-5. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and overall survival among CRC patients in men and women in the Multiethnic Cohort



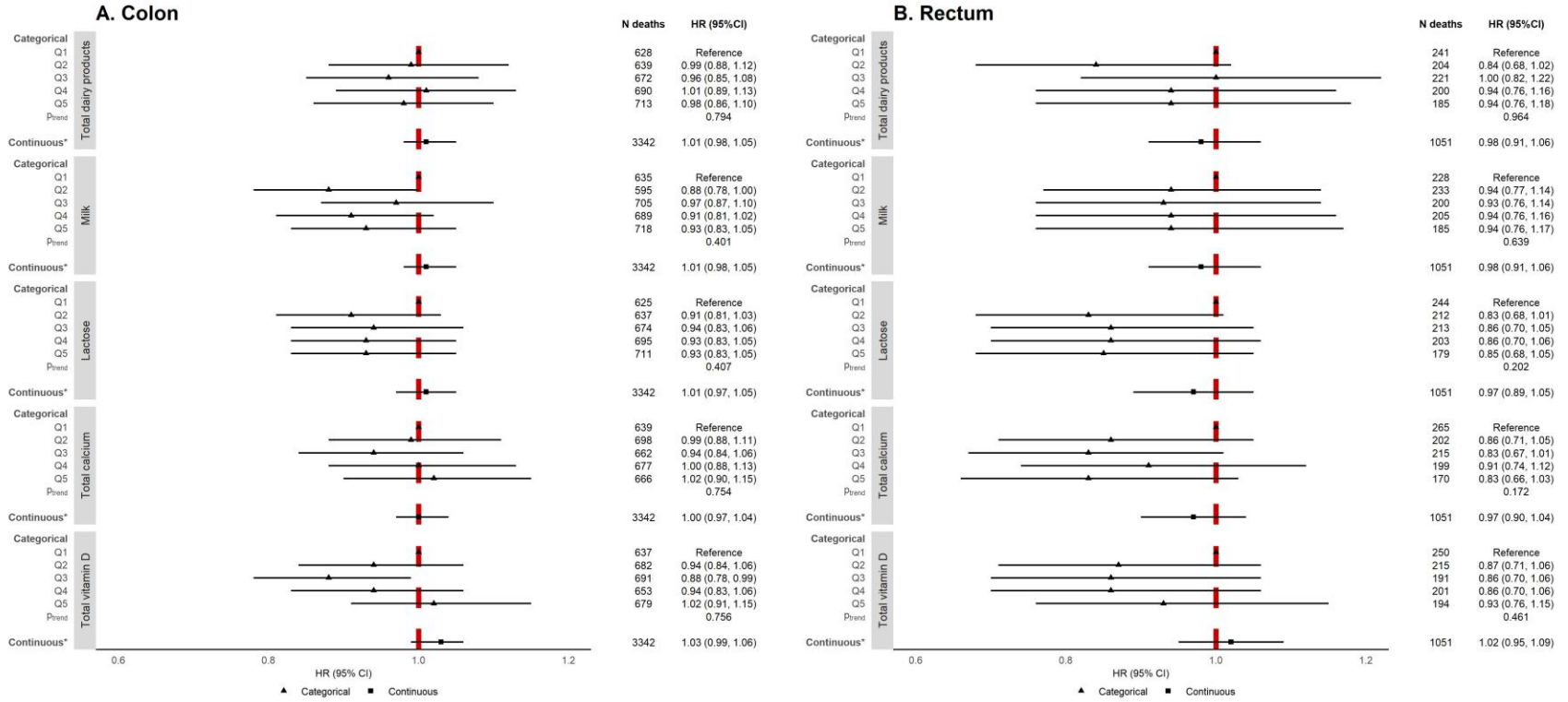
[a] The Cox proportional hazards models adjusted for race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.
* Unit for the continuous intake is per SD increase.
Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.
Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; NASIDs, non-steroidal anti-inflammatory drugs.

Figure 4-6. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and CRC-specific survival among CRC patients in men and women in the Multiethnic Cohort



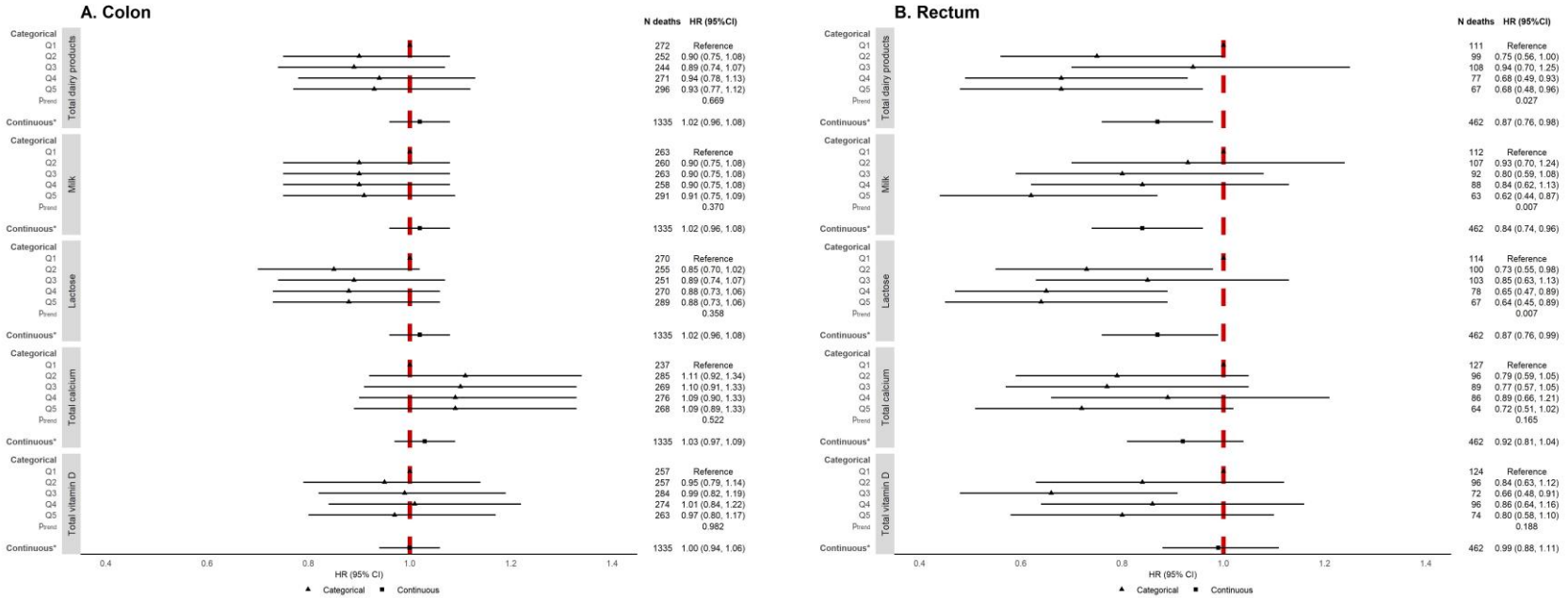
[a] The Cox proportional hazards models adjusted for race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.
* Unit for the continuous intake is per SD increase.
Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.
Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; NASIDs, non-steroidal anti-inflammatory drugs.

Figure 4-7. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and the overall survival among colon and rectal cancer patients in the Multiethnic Cohort



[a] The Cox proportional hazards models adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.
* Unit for the continuous intake is per SD increase.
Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.
Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; NASIDs, non-steroidal anti-inflammatory drugs.

Figure 4-8. Forest plot for association[a] between intakes of total dairy products, milk, lactose, calcium, vitamin D and the CRC-specific survival among colon and rectal cancer patients in the Multiethnic Cohort



[a] The Cox proportional hazards models adjusted for sex, race and ethnicity group as strata variables and the following variables as proportional hazards covariates: age at diagnosis in 10-year age group, family history of colorectal cancer, education, BMI, smoking status and pack-years, physical activity, use of NSAIDs, total energy intake, comorbidity, radiation therapy, and chemotherapy.
* Unit for the continuous intake is per SD increase.
Unit of each exposure intake: total dairy products, g/1000 kcal/day; milk, g/1000 kcal/day; lactose, g/1000 kcal/day; total calcium, mg/1000 kcal/day; total vitamin D, IU/1000 kcal/day.
Abbreviations: CRC, colorectal cancer; SD, standard deviation; HR, hazard ratio; CI, confidence interval; NASIDs, non-steroidal anti-inflammatory drugs.

Figure 5-1. Flowchart of the study

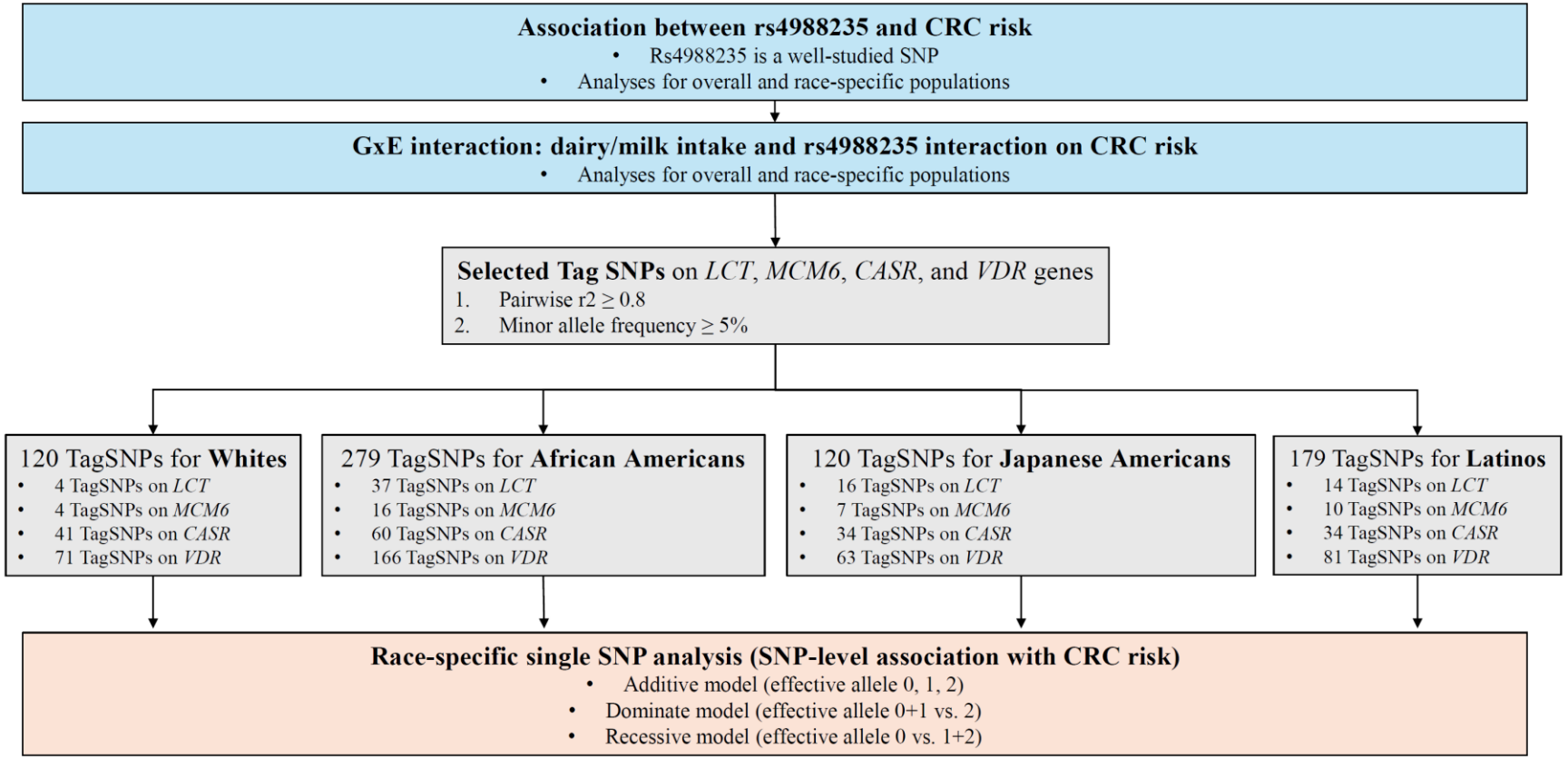
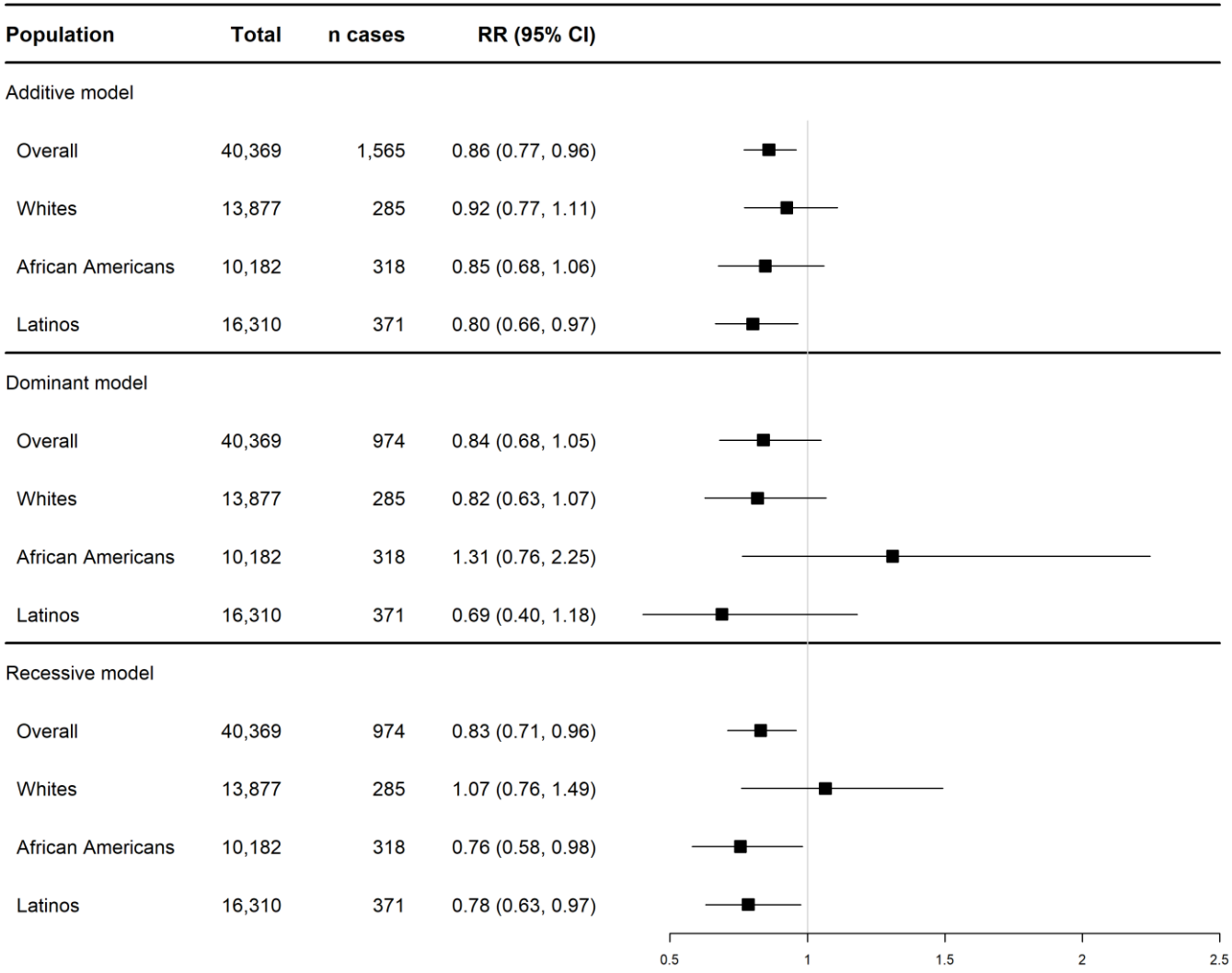
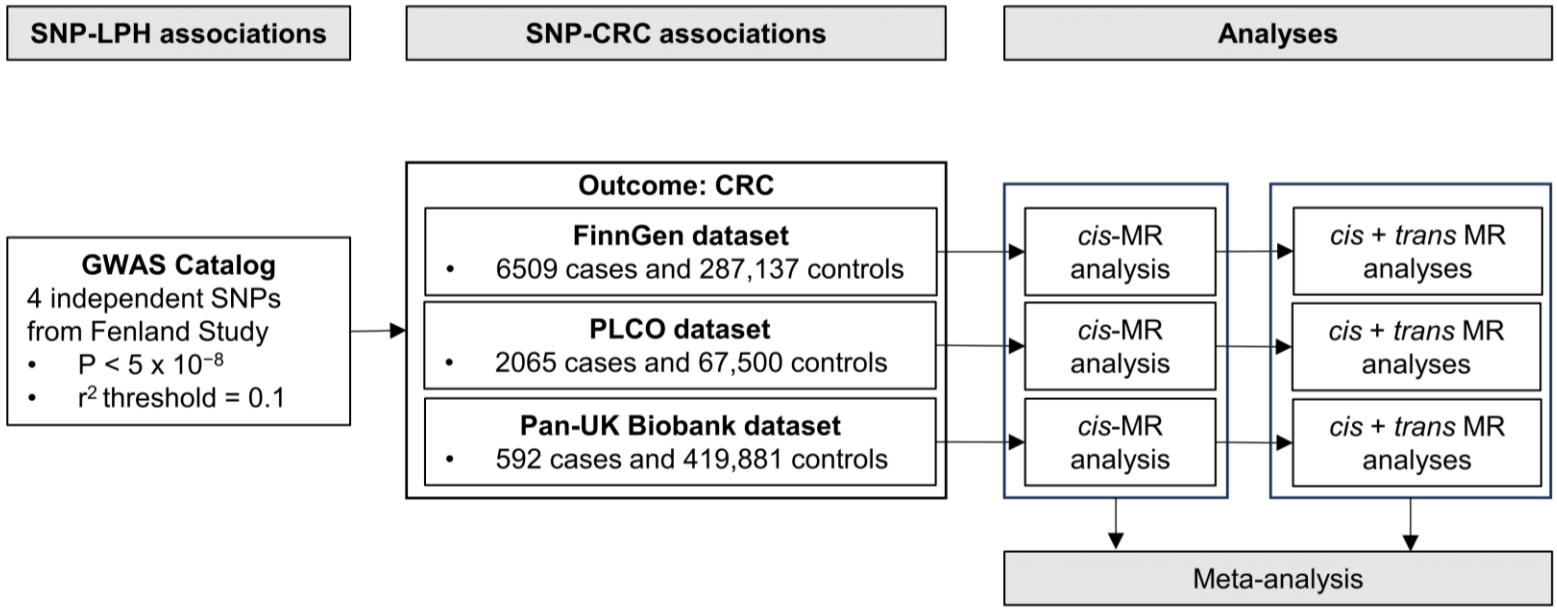


Figure 5-2. Association between rs4988235 and risk of CRC among participants in the Multiethnic cohort



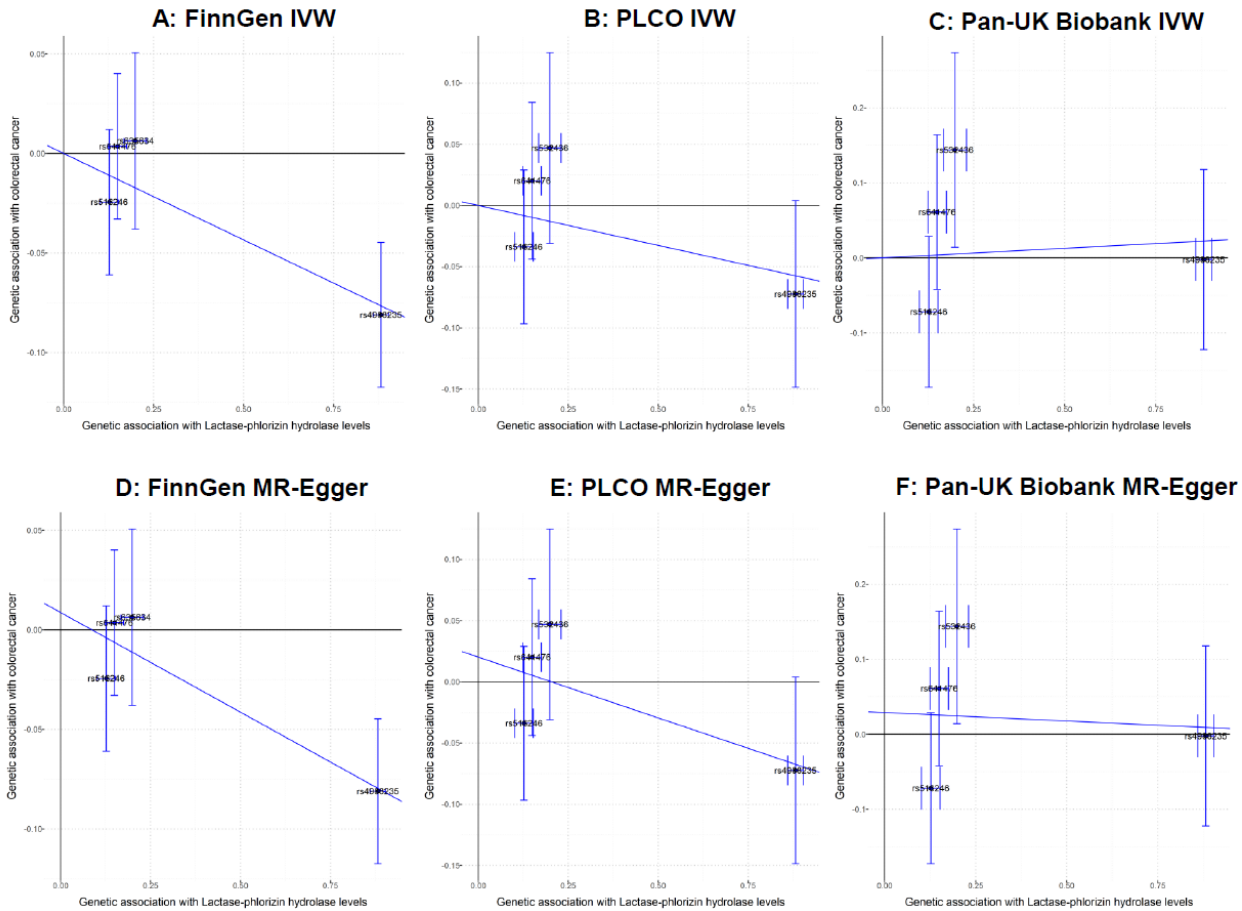
Abbreviations: RR, relative risk; CRC, colorectal cancer; CI, confidence interval.

Figure 6-1. Schematic overview of the study design



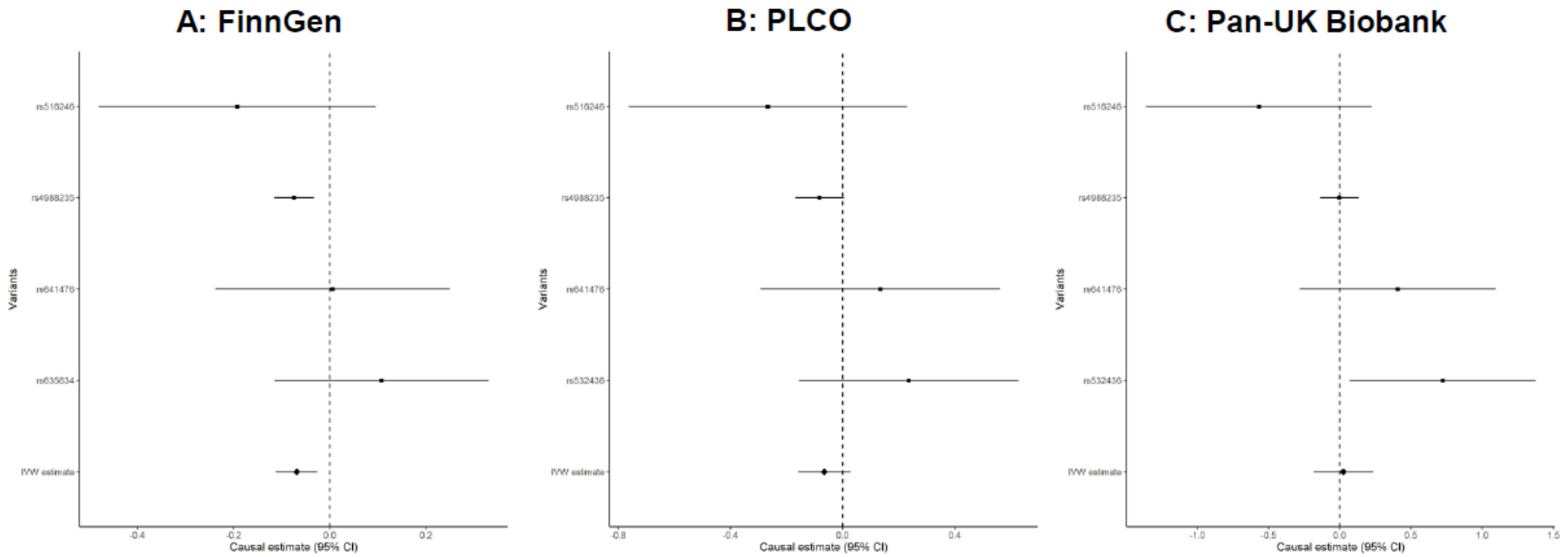
Abbreviations: SNP, single nucleotide polymorphism; LPH, lactase-phlorizin hydrolase; CRC, colorectal cancer; GWAS, genome-wide association study; MR, Mendelian Randomization

Figure 6-2. Scatter plots of the IVW and MR-Egger methods investigating the effect of elevated LPH levels on CRC in the FinnGen, PLCO, and Pan-UK Biobank datasets



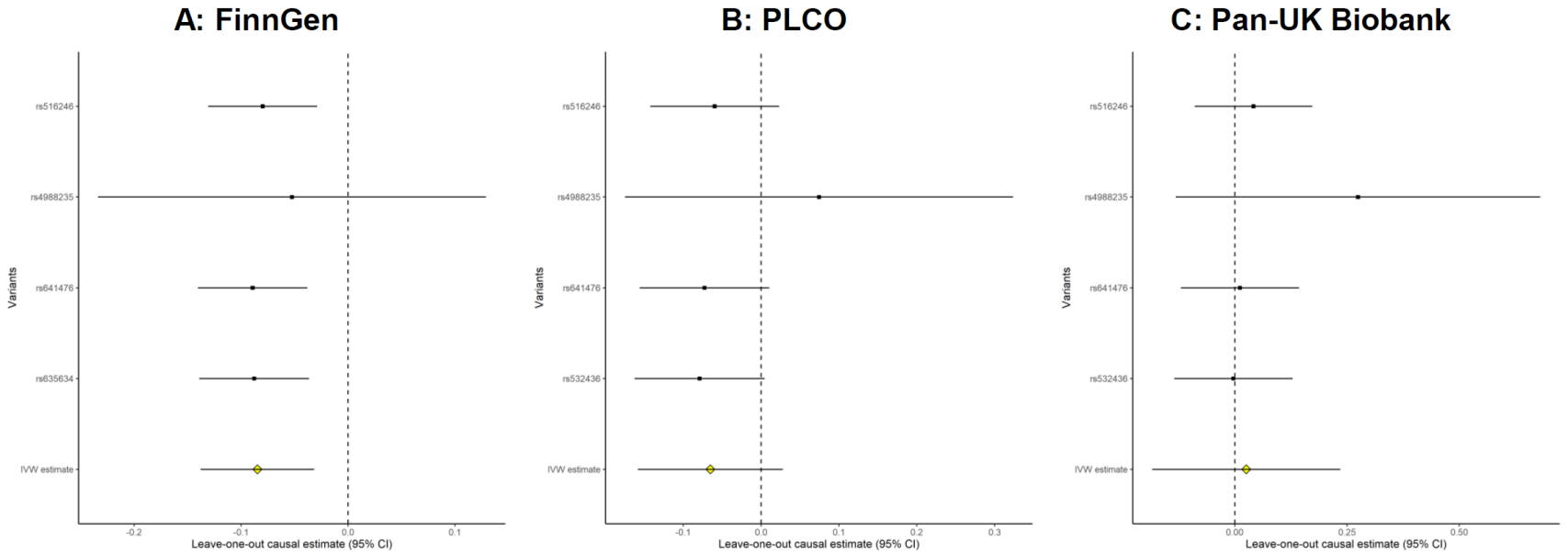
The x-axis represents the genetic association with LPH; the y-axis represents the genetic association with risk of CRC. A: FinnGen dataset using IVW method; B: PLCO dataset using IVW method; C: Pan-UK Biobank dataset using IVW method; D: FinnGen dataset using MR-Egger method; E: PLCO dataset using MR-Egger method; F: Pan-UK Biobank dataset using MR-Egger method. IVW, inverse-variance weighted; MR: Mendelian Randomization; LPH, lactase-phlorizin hydrolase; CRC, colorectal cancer

Figure 6-3. Forest plots of the IVW estimates on the association between genetically predicted LPH levels and CRC risk for each genetic instrument in the FinnGen, PLCO, and Pan-UK Biobank datasets



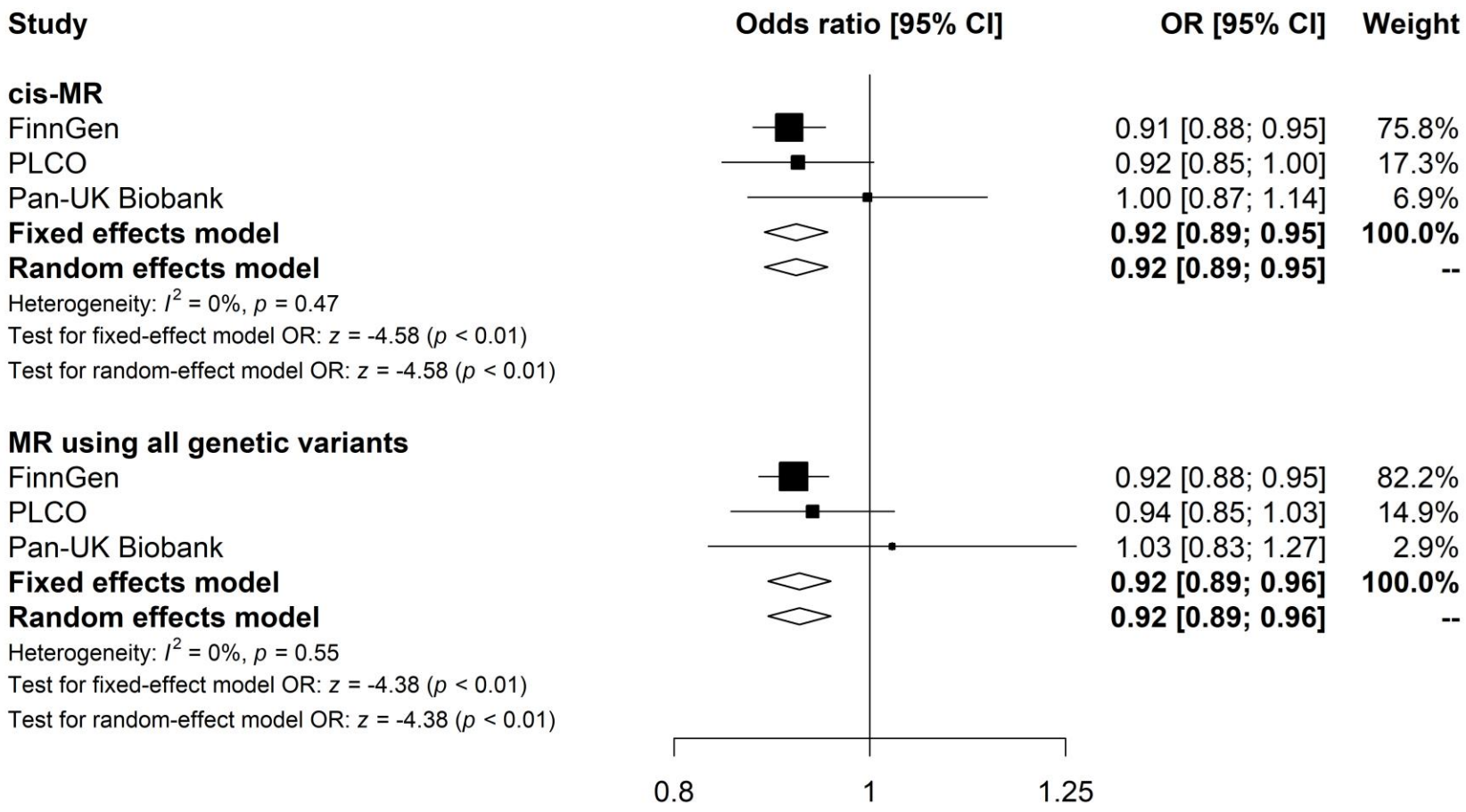
The x-axis represents the IVW causal estimate with its 95% CI; the y-axis represents genetic variant. A: FinnGen dataset; B: PLCO dataset; C: Pan-UK Biobank dataset. IVW, inverse-variance weighted; LPH, lactase-phlorizin hydrolase; CRC, colorectal cancer; CI: confidence interval

Figure 6-4. Leave-one-out analyses for the MR analysis on LPH levels and CRC risk in the FinnGen, PLCO, and Pan-UK Biobank datasets



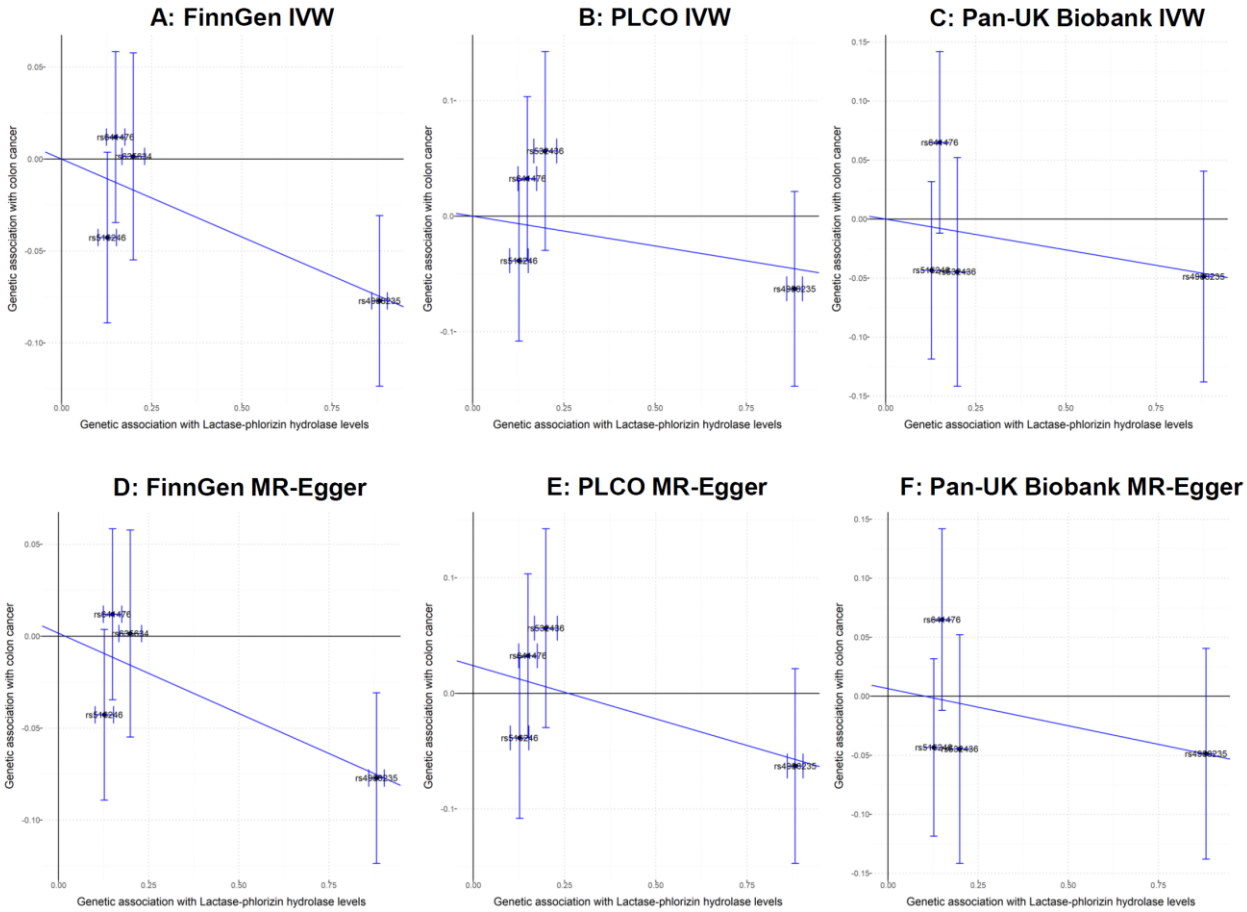
The x-axis represents the IVW estimate after removing the corresponding SNP, the Y-axis represents which genetic variant was removed from the MR analysis. MR, Mendelian Randomization; LPH, lactase-phlorizin hydrolase; CRC, colorectal cancer; IVW: inverse variance weighted; SNP, single nucleotide polymorphism; CI, confidence interval

Figure 6-5. Meta-analysis results for the association of LPH levels with CRC risk



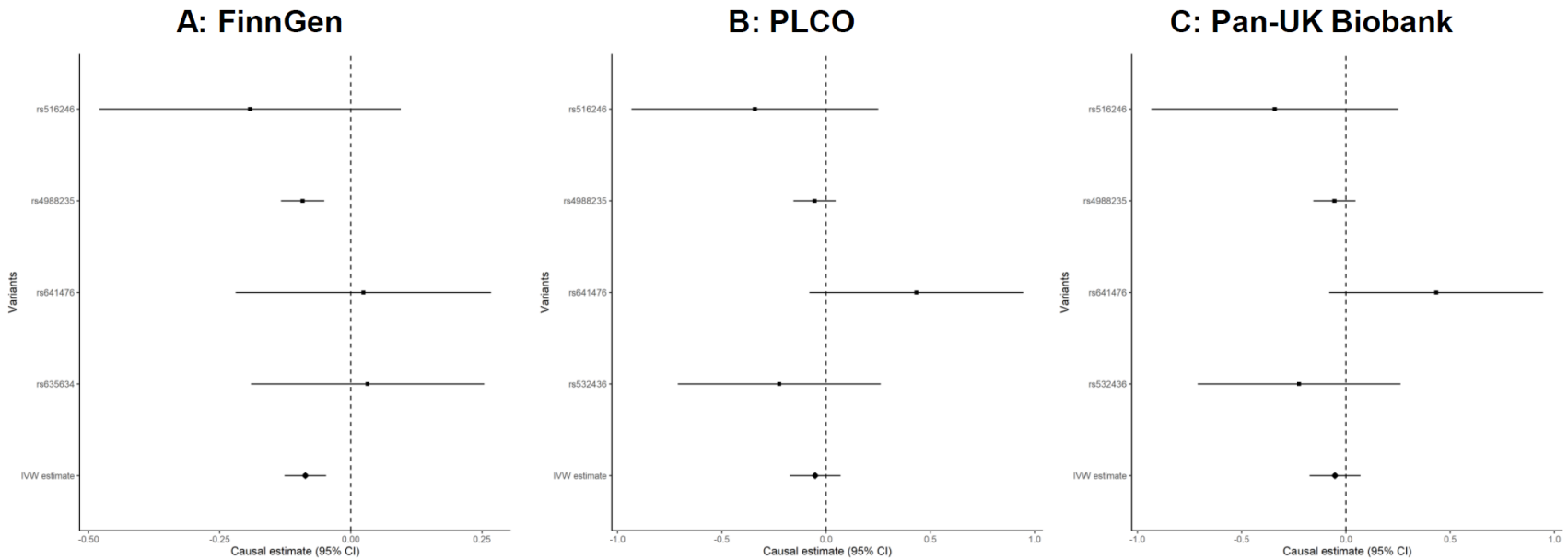
Forest plots show results from *cis*-MR and MR using all genetic variants. Squares represent study-specific MR estimates. Diamonds represent meta-analyzed MR estimates using fixed and random effects models. Abbreviations: LPH, lactase-phlorizin hydrolase; CRC, colorectal cancer; MR, Mendelian Randomization.

Figure 6-6. Scatter plots of the IVW and MR-Egger methods investigating the effect of elevated LPH levels on colon cancer in the FinnGen, PLCO, and Pan-UK Biobank datasets



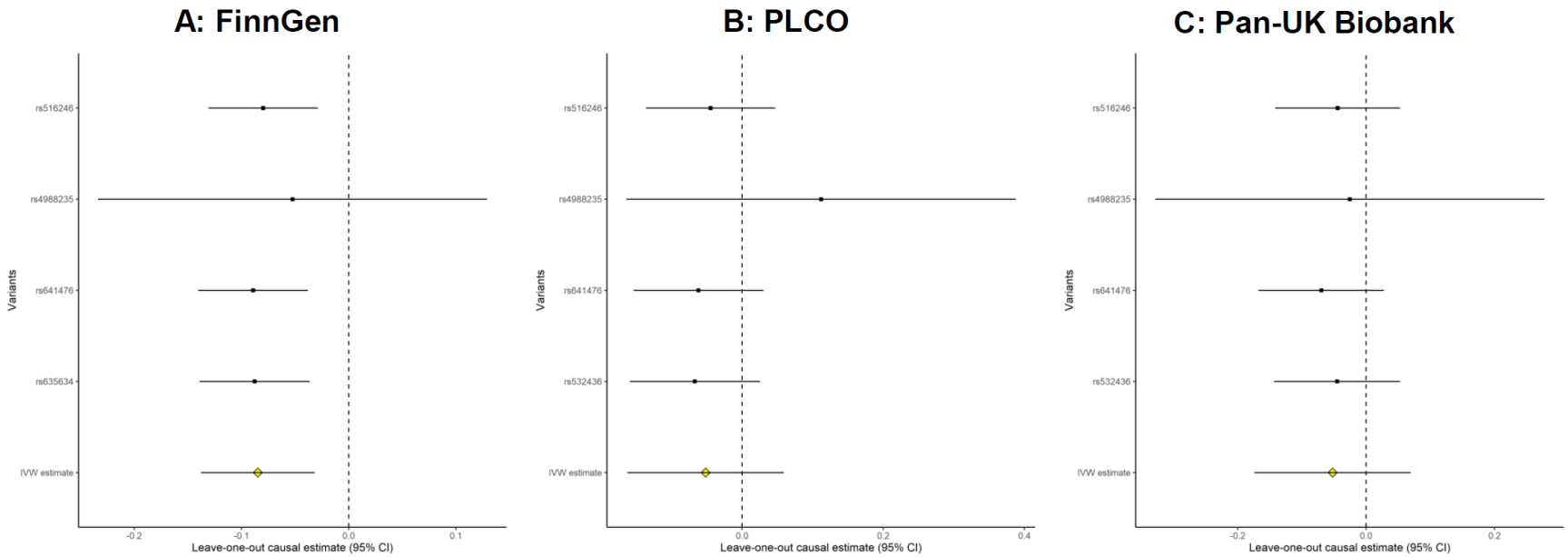
The x-axis represents the genetic association with LPH; the y-axis represents the genetic association with risk of colon cancer. A: FinnGen dataset using IVW method; B: PLCO dataset using IVW method; C: Pan-UK Biobank dataset using IVW method; D: FinnGen dataset using MR-Egger method; E: PLCO dataset using MR-Egger method; F: Pan-UK Biobank dataset using MR-Egger method. IVW, inverse-variance weighted; MR: Mendelian Randomization; LPH, lactase-phlorizin hydrolase

Figure 6-7. Forest plots of the IVW estimates on the association between genetically predicted LPH levels and colon cancer risk for each genetic instrument in the FinnGen, PLCO, and Pan-UK Biobank datasets



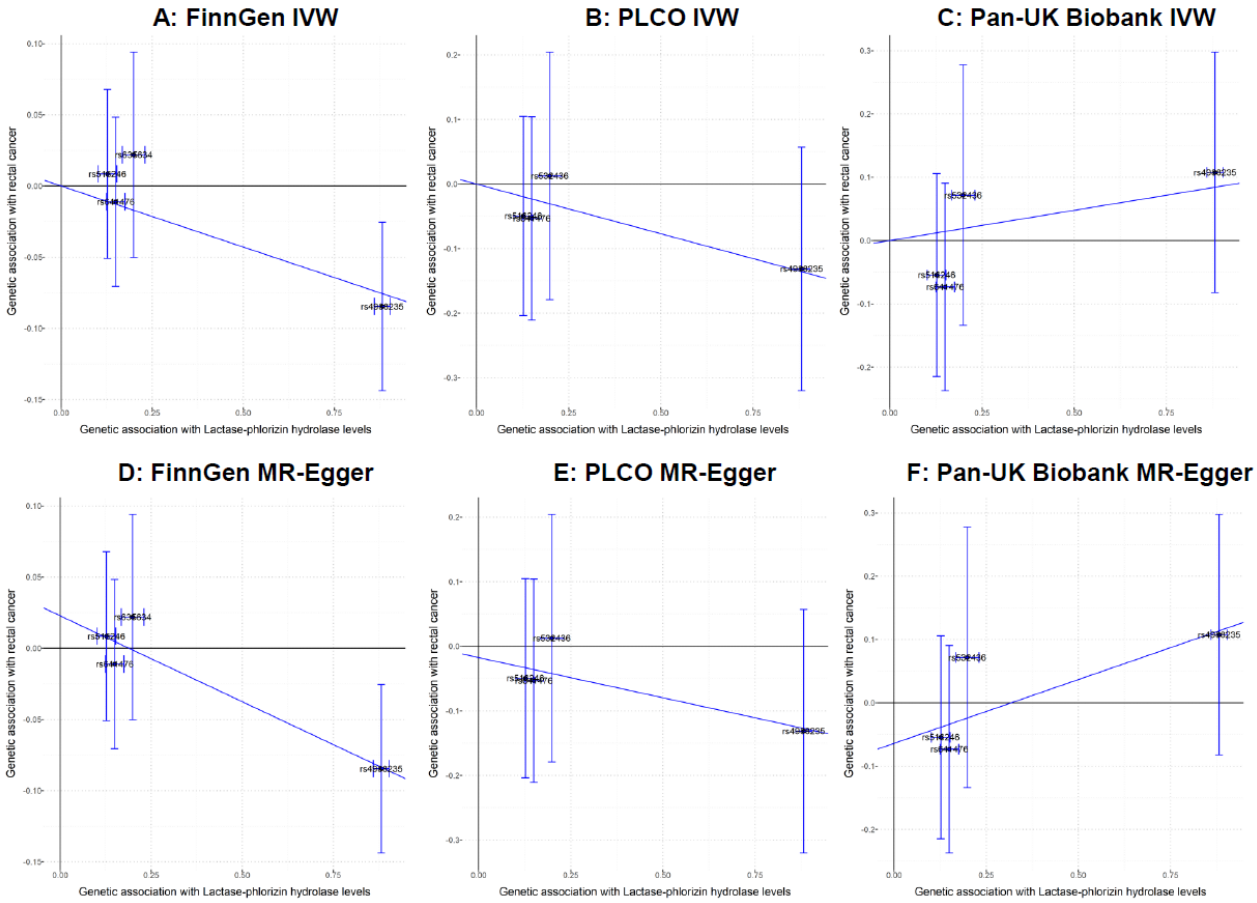
The x-axis represents the IVW causal estimate with its 95% CI; the y-axis represents genetic variant. A: FinnGen dataset; B: PLCO dataset; C: Pan-UK Biobank dataset. IVW, inverse-variance weighted; LPH, lactase-phlorizin hydrolase; CI: confidence interval

Figure 6-8. Leave-one-out analyses for the MR analysis on LPH levels and colon cancer risk in the FinnGen, PLCO, and Pan-UK Biobank datasets



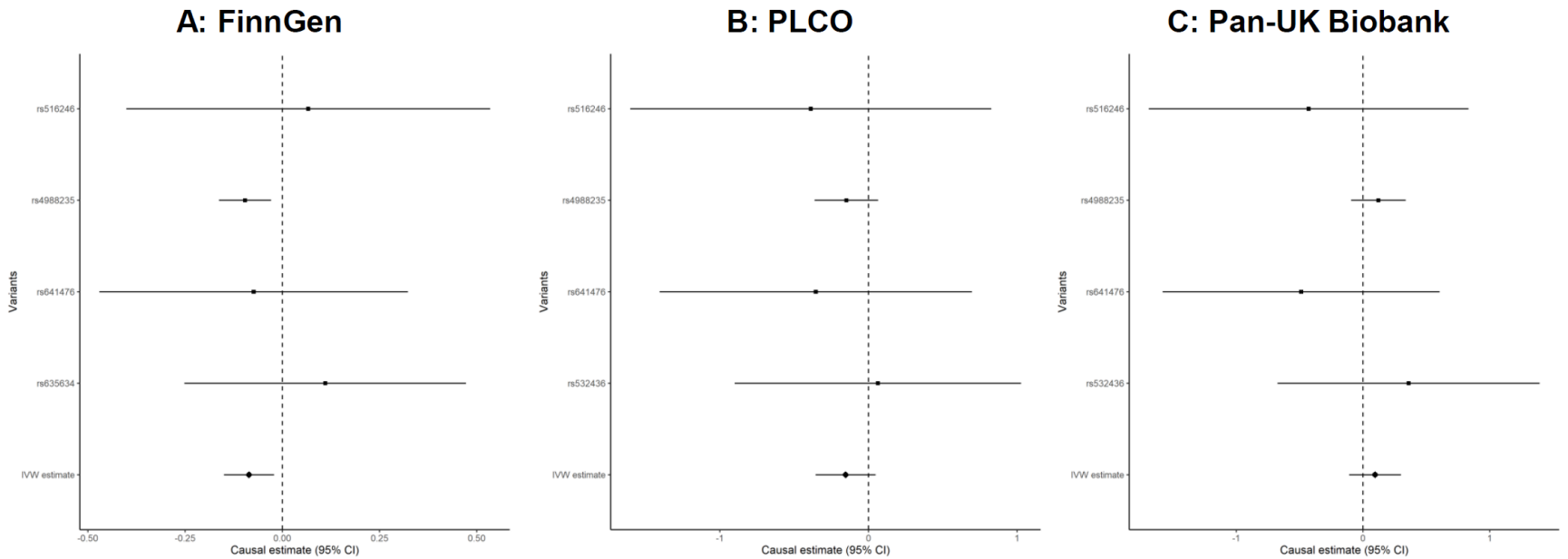
The x-axis represents the IVW estimate after removing the corresponding SNP, the Y-axis represents which genetic variant was removed from the MR analysis. MR, Mendelian Randomization; LPH, lactase-phlorizin hydrolase; IVW: inverse variance weighted; SNP, single nucleotide polymorphism; CI, confidence interval

Figure 6-9. Scatter plots of the IVW and MR-Egger methods investigating the effect of elevated LPH levels on rectal cancer in the FinnGen, PLCO, and Pan-UK Biobank datasets



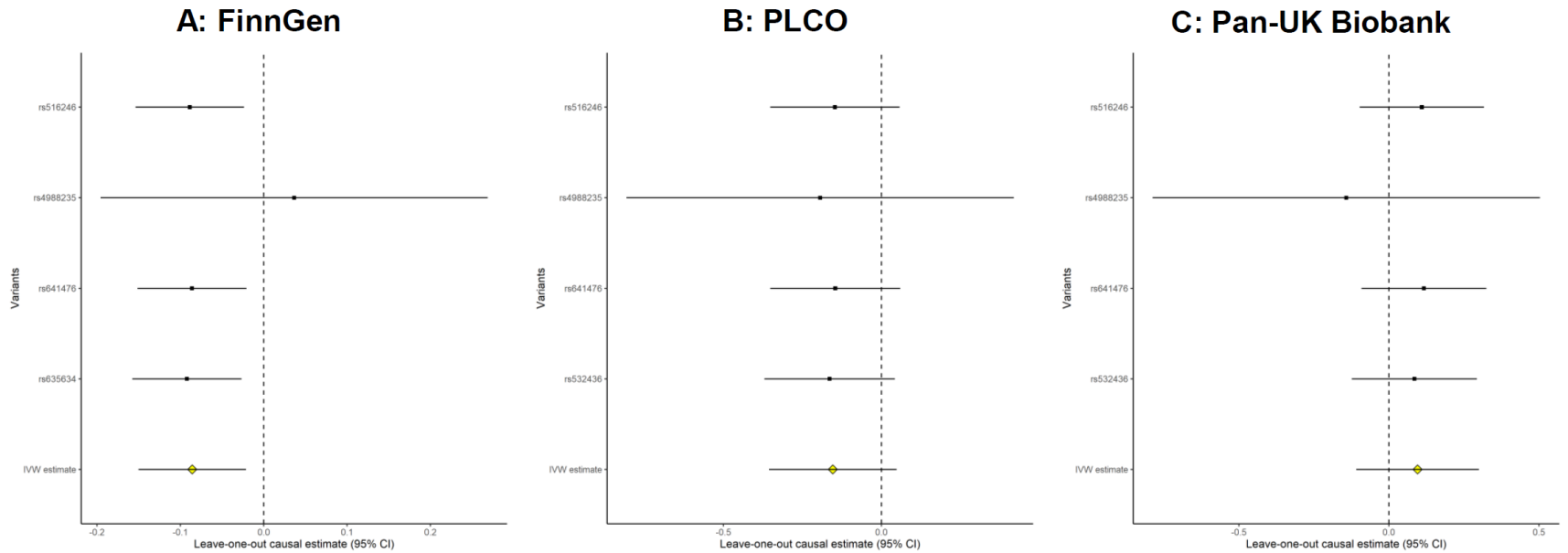
The x-axis represents the genetic association with LPH; the y-axis represents the genetic association with risk of rectal cancer. A: FinnGen dataset using IVW method; B: PLCO dataset using IVW method; C: Pan-UK Biobank dataset using IVW method; D: FinnGen dataset using MR-Egger method; E: PLCO dataset using MR-Egger method; F: Pan-UK Biobank dataset using MR-Egger method. IVW, inverse-variance weighted; MR: Mendelian Randomization; LPH, lactase-phlorizin hydrolase.

Figure 6-10. Forest plots of the IVW estimate on the association between genetically predicted LPH levels and rectal cancer risk for each genetic instrument in the FinnGen, PLCO, and Pan-UK Biobank datasets



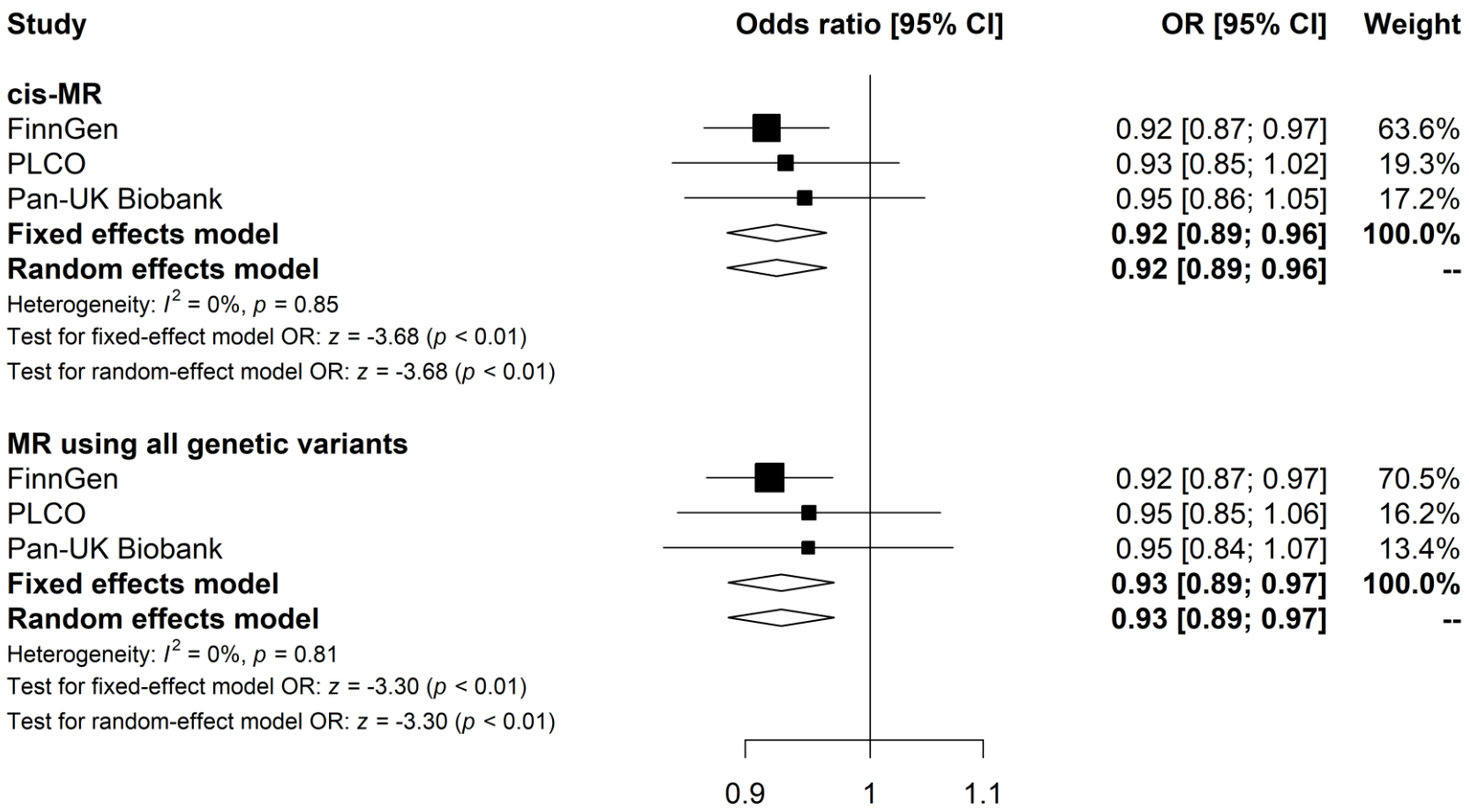
The x-axis represents the IVW causal estimate with its 95% CI; the y-axis represents genetic variant. A: FinnGen dataset; B: PLCO dataset; C: Pan-UK Biobank dataset. IVW, inverse-variance weighted; LPH, lactase-phlorizin hydrolase; CI: confidence interval

Figure 6-11. Leave-one-out analyses for the MR analysis on LPH levels and rectal cancer risk in the FinnGen, PLCO, and Pan-UK Biobank datasets



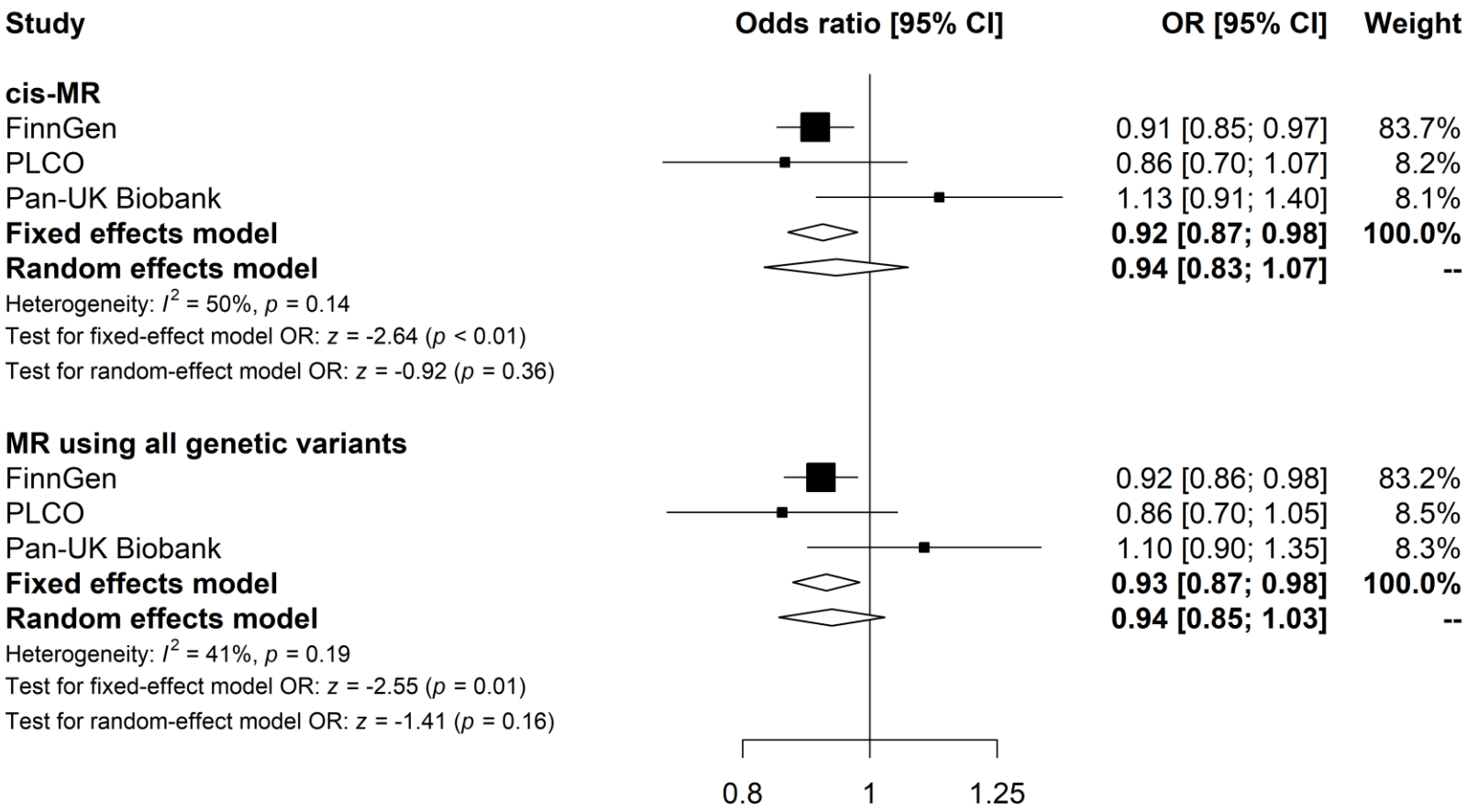
The x-axis represents the IVW estimate after removing the corresponding SNP, the Y-axis represents which genetic variant was removed from the MR analysis. MR, Mendelian Randomization; LPH, lactase-phlorizin hydrolase; CRC, colorectal cancer; IVW: inverse variance weighted; SNP, single nucleotide polymorphism; CI, confidence interval.

Figure 6-12. Meta-analysis results for the association of elevated LPH levels with colon cancer risk using *cis*-MR and MR using all genetic variants



Forest plots show results from *cis*-MR and MR using all genetic variants. Squares represent study-specific MR estimates. Diamonds represent meta-analyzed MR estimates using fixed and random effects models. LPH, lactase-phlorizin hydrolase; MR, Mendelian Randomization

Figure 6-13. Meta-analysis results for the association of elevated LPH levels with rectal cancer risk using *cis*-MR and MR using all genetic variants



Forest plots show results from *cis*-MR and MR using all genetic variants. Squares represent study-specific MR estimates. Diamonds represent meta-analyzed MR estimates using fixed and random effects models. LPH, lactase-phlorizin hydrolase; MR, Mendelian Randomization

Figure 7-1. Schematic overview of the study design

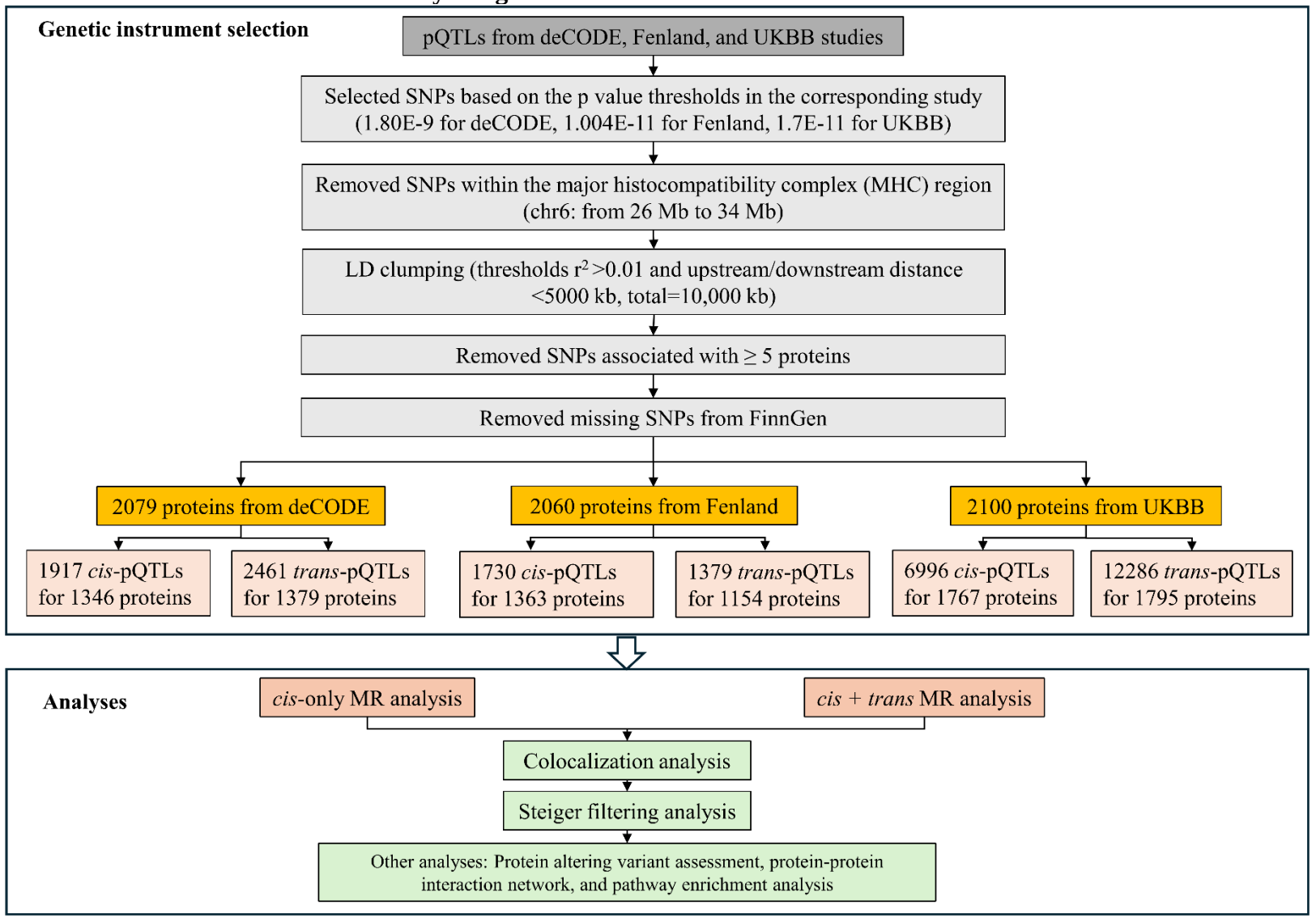
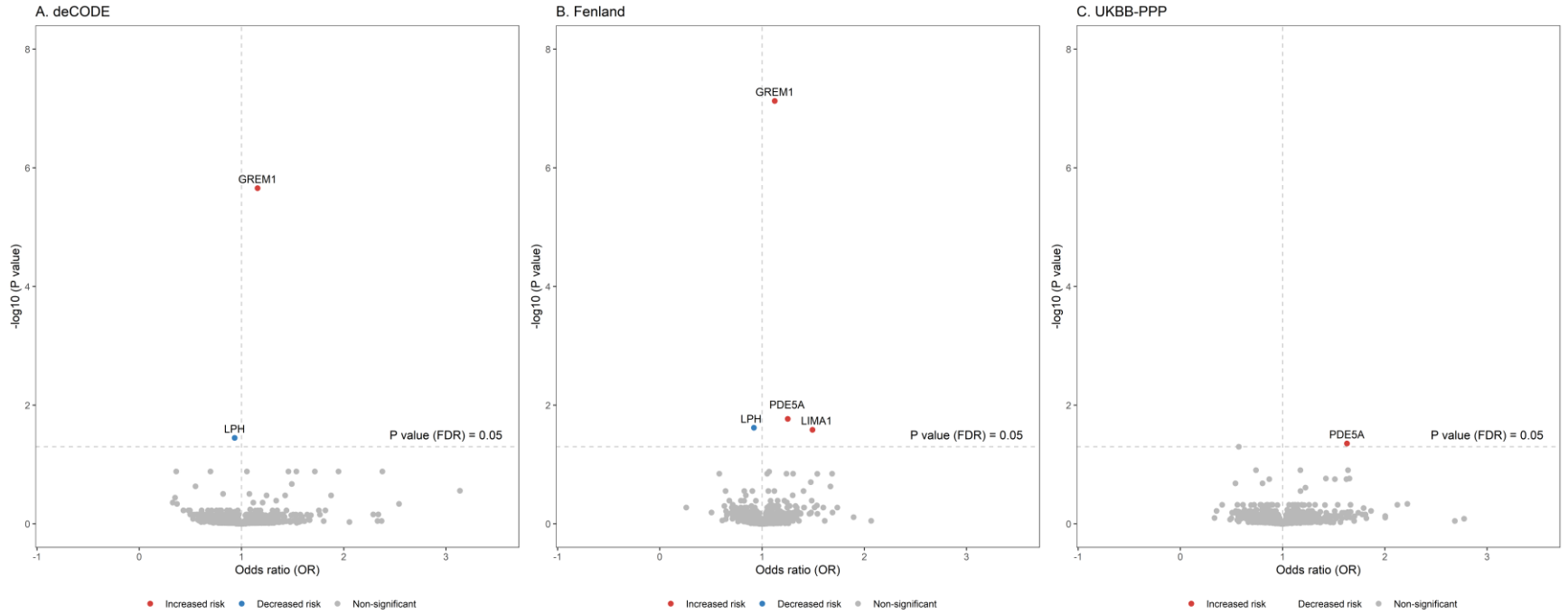
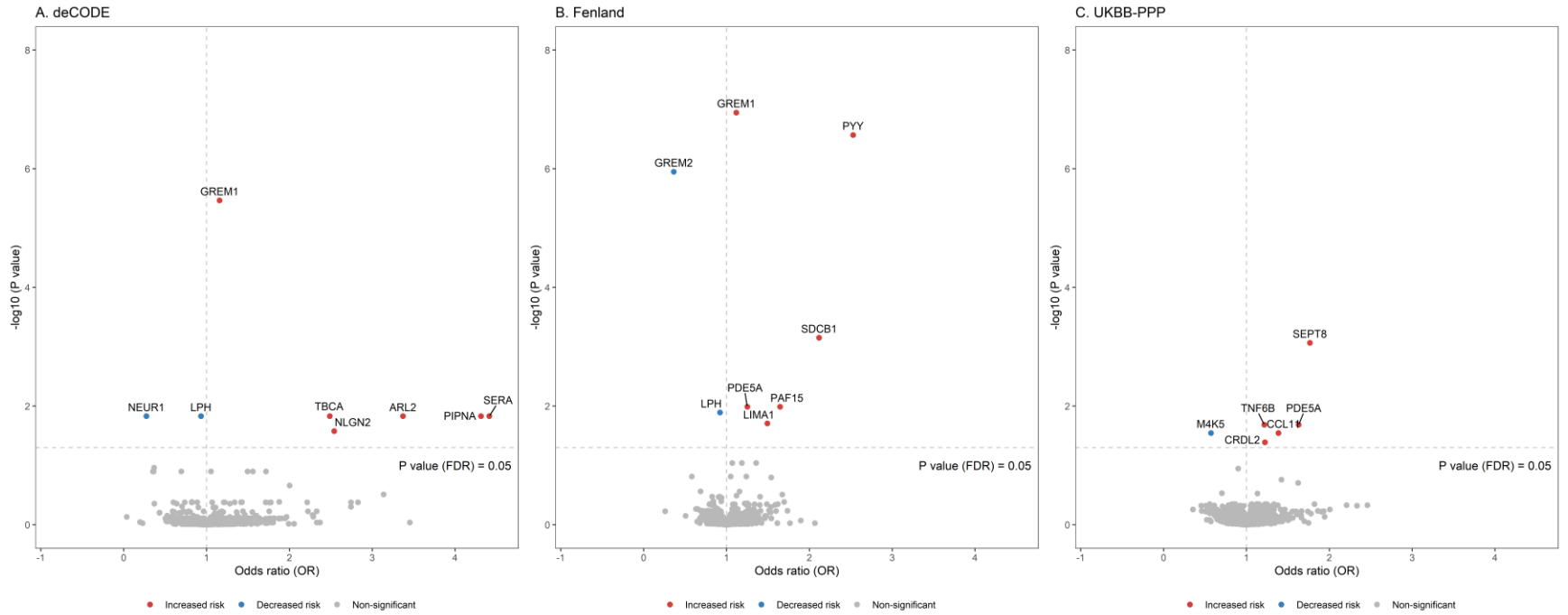


Figure 7-2. Volcano plot of the *cis*-only MR analyses for the association between plasma proteins and CRC risk



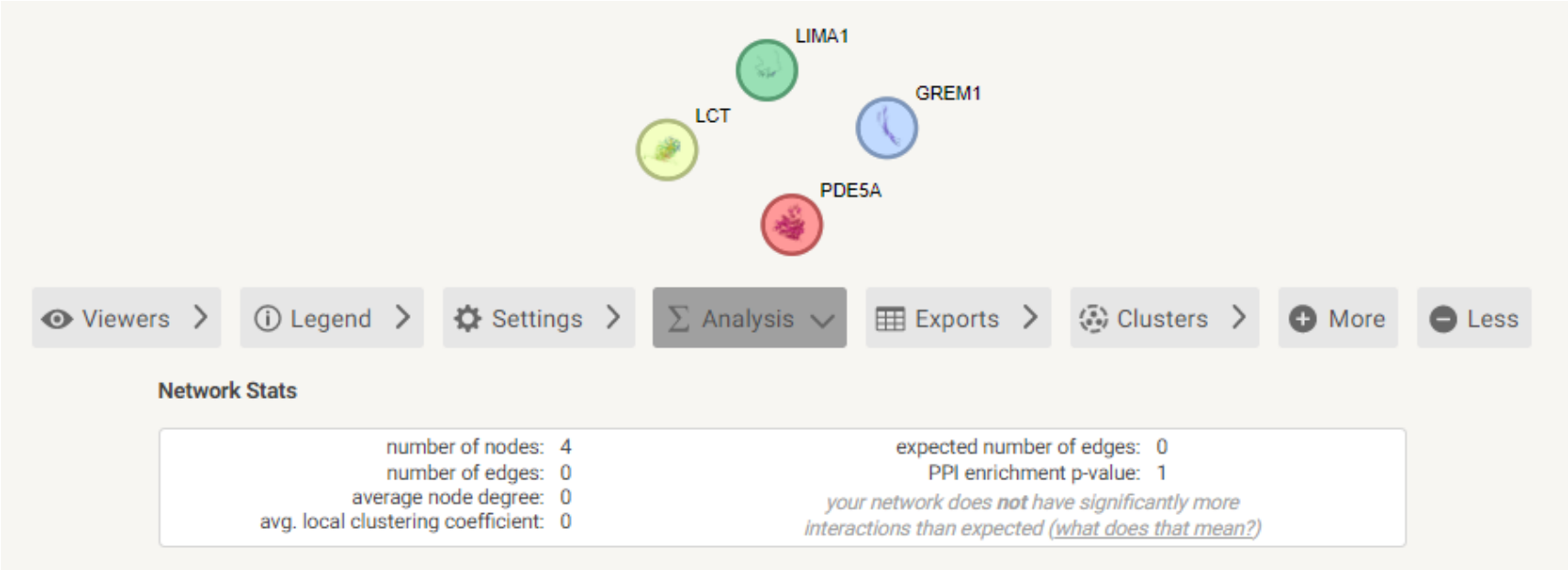
P values were FDR corrected. The horizontal line represents the threshold of P value (FDR) = 0.05.
A. Association between plasma proteins and CRC risk using *cis*-pQTLs from deCODE study; B. Association between plasma proteins and CRC risk using *cis*-pQTLs from Fenland study; C. Association between plasma proteins and CRC risk using *cis*-pQTLs from UKBB-PPP study.
Abbreviations: MR, Mendelian Randomization; CRC, colorectal cancer; UKBB-PPP, The UK Biobank Pharma Proteomics Project; GREM1, Gremlin-1; LPH, Lactase/phlorizin hydrolase; PDE5A, cGMP-specific 3',5'-cyclic phosphodiesterase; LIMA1, LIM domain and actin-binding protein 1; pQTL, protein quantitative trait loci.

Figure 7-3. Volcano plot of the MR analyses with all pQTLs (*cis+trans*) for the association between plasma proteins and CRC risk



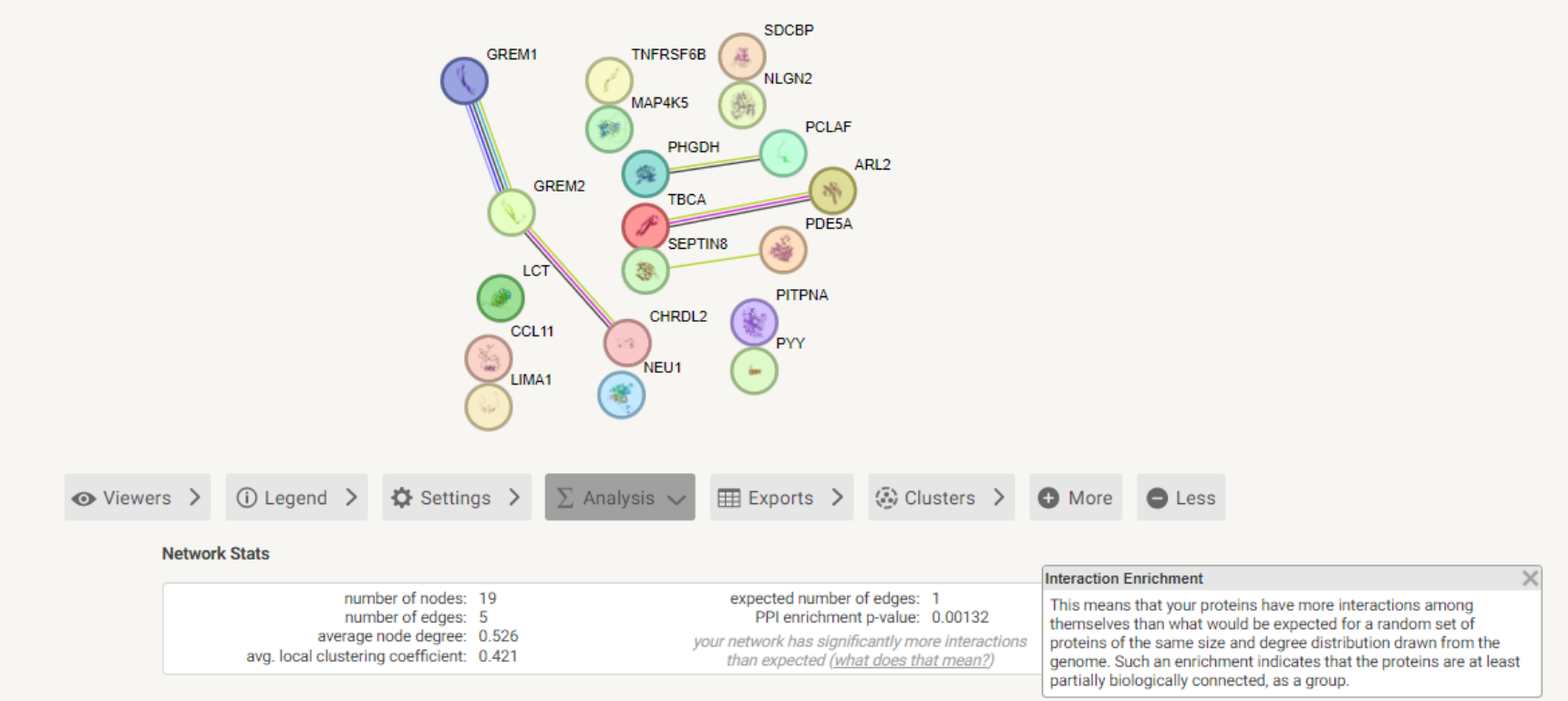
P values were FDR corrected. The horizontal line represents the threshold of P value (FDR) = 0.05.
A. Association between plasma proteins and CRC risk using all pQTLs from deCODE study; B. Association between plasma proteins and CRC risk using all pQTLs from Fenland study; C. Association between plasma proteins and CRC risk using all pQTLs from UKBB-PPP study.
Abbreviations: MR, Mendelian Randomization; pQTL, protein quantitative trait loci; CRC, colorectal cancer; UKBB-PPP, The UK Biobank Pharma Proteomics Project; GREM1, Gremlin-1; LPH, Lactase/phlorizin hydrolase; PDE5A, cGMP-specific 3',5'-cyclic phosphodiesterase; LIMA1, LIM domain and actin-binding protein 1; TBCA, Tubulin-specific chaperone A; PIPNA, Phosphatidylinositol transfer protein alpha isoform; SERA, D-3-phosphoglycerate dehydrogenase; ARL2, ADP-ribosylation factor-like protein 2; NEUR1, Sialidase-1; NLGN2, Neuroligin-2; PYY, Peptide YY; GREM2, Gremlin-2; SDCB1, Syntenin-1; PAF15, PCNA-associated factor; SEPT8, Septin-8; TNF6B, Tumor necrosis factor receptor superfamily member 6B; CCL11, Eotaxin; M4K5, Mitogen-activated protein kinase kinase kinase kinase 5; CRDL2, Chordin-like protein 2

Figure 7-4. PPI network for *cis*-prioritized proteins associated with CRC risk



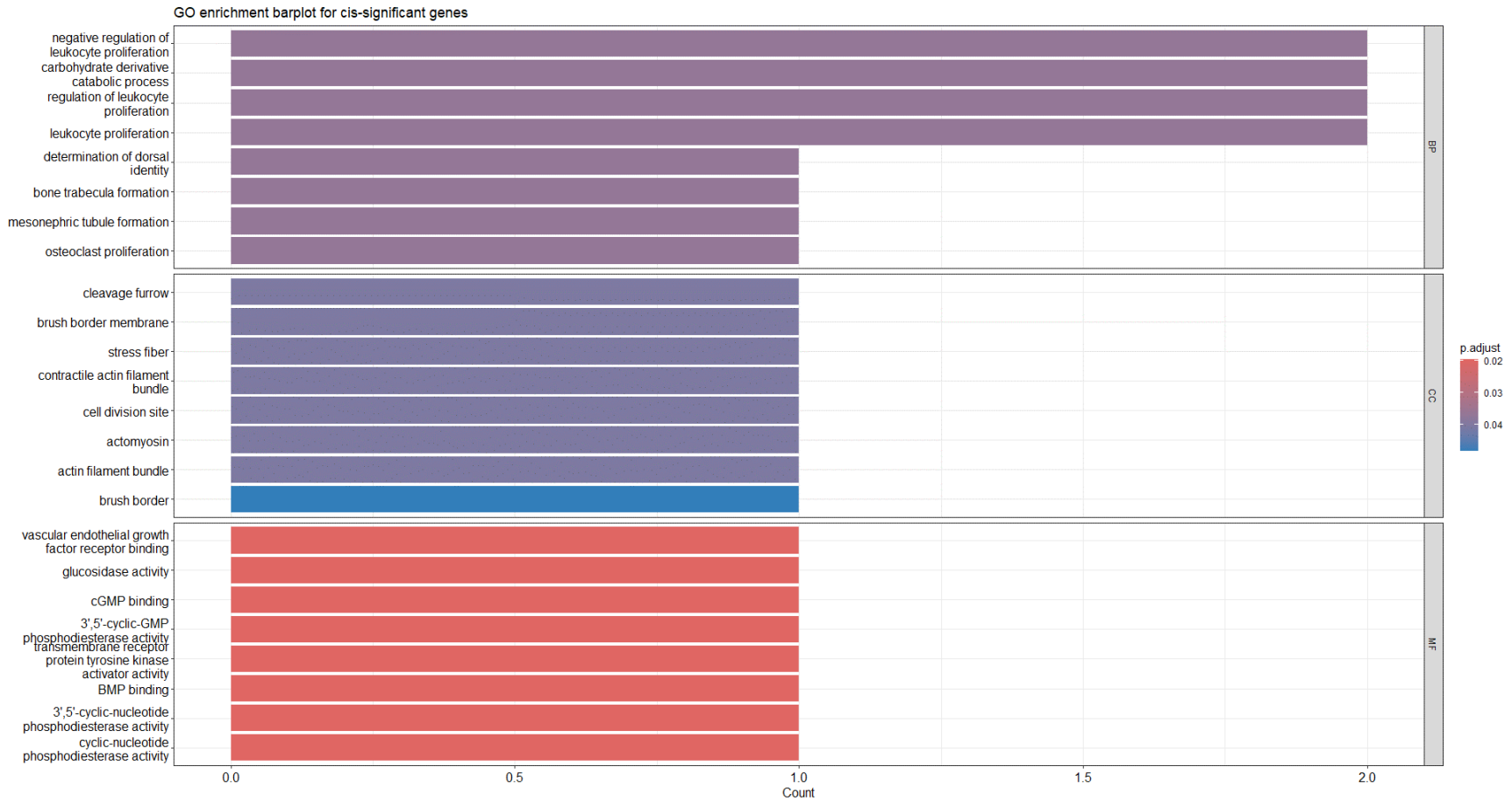
Abbreviations: MR, Mendelian Randomization; CRC, colorectal cancer.

Figure 7-5. PPI network for all MR significant proteins associated with CRC risk



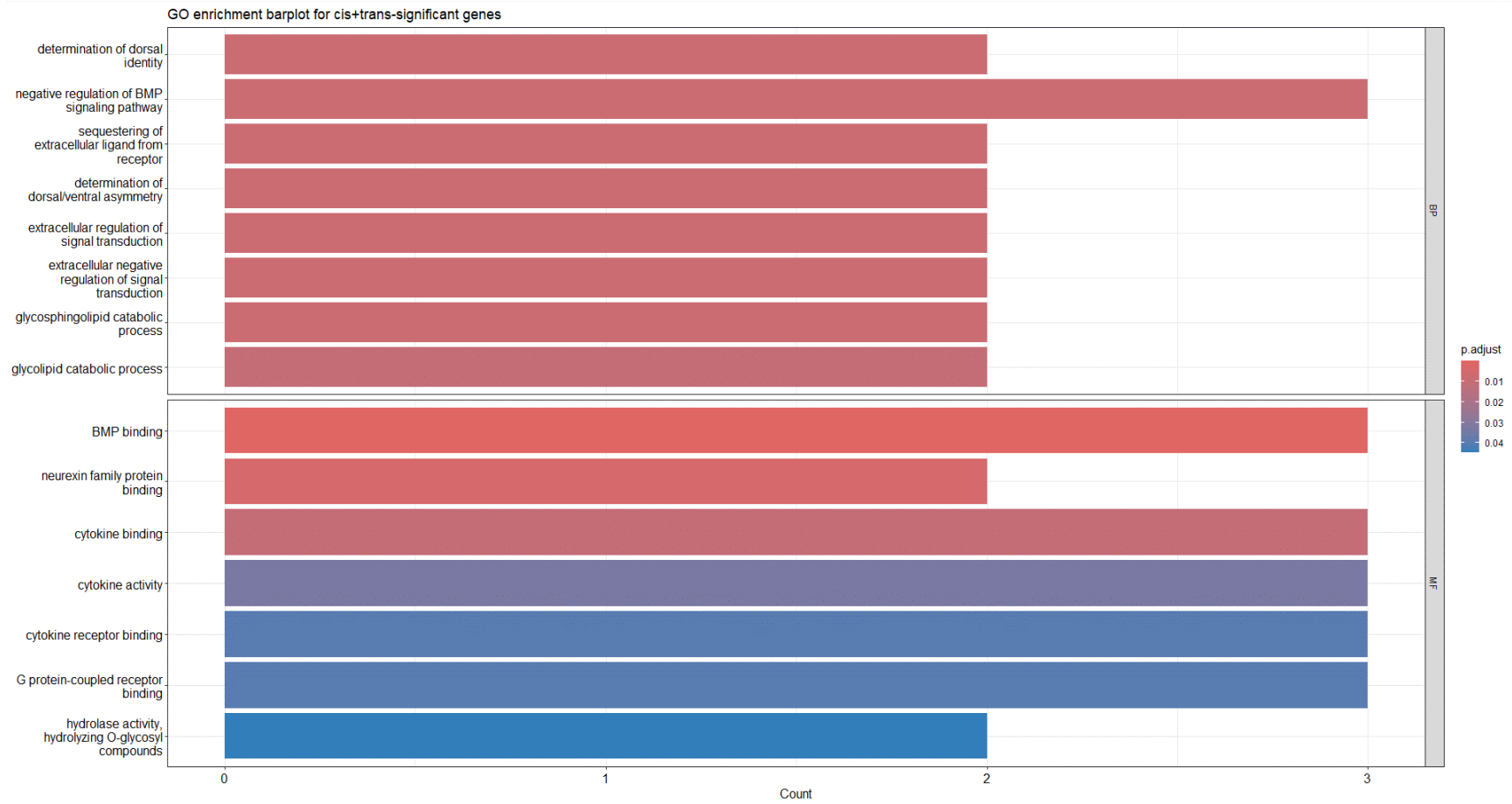
Abbreviations: MR, Mendelian Randomization; CRC, colorectal cancer.

Figure 7-6. GO enrichment analysis for *cis*-MR prioritized proteins



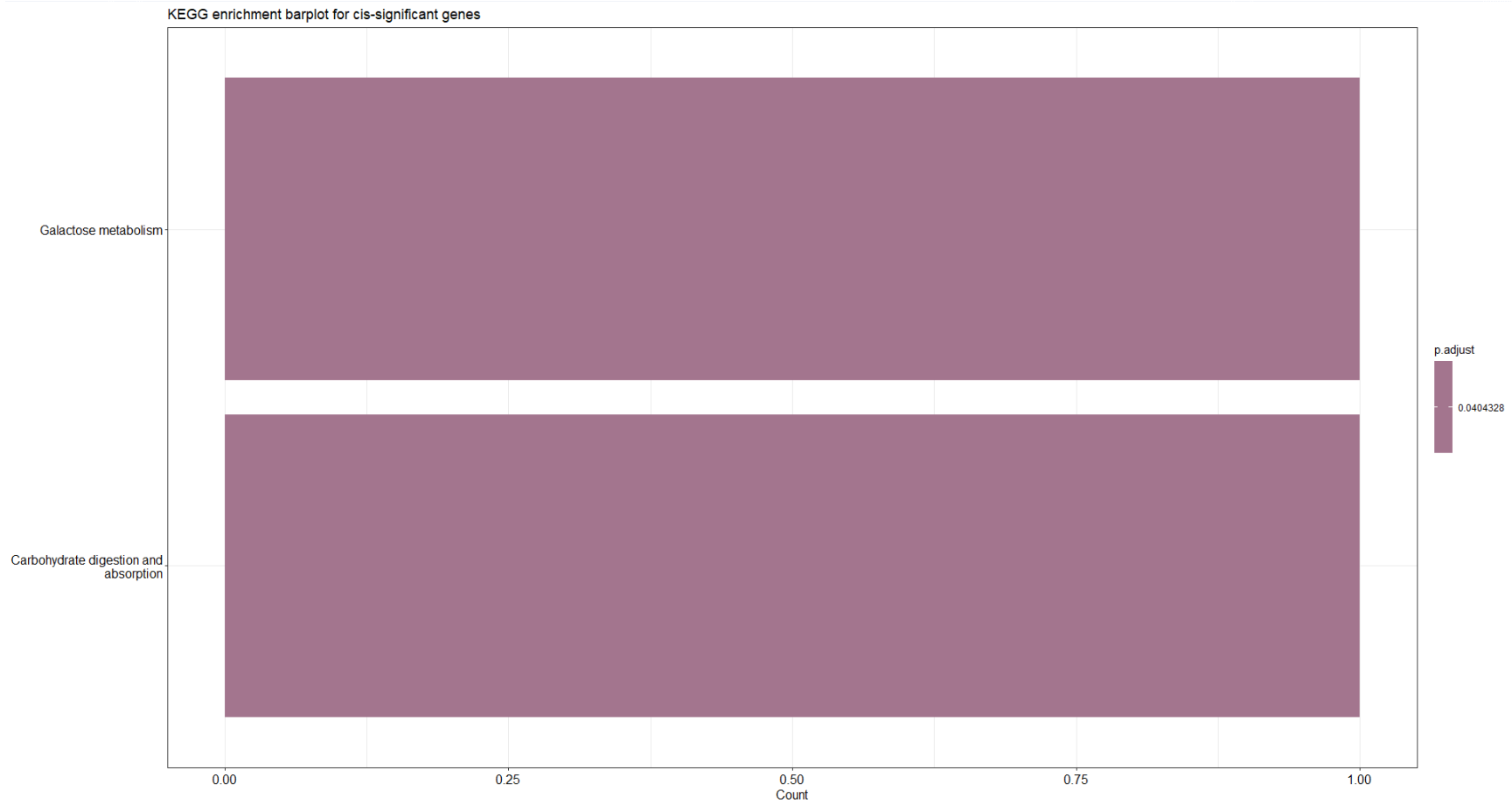
Abbreviations: GO, Gene Ontology; MR, Mendelian Randomization.

Figure 7-7. GO enrichment analysis for *cis+trans* MR prioritized proteins



Abbreviations: GO, Gene Ontology; MR, Mendelian Randomization.

Figure 7-8. KEGG enrichment analysis for *cis*-MR prioritized proteins



Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; MR, Mendelian Randomization.

REFERENCES

1. Keum, N. and E. Giovannucci, *Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies.* Nature reviews Gastroenterology & hepatology, 2019. **16**(12): p. 713-732.
2. Ferlay J, E.M., Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. *Global Cancer Observatory: Cancer Today (version 1.1)*. 2024 [cited 2024 March 14]; Available from: https://gco.iarc.who.int/today.
3. Dekker, E., et al., *Colorectal cancer.* Lancet, 2019. **394**(10207): p. 1467-1480.
4. Xi, Y. and P. Xu, *Global colorectal cancer burden in 2020 and projections to 2040.* Translational oncology, 2021. **14**(10): p. 101174.
5. Siegel, R.L., A.N. Giaquinto, and A. Jemal, *Cancer statistics, 2024.* CA Cancer J Clin, 2024. **74**(1): p. 12-49.
6. Meissner, H.I., et al., *Patterns of colorectal cancer screening uptake among men and women in the United States.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(2): p. 389-94.
7. Murphy, G., et al., *Sex disparities in colorectal cancer incidence by anatomic subsite, race and age.* Int J Cancer, 2011. **128**(7): p. 1668-75.
8. Siegel, R.L., et al., *Colorectal cancer statistics, 2023.* CA: a cancer journal for clinicians, 2023. **73**(3): p. 233-254.
9. Murphy, C.C., et al., *Decrease in Incidence of Colorectal Cancer Among Individuals 50 Years or Older After Recommendations for Population-based Screening.* Clin Gastroenterol Hepatol, 2017. **15**(6): p. 903-909.e6.
10. Mauri, G., et al., *Early-onset colorectal cancer in young individuals.* Molecular oncology, 2019. **13**(2): p. 109-131.

11. Arnold, M., et al., *Global patterns and trends in colorectal cancer incidence and mortality.* Gut, 2017. **66**(4): p. 683-691.

12. Wong, M.C., et al., *Differences in incidence and mortality trends of colorectal cancer worldwide based on sex, age, and anatomic location.* Clinical Gastroenterology and Hepatology, 2021. **19**(5): p. 955-966. e61.

13. Sawicki, T., et al., *A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis.* Cancers, 2021. **13**(9): p. 2025.

14. Miller, K.D., et al., *Cancer treatment and survivorship statistics, 2022.* CA: a cancer journal for clinicians, 2022. **72**(5): p. 409-436.

15. Murphy, C.C., et al., *Patterns of Colorectal Cancer Care in the United States: 1990-2010.* J Natl Cancer Inst, 2015. **107**(10).

16. Chu, K.C., et al., *Temporal patterns in colorectal cancer incidence, survival, and mortality from 1950 through 1990.* J Natl Cancer Inst, 1994. **86**(13): p. 997-1006.

17. Gross, C.P., et al., *Relation between Medicare screening reimbursement and stage at diagnosis for older patients with colon cancer.* Jama, 2006. **296**(23): p. 2815-22.

18. Kim, T.J., et al., *Long-Term Outcome and Prognostic Factors of Sporadic Colorectal Cancer in Young Patients: A Large Institutional-Based Retrospective Study.* Medicine (Baltimore), 2016. **95**(19): p. e3641.

19. Dow, L.E., et al., *Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer.* Cell, 2015. **161**(7): p. 1539-1552.

20. Bardelli, A., et al., *Mutational analysis of the tyrosine kinome in colorectal cancers.* Science, 2003. **300**(5621): p. 949.

21. Parsons, D.W., et al., *Colorectal cancer: mutations in a signalling pathway.* Nature, 2005. **436**(7052): p. 792.

22. Armaghany, T., et al., *Genetic alterations in colorectal cancer.* Gastrointest Cancer Res, 2012. **5**(1): p. 19-27.

23. Pino, M.S. and D.C. Chung, *The chromosomal instability pathway in colon cancer.* Gastroenterology, 2010. **138**(6): p. 2059-72.

24. Leggett, B. and V. Whitehall, *Role of the serrated pathway in colorectal cancer pathogenesis.* Gastroenterology, 2010. **138**(6): p. 2088-100.

25. Kedrin, D. and M.K. Gala, *Genetics of the serrated pathway to colorectal cancer.* Clin Transl Gastroenterol, 2015. **6**(4): p. e84.

26. Yang, S., et al., *BRAF and KRAS Mutations in hyperplastic polyps and serrated adenomas of the colorectum: relationship to histology and CpG island methylation status.* Am J Surg Pathol, 2004. **28**(11): p. 1452-9.

27. Kim, K.M., et al., *Molecular features of colorectal hyperplastic polyps and sessile serrated adenoma/polyps from Korea.* Am J Surg Pathol, 2011. **35**(9): p. 1274-86.

28. O'Brien, M.J., et al., *Comparison of microsatellite instability, CpG island methylation phenotype, BRAF and KRAS status in serrated polyps and traditional adenomas indicates separate pathways to distinct colorectal carcinoma end points.* Am J Surg Pathol, 2006. **30**(12): p. 1491-501.

29. Itzkowitz, S.H. and X. Yio, *Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation.* Am J Physiol Gastrointest Liver Physiol, 2004. **287**(1): p. G7-17.

30. Aust, D.E., et al., *The APC/β-catenin pathway in ulcerative colitis–related colorectal carcinomas: A mutational analysis.* Cancer, 2002. **94**(5): p. 1421-1427.

31. Brentnall, T.A., et al., *Mutations in the p53 gene: an early marker of neoplastic progression in ulcerative colitis.* Gastroenterology, 1994. **107**(2): p. 369-78.

32. Burmer, G.C., et al., *Neoplastic progression in ulcerative colitis: histology, DNA content, and loss of a p53 allele.* Gastroenterology, 1992. **103**(5): p. 1602-10.

33. Redston, M.S., et al., *Common occurrence of APC and K-ras gene mutations in the spectrum of colitis-associated neoplasias.* Gastroenterology, 1995. **108**(2): p. 383-92.

34. Tarmin, L., et al., *Adenomatous polyposis coli gene mutations in ulcerative colitis-associated dysplasias and cancers versus sporadic colon neoplasms.* Cancer Res, 1995. **55**(10): p. 2035-8.

35. Hussain, S.P., L.J. Hofseth, and C.C. Harris, *Radical causes of cancer.* Nat Rev Cancer, 2003. **3**(4): p. 276-85.

36. Jess, T., C. Rungoe, and L. Peyrin-Biroulet, *Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies.* Clin Gastroenterol Hepatol, 2012. **10**(6): p. 639-45.

37. Kuipers, E.J., et al., *Colorectal cancer.* Nat Rev Dis Primers, 2015. **1**: p. 15065.

38. Thanikachalam, K. and G. Khan, *Colorectal Cancer and Nutrition.* Nutrients, 2019. **11**(1).

39. Dinarvand, P., et al., *Familial Adenomatous Polyposis Syndrome: An Update and Review of Extraintestinal Manifestations.* Arch Pathol Lab Med, 2019. **143**(11): p. 1382-1398.

40. Mármol, I., et al., *Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer.* Int J Mol Sci, 2017. **18**(1).

41. Johns, L.E. and R.S. Houlston, *A systematic review and meta-analysis of familial colorectal cancer risk.* Am J Gastroenterol, 2001. **96**(10): p. 2992-3003.

42. Eaden, J.A., K.R. Abrams, and J.F. Mayberry, *The risk of colorectal cancer in ulcerative colitis: a meta-analysis.* Gut, 2001. **48**(4): p. 526-35.

43. Canavan, C., K.R. Abrams, and J. Mayberry, *Meta-analysis: colorectal and small bowel cancer risk in patients with Crohn's disease.* Aliment Pharmacol Ther, 2006. **23**(8): p. 1097-104.

44. Butterworth, A.S., J.P. Higgins, and P. Pharoah, *Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis.* Eur J Cancer, 2006. **42**(2): p. 216-27.

45. Huyghe, J.R., et al., *Discovery of common and rare genetic risk variants for colorectal cancer.* Nat Genet, 2019. **51**(1): p. 76-87.

46. Law, P.J., et al., *Association analyses identify 31 new risk loci for colorectal cancer susceptibility.* Nat Commun, 2019. **10**(1): p. 2154.

47. Jasperson, K.W., et al., *Hereditary and familial colon cancer.* Gastroenterology, 2010. **138**(6): p. 2044-58.

48. Research, W.C.R.F.I.A.I.f.C., *Continuous Update Project Report: Diet, Nutrition, Physical Activity and Colorectal Cancer*. 2017, World Cancer Research Fund/American Institute for Cancer Research Washington, DC.

49. Botteri, E., et al., *Smoking and colorectal cancer: a meta-analysis.* Jama, 2008. **300**(23): p. 2765-78.

50. Cai, S., et al., *Alcohol drinking and the risk of colorectal cancer death: a meta-analysis.* Eur J Cancer Prev, 2014. **23**(6): p. 532-9.

51. Kyrgiou, M., et al., *Adiposity and cancer at major anatomical sites: umbrella review of the literature.* Bmj, 2017. **356**: p. j477.

52. Tabung, F.K., L.S. Brown, and T.T. Fung, *Dietary Patterns and Colorectal Cancer Risk: A Review of 17 Years of Evidence (2000-2016).* Curr Colorectal Cancer Rep, 2017. **13**(6): p. 440-454.

53. Garcia-Larsen, V., et al., *Dietary patterns derived from principal component analysis (PCA) and risk of colorectal cancer: a systematic review and meta-analysis.* Eur J Clin Nutr, 2019. **73**(3): p. 366-386.

54. Keum, N., et al., *Calcium intake and colorectal cancer risk: dose-response meta-analysis of prospective observational studies.* Int J Cancer, 2014. **135**(8): p. 1940-8.

55. Bristow, S.M., et al., *Calcium supplements and cancer risk: a meta-analysis of randomised controlled trials.* Br J Nutr, 2013. **110**(8): p. 1384-93.

56. Lamprecht, S.A. and M. Lipkin, *Chemoprevention of colon cancer by calcium, vitamin D and folate: molecular mechanisms.* Nat Rev Cancer, 2003. **3**(8): p. 601-14.

57. Tennakoon, S., A. Aggarwal, and E. Kállay, *The calcium-sensing receptor and the hallmarks of cancer.* Biochim Biophys Acta, 2016. **1863**(6 Pt B): p. 1398-407.

58. McCullough, M.L., et al., *Calcium, vitamin D, dairy products, and risk of colorectal cancer in the Cancer Prevention Study II Nutrition Cohort (United States).* Cancer Causes Control, 2003. **14**(1): p. 1-12.

59. Järvinen, R., et al., *Prospective study on milk products, calcium and cancers of the colon and rectum.* Eur J Clin Nutr, 2001. **55**(11): p. 1000-7.

60. Cho, E., et al., *Dairy foods, calcium, and colorectal cancer: a pooled analysis of 10 cohort studies.* J Natl Cancer Inst, 2004. **96**(13): p. 1015-22.

61. Norat, T. and E. Riboli, *Dairy products and colorectal cancer. A review of possible mechanisms and epidemiological evidence.* European Journal of Clinical Nutrition, 2003. **57**(1): p. 1-17.

62. Ursin, G., et al., *Milk consumption and cancer incidence: a Norwegian prospective study.* Br J Cancer, 1990. **61**(3): p. 454-9.

63. Kampman, E., et al., *Calcium, vitamin D, dairy foods, and the occurrence of colorectal adenomas among men and women in two prospective studies.* Am J Epidemiol, 1994. **139**(1): p. 16-29.

64. Park, Y., et al., *Dairy food, calcium, and risk of cancer in the NIH-AARP Diet and Health Study.* Arch Intern Med, 2009. **169**(4): p. 391-401.

65. Murphy, N., et al., *Consumption of dairy products and colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC).* PLoS One, 2013. **8**(9): p. e72715.

66. Barrubés, L., et al., *Association Between Dairy Product Consumption and Colorectal Cancer Risk in Adults: A Systematic Review and Meta-Analysis of Epidemiologic Studies.* Adv Nutr, 2019. **10**(suppl_2): p. S190-s211.

67. Martinez-Useros, J. and J. Garcia-Foncillas, *Obesity and colorectal cancer: molecular features of adipose tissue.* J Transl Med, 2016. **14**: p. 21.

68. Giovannucci, E. and M.E. Martínez, *Tobacco, colorectal cancer, and adenomas: a review of the evidence.* J Natl Cancer Inst, 1996. **88**(23): p. 1717-30.

69. Johnson, C.M., et al., *Meta-analyses of colorectal cancer risk factors.* Cancer Causes Control, 2013. **24**(6): p. 1207-22.

70. Cancer, I.A.f.R.o., *Agents classified by the IARC monographs, volumes 1–122 https://monographs. iarc. fr/agents-classified-by-the-iarc.* Accessed September, 2018. **1**.

71. Choi, Y.J., S.K. Myung, and J.H. Lee, *Light Alcohol Drinking and Risk of Cancer: A Meta-Analysis of Cohort Studies.* Cancer Res Treat, 2018. **50**(2): p. 474-487.

72. Renehan, A.G., et al., *Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies.* Lancet, 2008. **371**(9612): p. 569-78.

73. Dong, Y., et al., *Abdominal obesity and colorectal cancer risk: systematic review and meta-analysis of prospective studies.* Biosci Rep, 2017. **37**(6).

74. Song, M., et al., *Long-term status and change of body fat distribution, and risk of colorectal cancer: a prospective cohort study.* Int J Epidemiol, 2016. **45**(3): p. 871-83.

75. Moore, L.L., et al., *BMI and waist circumference as predictors of lifetime colon cancer risk in Framingham Study adults.* Int J Obes Relat Metab Disord, 2004. **28**(4): p. 559-67.

76. Thrift, A.P., et al., *Mendelian Randomization Study of Body Mass Index and Colorectal Cancer Risk.* Cancer Epidemiol Biomarkers Prev, 2015. **24**(7): p. 1024-31.

77. Song, M., W.S. Garrett, and A.T. Chan, *Nutrients, foods, and colorectal cancer prevention.* Gastroenterology, 2015. **148**(6): p. 1244-60.e16.

78. Ahn, J., et al., *Human gut microbiome and risk for colorectal cancer.* J Natl Cancer Inst, 2013. **105**(24): p. 1907-11.

79. Zackular, J.P., et al., *The gut microbiome modulates colon tumorigenesis.* mBio, 2013. **4**(6): p. e00692-13.

80. Warren, R.L., et al., *Co-occurrence of anaerobic bacteria in colorectal carcinomas.* Microbiome, 2013. **1**(1): p. 16.

81. Wang, T., et al., *Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers.* Isme j, 2012. **6**(2): p. 320-9.

82. Sanapareddy, N., et al., *Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans.* Isme j, 2012. **6**(10): p. 1858-68.

83. Shen, X.J., et al., *Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas.* Gut Microbes, 2010. **1**(3): p. 138-47.

84. Wu, G.D., et al., *Linking long-term dietary patterns with gut microbial enterotypes.* Science, 2011. **334**(6052): p. 105-8.

85. Claesson, M.J., et al., *Gut microbiota composition correlates with diet and health in the elderly.* Nature, 2012. **488**(7410): p. 178-84.

86. David, L.A., et al., *Diet rapidly and reproducibly alters the human gut microbiome.* Nature, 2014. **505**(7484): p. 559-63.

87. De Filippo, C., et al., *Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa.* Proc Natl Acad Sci U S A, 2010. **107**(33): p. 14691-6.

88. Turnbaugh, P.J., et al., *The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice.* Sci Transl Med, 2009. **1**(6): p. 6ra14.

89. Veettil, S.K., et al., *Role of diet in colorectal cancer incidence: umbrella review of meta-analyses of prospective observational studies.* JAMA network open, 2021. **4**(2): p. e2037341-e2037341.

90. Clinton, S.K., E.L. Giovannucci, and S.D. Hursting, *The world cancer research fund/American institute for cancer research third expert report on diet, nutrition, physical*

activity, and cancer: impact and future directions. The Journal of nutrition, 2020. **150**(4): p. 663-671.

91. Ursin, G., et al., *Milk consumption and cancer incidence: a Norwegian prospective study.* British journal of cancer, 1990. **61**(3): p. 456-459.

92. Kampman, E., et al., *Fermented dairy products, calcium, and colorectal cancer in The Netherlands Cohort Study.* Cancer Research, 1994. **54**(12): p. 3186-3190.

93. Gaard, M., S. Tretli, and E. Løken, *Dietary factors and risk of colon cancer: a prospective study of 50,535 young Norwegian men and women.* European journal of cancer prevention, 1996: p. 445-454.

94. Keamey, J., et al., *Calcium, vitamin D, and dairy foods and the occurrence of colon cancer in men.* American journal of epidemiology, 1996. **143**(9): p. 907-917.

95. Martinez, M.E., et al., *Calcium, vitamin D, and the occurrence of colorectal cancer among women.* JNCI: Journal of the National Cancer Institute, 1996. **88**(19): p. 1375-1382.

96. Kato, I., et al., *Prospective study of diet and female colorectal cancer: the New York University Women's Health Study.* 1997.

97. Singh, P.N. and G.E. Fraser, *Dietary risk factors for colon cancer in a low-risk population.* American journal of epidemiology, 1998. **148**(8): p. 761-774.

98. Pietinen, P., et al., *Diet and risk of colorectal cancer in a cohort of Finnish men.* Cancer causes & control, 1999. **10**: p. 387-396.

99. Järvinen, R., et al., *Prospective study on milk products, calcium and cancers of the colon and rectum.* European journal of clinical nutrition, 2001. **55**(11): p. 1000-1007.

100. Terry, P., et al., *Dietary calcium and vitamin D intake and risk of colorectal cancer: a prospective cohort study in women.* Nutrition and cancer, 2002. **43**(1): p. 39-46.

101. Wu, K., et al., *Calcium intake and risk of colon cancer in women and men.* Journal of the national cancer institute, 2002. **94**(6): p. 437-446.

102. McCullough, M.L., et al., *Calcium, vitamin D, dairy products, and risk of colorectal cancer in the Cancer Prevention Study II Nutrition Cohort (United States).* Cancer Causes & Control, 2003. **14**: p. 1-12.

103. Sanjoaquin, M.A., et al., *Nutrition, lifestyle and colorectal cancer incidence: a prospective investigation of 10 998 vegetarians and non-vegetarians in the United Kingdom.* British journal of cancer, 2004. **90**(1): p. 118-121.

104. Lin, J., et al., *Intakes of calcium and vitamin D and risk of colorectal cancer in women.* American journal of epidemiology, 2005. **161**(8): p. 755-764.

105. Larsson, S.C., L. Bergkvist, and A. Wolk, *High-fat dairy food and conjugated linoleic acid intakes in relation to colorectal cancer incidence in the Swedish Mammography Cohort.* The American journal of clinical nutrition, 2005. **82**(4): p. 894-900.

106. Larsson, S.C., et al., *Calcium and dairy food intakes are inversely associated with colorectal cancer risk in the Cohort of Swedish Men.* The American journal of clinical nutrition, 2006. **83**(3): p. 667-673.

107. Park, S.-Y., et al., *Calcium and vitamin D intake and risk of colorectal cancer: the Multiethnic Cohort Study.* American journal of epidemiology, 2007. **165**(7): p. 784-793.

108. Butler, L., et al., *Prospective study of dietary patterns and colorectal cancer among Singapore Chinese.* British journal of cancer, 2008. **99**(9): p. 1511-1516.

109. Lee, S.-A., et al., *Animal origin foods and colorectal cancer risk: a report from the Shanghai Women's Health Study.* Nutrition and cancer, 2009. **61**(2): p. 194-205.

110. Park, Y., et al., *Dairy food, calcium, and risk of cancer in the NIH-AARP Diet and Health Study.* Archives of internal medicine, 2009. **169**(4): p. 391-401.

111. Simons, C.C., et al., *Fluid intake and colorectal cancer risk in the Netherlands Cohort Study.* Nutrition and cancer, 2010. **62**(3): p. 307-321.

112. Ruder, E.H., et al., *Adolescent and mid-life diet: risk of colorectal cancer in the NIH-AARP Diet and Health Study.* The American journal of clinical nutrition, 2011. **94**(6): p. 1607-1619.

113. Murphy, N., et al., *Consumption of dairy products and colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC).* PloS one, 2013. **8**(9): p. e72715.

114. Tantamango-Bartley, Y., et al., *Independent associations of dairy and calcium intakes with colorectal cancers in the Adventist Health Study-2 cohort.* Public health nutrition, 2017. **20**(14): p. 2577-2586.

115. Bakken, T., et al., *Milk and risk of colorectal, colon and rectal cancer in the Norwegian Women and Cancer (NOWAC) Cohort Study.* British Journal of Nutrition, 2018. **119**(11): p. 1274-1285.

116. Vulcan, A., et al., *A colorectal cancer diet quality index is inversely associated with colorectal cancer in the Malmö diet and cancer study.* European Journal of Cancer Prevention, 2019. **28**(6): p. 463-471.

117. Um, C.Y., et al., *Associations of calcium, vitamin D, and dairy product intakes with colorectal cancer risk among older women: the Iowa women's health study.* Nutrition and cancer, 2019. **71**(5): p. 739-748.

118. Nilsson, L.M., et al., *Dairy products and cancer risk in a Northern Sweden population.* Nutrition and cancer, 2020. **72**(3): p. 409-420.

119. Jin, S., Y. Kim, and Y. Je, *Dairy consumption and risks of colorectal cancer incidence and mortality: a meta-analysis of prospective cohort studies.* Cancer Epidemiology, Biomarkers & Prevention, 2020. **29**(11): p. 2309-2322.

120. Siegel, R.L., et al., *Colorectal cancer statistics, 2020.* CA: a cancer journal for clinicians, 2020. **70**(3): p. 145-164.

121. Cruz-Pierard, S.M., T. Nestares, and F.J. Amaro-Gahete, *Vitamin D and Calcium as Key Potential Factors Related to Colorectal Cancer Prevention and Treatment: A Systematic Review.* Nutrients, 2022. **14**(22).

122. Yang, W., et al., *Calcium intake and risk of colorectal cancer according to expression status of calcium-sensing receptor (CASR).* Gut, 2018. **67**(8): p. 1475-1483.

123. Fleet, J.C., et al., *Vitamin D and cancer: a review of molecular mechanisms.* Biochemical Journal, 2012. **441**(1): p. 61-76.

124. Arayici, M.E., Y. Basbinar, and H. Ellidokuz, *Vitamin D Intake, Serum 25-Hydroxyvitamin-D (25 (OH) D) Levels, and Cancer Risk: A Comprehensive Meta-Meta-Analysis Including Meta-Analyses of Randomized Controlled Trials and Observational Epidemiological Studies.* Nutrients, 2023. **15**(12): p. 2722.

125. Kozu, T., et al., *Effect of orally administered bovine lactoferrin on the growth of adenomatous colorectal polyps in a randomized, placebo-controlled clinical trial.* Cancer Prevention Research, 2009. **2**(11): p. 975-983.

126. Pufulete, M., *Intake of dairy products and risk of colorectal neoplasia.* Nutrition research reviews, 2008. **21**(1): p. 56-67.

127. Kolonel, L.N., et al., *A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics.* American journal of epidemiology, 2000. **151**(4): p. 346-357.

128. Stram, D.O., et al., *Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles.* American journal of epidemiology, 2000. **151**(4): p. 358-370.

129. Sharma, S., et al., *Extending a multiethnic food composition table to include standardized food group servings.* Journal of Food Composition and Analysis, 2003. **16**(4): p. 485-495.

130. Kipnis, V., et al., *Structure of dietary measurement error: results of the OPEN biomarker study.* American journal of epidemiology, 2003. **158**(1): p. 14-21.

131. R Core Team. *R: A Language and Environment for Statistical Computing*. 2021; Available from: https://www.R-project.org/.

132. Liu, Y., et al., *Vitamin and multiple-vitamin supplement intake and incidence of colorectal cancer: a meta-analysis of cohort studies.* Medical Oncology, 2015. **32**: p. 1-10.

133. Catanzaro, R., M. Sciuto, and F. Marotta, *Lactose intolerance: An update on its pathogenesis, diagnosis, and treatment.* Nutrition Research, 2021. **89**: p. 23-34.

134. Venema, K., *Intestinal fermentation of lactose and prebiotic lactose derivatives, including human milk oligosaccharides.* International Dairy Journal, 2012. **22**(2): p. 123-140.

135. Wu, X., et al., *Effects of the intestinal microbial metabolite butyrate on the development of colorectal cancer.* J Cancer, 2018. **9**(14): p. 2510-2517.

136. Meng, Y., et al., *Dietary intakes of calcium, iron, magnesium, and potassium elements and the risk of colorectal cancer: a meta-analysis.* Biological trace element research, 2019. **189**: p. 325-335.

137. Barnung, R.B., et al., *High lactose whey cheese consumption and risk of colorectal cancer-The Norwegian Women and Cancer Study.* Scientific Reports, 2019. **9**(1): p. 296.

138. Jayasekara, H., et al., *Associations of alcohol intake, smoking, physical activity and obesity with survival following colorectal cancer diagnosis by stage, anatomic site and tumor molecular subtype.* International journal of cancer, 2018. **142**(2): p. 238-250.

139. Yang, B., et al., *Calcium, vitamin D, dairy products, and mortality among colorectal cancer survivors: the Cancer Prevention Study-II Nutrition Cohort.* J Clin Oncol, 2014. **32**(22): p. 2335-43.

140. Shultz, T., et al., *Inhibitory effect of conjugated dienoic derivatives of linoleic acid and β-carotene on the in vitro growth of human cancer cells.* Cancer letters, 1992. **63**(2): p. 125-133.

141. Fedirko, V., et al., *Effects of vitamin D and calcium supplementation on markers of apoptosis in normal colon mucosa: a randomized, double-blind, placebo-controlled clinical trial.* Cancer Prevention Research, 2009. **2**(3): p. 213-223.

142. Beasley, J.M., et al., *Associations of serum insulin-like growth factor-I and insulin-like growth factor-binding protein 3 levels with biomarker-calibrated protein, dairy product and milk intake in the Women's Health Initiative.* British journal of nutrition, 2014. **111**(5): p. 847-853.

143. Fuchs, C.S., et al., *Plasma insulin-like growth factors, insulin-like binding protein-3, and outcome in metastatic colorectal cancer: results from intergroup trial N9741.* Clinical Cancer Research, 2008. **14**(24): p. 8263-8269.

144. Wolpin, B.M., et al., *Insulin, the insulin-like growth factor axis, and mortality in patients with nonmetastatic colorectal cancer.* Journal of Clinical Oncology, 2009. **27**(2): p. 176.

145. Haydon, A.M., et al., *Physical activity, insulin-like growth factor 1, insulin-like growth factor binding protein 3, and survival from colorectal cancer.* Gut, 2006. **55**(5): p. 689-694.

146. Yu, H. and T. Rohan, *Role of the insulin-like growth factor family in cancer development and progression.* Journal of the National Cancer Institute, 2000. **92**(18): p. 1472-1489.

147. Warren, R.S., et al., *Induction of vascular endothelial growth factor by insulin-like growth factor 1 in colorectal carcinoma.* Journal of Biological Chemistry, 1996. **271**(46): p. 29483-29488.

148. Freier, S., et al., *Expression of the insulin-like growth factors and their receptors in adenocarcinoma of the colon.* Gut, 1999. **44**(5): p. 704-708.

149. National Center for Biotechnology Information. *CASR calcium sensing receptor [Homo sapiens (human)]*. 2024 [cited 2024 23 March]; Available from: https://www.ncbi.nlm.nih.gov/gene/846.

150. National Center for Biotechnology Information. *vitamin D receptor [Homo sapiens (human)]*. 2024 [cited 2024 23 March]; Available from: https://www.ncbi.nlm.nih.gov/gene/7421.

151. Kuokkanen, M., et al., *Transcriptional regulation of the lactase-phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia.* Gut, 2003. **52**(5): p. 647-52.

152. Anguita-Ruiz, A., C.M. Aguilera, and Á. Gil, *Genetics of Lactose Intolerance: An Updated Review and Online Interactive World Maps of Phenotype and Genotype Frequencies.* Nutrients, 2020. **12**(9).

153. Iamartino, L., et al., *Calcium-sensing receptor in colorectal inflammation and cancer: Current insights and future perspectives.* World Journal of Gastroenterology, 2018. **24**(36): p. 4119.

154. Aggarwal, A., et al., *The calcium-sensing receptor: A promising target for prevention of colorectal cancer.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2015. **1853**(9): p. 2158-2167.

155. Touvier, M., et al., *Meta-analyses of vitamin D intake, 25-hydroxyvitamin D status, vitamin D receptor polymorphisms, and colorectal cancer risk.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(5): p. 1003-16.

156. Rasinperä, H., et al., *The c/c− 13910 genotype of adult-type hypolactasia is associated with an increased risk of colorectal cancer in the Finnish population.* Gut, 2005. **54**(5): p. 643-647.

157. Piepoli, A., et al., *Genotyping of the lactase-phlorizin hydrolase c/t-13910 polymorphism by means of a new rapid denaturing high-performance liquid chromatography-based assay in healthy subjects and colorectal cancer patients.* Journal of biomolecular screening, 2007. **12**(5): p. 733-739.

158. Gençdal, G., et al., *Association of LCT-13910 C/T polymorphism and colorectal cancer.* Annals of Coloproctology, 2017. **33**(5): p. 169.

159. Bácsi, K., et al., *Effects of the lactase 13910 C/T and calcium-sensor receptor A986S G/T gene polymorphisms on the incidence and recurrence of colorectal cancer in Hungarian population.* BMC cancer, 2008. **8**: p. 1-8.

160. Tarabra, E., et al., *LCT-13910C> T polymorphism-associated lactose malabsorption and risk for colorectal cancer in Italy.* Digestive and liver disease, 2010. **42**(10): p. 741-743.

161. The Multiethnic Cohort Study. *Composition of the Cohort*. 2023 [cited 2024 12 March]; Available from: https://www.uhcancercenter.org/for-researchers/mec-cohort-composition.

162. Le Marchand, L., et al., *Feasibility of collecting buccal cell DNA by mail in a cohort study.* Cancer Epidemiol Biomarkers Prev, 2001. **10**(6): p. 701-3.

163. Chin, E.L., et al., *Association of lactase persistence genotypes (rs4988235) and ethnicity with dairy intake in a healthy US population.* Nutrients, 2019. **11**(8): p. 1860.

164. De Bakker, P.I., et al., *Efficiency and power in genetic association studies.* Nature genetics, 2005. **37**(11): p. 1217-1223.

165. Tsilidis, K.K., et al., *Association of common polymorphisms in IL10, and in other genes related to inflammatory response and obesity with colorectal cancer.* Cancer Causes & Control, 2009. **20**: p. 1739-1751.

166. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the Royal statistical society: series B (Methodological), 1995. **57**(1): p. 289-300.

167. Kim, J.-W., *Lactose intolerance and colorectal cancer.* Annals of Coloproctology, 2017. **33**(5): p. 157.

168. Campbell, B.J., et al., *Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic adenocarcinoma and ulcerative colitis mucin and its concealment in normal mucin.* The Journal of clinical investigation, 1995. **95**(2): p. 571-576.

169. Evans, R.C., et al., *Diet and colorectal cancer: an investigation of the lectin/galactose hypothesis.* Gastroenterology, 2002. **122**(7): p. 1784-1792.

170. Heine-Bröring, R.C., et al., *Dietary supplement use and colorectal cancer risk: a systematic review and meta-analyses of prospective cohort studies.* International Journal of Cancer, 2015. **136**(10): p. 2388-2401.

171. Chau, R., et al., *Multivitamin, calcium and folic acid supplements and the risk of colorectal cancer in Lynch syndrome.* International journal of epidemiology, 2016. **45**(3): p. 940-953.

172. Cruz-Pierard, S.M., T. Nestares, and F.J. Amaro-Gahete, *Vitamin D and calcium as key potential factors related to colorectal cancer prevention and treatment: a systematic review*. Nutrients, 2022. **14**(22): p. 4934.

173. Chakrabarty, S., et al., *Extracellular calcium and calcium sensing receptor function in human colon carcinomas: promotion of E-cadherin expression and suppression of β-catenin/TCF activation.* Cancer research, 2003. **63**(1): p. 67-71.

174. Kállay, E., et al., *Calcium-Dependent c-mycProto-Oncogene Expression and Proliferation of CACO-2 Cells: A Role for a Luminal Extracellular Calcium-Sensing Receptor.* Biochemical and biophysical research communications, 1997. **232**(1): p. 80-83.

175. Kallay, E., et al., *Dietary calcium and growth modulation of human colon cancer cells: role of the extracellular calcium-sensing receptor.* Cancer detection and prevention, 2000. **24**(2): p. 127-136.

176. Lamprecht, S.A. and M. Lipkin, *Chemoprevention of colon cancer by calcium, vitamin D and folate: molecular mechanisms.* Nature reviews cancer, 2003. **3**(8): p. 601-614.

177. Buras, R., et al., *The effect of extracellular calcium on colonocytes: evidence for differential responsiveness based upon degree of cell differentiation.* Cell proliferation, 1995. **28**(4): p. 245-262.

178. Brown, E.M. and R.J. MacLeod, *Extracellular calcium sensing and extracellular calcium signaling.* Physiological reviews, 2001. **81**(1): p. 239-297.

179. Hobson, S.A., et al., *Activation of the MAP kinase cascade by exogenous calcium-sensing receptor.* Molecular and cellular endocrinology, 2003. **200**(1-2): p. 189-198.

180. Garland, C.F., F.C. Garland, and E.D. Gorham, *Calcium and vitamin D: their potential roles in colon and breast cancer prevention.* Annals of the New York Academy of Sciences, 1999. **889**(1): p. 107-119.

181. Miyamoto, K.-i., et al., *Structural organization of the human vitamin D receptor chromosomal gene and its promoter.* Molecular endocrinology, 1997. **11**(8): p. 1165-1179.

182. Whitfield, G.K., et al., *Functionally relevant polymorphisms in the human nuclear vitamin D receptor gene.* Molecular and cellular endocrinology, 2001. **177**(1-2): p. 145-159.

183. Bai, Y.-H., et al., *Vitamin D receptor gene polymorphisms and colorectal cancer risk: a systematic meta-analysis.* World Journal of Gastroenterology: WJG, 2012. **18**(14): p. 1672.

184. Eklund, E.A., L. Bode, and H.H. Freeze, *4.19 - Diseases Associated with Carbohydrates/Glycoconjugates**, in *Comprehensive Glycoscience*, H. Kamerling, Editor. 2007, Elsevier: Oxford. p. 339-371.

185. Carroccio, A., et al., *Lactose intolerance and self-reported milk intolerance: relationship with lactose maldigestion and nutrient intake. Lactase Deficiency Study Group.* J Am Coll Nutr, 1998. **17**(6): p. 631-6.

186. Dewiasty, E., et al., *Prevalence of lactose intolerance and nutrients intake in an older population regarded as lactase non-persistent.* Clinical Nutrition ESPEN, 2021. **43**: p. 317-321.

187. Kato, K., et al., *Association between functional lactase variants and a high abundance of Bifidobacterium in the gut of healthy Japanese people.* PLoS One, 2018. **13**(10): p. e0206189.

188. Heine-Bröring, R.C., et al., *Dietary supplement use and colorectal cancer risk: a systematic review and meta-analyses of prospective cohort studies.* Int J Cancer, 2015. **136**(10): p. 2388-401.

189. Chau, R., et al., *Multivitamin, calcium and folic acid supplements and the risk of colorectal cancer in Lynch syndrome.* Int J Epidemiol, 2016. **45**(3): p. 940-53.

190. Arayici, M.E., Y. Basbinar, and H. Ellidokuz, *Vitamin D Intake, Serum 25-Hydroxyvitamin-D (25(OH)D) Levels, and Cancer Risk: A Comprehensive Meta-Meta-Analysis Including Meta-Analyses of Randomized Controlled Trials and Observational Epidemiological Studies.* Nutrients, 2023. **15**(12).

191. Fratila, T.D., A. Ismaiel, and D.L. Dumitrascu, *Microbiome modulation in the prevention and management of colorectal cancer: a systematic review of clinical interventions.* Med Pharm Rep, 2023. **96**(2): p. 131-145.

192. Tsilidis, K.K., et al., *C-reactive protein and colorectal cancer risk: A systematic review of prospective studies.* International journal of cancer, 2008. **123**(5): p. 1133-1140.

193. Murphy, N., et al., *Circulating levels of insulin-like growth factor 1 and insulin-like growth factor binding protein 3 associate with risk of colorectal cancer based on serologic and mendelian randomization analyses.* Gastroenterology, 2020. **158**(5): p. 1300-1312. e20.

194. Mehta, R.S., et al., *A prospective study of macrophage inhibitory cytokine-1 (MIC-1/GDF15) and risk of colorectal cancer.* Journal of the National Cancer Institute, 2014. **106**(4): p. dju016.

195. Kakourou, A., et al., *Interleukin-6 and risk of colorectal cancer: results from the CLUE II cohort and a meta-analysis of prospective studies.* Cancer Causes & Control, 2015. **26**: p. 1449-1460.

196. Piepoli, A., et al., *Genotyping of the lactase-phlorizin hydrolase c/t-13910 polymorphism by means of a new rapid denaturing high-performance liquid chromatography-based assay in healthy subjects and colorectal cancer patients.* J Biomol Screen, 2007. **12**(5): p. 733-9.

197. Rasinperä, H., et al., *The C/C-13910 genotype of adult-type hypolactasia is associated with an increased risk of colorectal cancer in the Finnish population.* Gut, 2005. **54**(5): p. 643-7.

198. Gençdal, G., et al., *Association of LCT-13910 C/T Polymorphism and Colorectal Cancer.* Ann Coloproctol, 2017. **33**(5): p. 169-172.

199. Bácsi, K., et al., *Effects of the lactase 13910 C/T and calcium-sensor receptor A986S G/T gene polymorphisms on the incidence and recurrence of colorectal cancer in Hungarian population.* BMC Cancer, 2008. **8**: p. 317.

200. Tarabra, E., et al., *LCT-13910C>T polymorphism-associated lactose malabsorption and risk for colorectal cancer in Italy.* Dig Liver Dis, 2010. **42**(10): p. 741-3.

201. Larsson, S.C., et al., *Genetically proxied milk consumption and risk of colorectal, bladder, breast, and prostate cancer: a two-sample Mendelian randomization study.* BMC Med, 2020. **18**(1): p. 370.

202. Lumsden, A.L., A. Mulugeta, and E. Hyppönen, *Milk consumption and risk of twelve cancers: A large-scale observational and Mendelian randomisation study.* Clinical Nutrition, 2023. **42**(1): p. 1-8.

203. Shepherd, S.J., M.C. Lomer, and P.R. Gibson, *Short-chain carbohydrates and functional gastrointestinal disorders.* Am J Gastroenterol, 2013. **108**(5): p. 707-17.

204. Misselwitz, B., et al., *Lactose malabsorption and intolerance: pathogenesis, diagnosis and treatment.* United European Gastroenterol J, 2013. **1**(3): p. 151-9.

205. Holma, R., et al., *Consumption of Lactose, Other FODMAPs and Diarrhoea during Adjuvant 5-Fluorouracil Chemotherapy for Colorectal Cancer.* Nutrients, 2020. **12**(2): p. 407.

206. Österlund, P., et al., *Lactose intolerance associated with adjuvant 5-fluorouracil-based chemotherapy for colorectal cancer*. Clinical Gastroenterology and Hepatology, 2004. **2**(8): p. 696-703.

207. Wright, E.M., M.n.G. Martín, and E. Turk, *Intestinal absorption in health and disease—sugars.* Best Practice & Research Clinical Gastroenterology, 2003. **17**(6): p. 943-956.

208. Grover, S., et al., *Mendelian Randomization.* Methods Mol Biol, 2017. **1666**: p. 581-628.

209. Smith, G.D., N. Timpson, and S. Ebrahim, *Strengthening causal inference in cardiovascular epidemiology through Mendelian randomization.* Ann Med, 2008. **40**(7): p. 524-41.

210. Hingorani, A. and S. Humphries, *Nature's randomised trials.* Lancet, 2005. **366**(9501): p. 1906-8.

211. Gill, D., et al., *Genetically determined FXI (factor XI) levels and risk of stroke.* Stroke, 2018. **49**(11): p. 2761-2763.

212. Gill, D. and S. Burgess, *Use of a genetic variant related to circulating FXa (Activated Factor X) levels to proxy the effect of FXa inhibition on cardiovascular outcomes.* Circulation: Genomic and Precision Medicine, 2020. **13**(5): p. 551-553.

213. Gill, D., et al., *Mendelian randomization for studying the effects of perturbing drug targets.* Wellcome open research, 2021. **6**.

214. Schmidt, A.F., et al., *Genetic drug target validation using Mendelian randomisation.* Nature communications, 2020. **11**(1): p. 3255.

215. Gkatzionis, A., S. Burgess, and P.J. Newcombe, *Statistical methods for cis-Mendelian randomization with two-sample summary-level data.* Genetic epidemiology, 2023. **47**(1): p. 3-25.

216. Emdin, C.A., A.V. Khera, and S. Kathiresan, *Mendelian randomization.* Jama, 2017. **318**(19): p. 1925-1926.

217. Sollis, E., et al., *The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource.* Nucleic Acids Res, 2023. **51**(D1): p. D977-d985.

218. Skrivankova, V.W., et al., *Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration.* BMJ, 2021. **375**: p. n2233.

219. Pietzner, M., et al., *Mapping the proteo-genomic convergence of human diseases.* Science, 2021. **374**(6569): p. eabj1541.

220. Pietzner, M., et al., *Genetic architecture of host proteins involved in SARS-CoV-2 infection.* Nat Commun, 2020. **11**(1): p. 6397.

221. Chen, Z., et al., *Revisiting the genome-wide significance threshold for common variant GWAS.* G3 (Bethesda), 2021. **11**(2).

222. Machiela, M.J. and S.J. Chanock, *LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants.* Bioinformatics, 2015. **31**(21): p. 3555-7.

223. Luo, J., et al., *Genetic Associations Between Modifiable Risk Factors and Alzheimer Disease.* JAMA network open, 2023. **6**(5): p. e2313734-e2313734.

224. Papadimitriou, N., et al., *Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis.* Nat Commun, 2020. **11**(1): p. 597.

225. Burgess, S. and S.G. Thompson, *Use of allele scores as instrumental variables for Mendelian randomization.* International Journal of Epidemiology, 2013. **42**(4): p. 1134-1144.

226. Machiela, M.J., et al., *GWAS Explorer: an open-source tool to explore, visualize, and access GWAS summary statistics in the PLCO Atlas.* Scientific Data, 2023. **10**(1): p. 25.

227. Kurki, M.I., et al., *FinnGen provides genetic insights from a well-phenotyped isolated population.* Nature, 2023. **613**(7944): p. 508-518.

228. Pan-UKB team. 2020; Available from: https://pan.ukbb.broadinstitute.org.

229. Lawlor, D.A., et al., *Mendelian randomization: using genes as instruments for making causal inferences in epidemiology.* Stat Med, 2008. **27**(8): p. 1133-63.

230. Bowden, J., et al., *Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator.* Genetic epidemiology, 2016. **40**(4): p. 304-314.

231. Burgess, S., et al., *Robust instrumental variable methods using multiple candidate instruments with application to Mendelian randomization.* arXiv preprint arXiv:1606.03729, 2016.

232. Slob, E.A.W. and S. Burgess, *A comparison of robust Mendelian randomization methods using summary data.* Genet Epidemiol, 2020. **44**(4): p. 313-329.

233. Bowden, J., G. Davey Smith, and S. Burgess, *Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression.* International journal of epidemiology, 2015. **44**(2): p. 512-525.

234. Burgess, S. and S.G. Thompson, *Interpreting findings from Mendelian randomization using the MR-Egger method.* European journal of epidemiology, 2017. **32**: p. 377-389.

235. Hartwig, F.P., G. Davey Smith, and J. Bowden, *Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption.* Int J Epidemiol, 2017. **46**(6): p. 1985-1998.

236. Rees, J.M.B., et al., *Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates.* PLoS One, 2019. **14**(9): p. e0222362.

237. Higgins, J.P., et al., *Measuring inconsistency in meta-analyses.* Bmj, 2003. **327**(7414): p. 557-560.

238. Yavorska, O.O. and S. Burgess, *MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data.* International journal of epidemiology, 2017. **46**(6): p. 1734-1739.

239. Balduzzi, S., G. Rücker, and G. Schwarzer, *How to perform a meta-analysis with R: a practical tutorial.* BMJ Ment Health, 2019. **22**(4): p. 153-160.

240. de la Chapelle, A., *Genetic predisposition to colorectal cancer.* Nature Reviews Cancer, 2004. **4**(10): p. 769-780.

241. Vissers, L.E.T., et al., *Dairy Product Intake and Risk of Type 2 Diabetes in EPIC-InterAct: A Mendelian Randomization Study.* Diabetes Care, 2019. **42**(4): p. 568-575.

242. Bergholdt, H.K.M., et al., *Lactase persistence, milk intake, hip fracture and bone mineral density: a study of 97 811 Danish individuals and a meta-analysis.* J Intern Med, 2018. **284**(3): p. 254-269.

243. Yuan, S., et al., *Health effects of milk consumption: phenome-wide Mendelian randomization study.* BMC Med, 2022. **20**(1): p. 455.

244. Aune, D., et al., *Dairy products and colorectal cancer risk: a systematic review and meta-analysis of cohort studies.* Ann Oncol, 2012. **23**(1): p. 37-45.

245. Thorning, T.K., et al., *Milk and dairy products: good or bad for human health? An assessment of the totality of scientific evidence.* Food Nutr Res, 2016. **60**: p. 32527.

246. Vieira, A.R., et al., *Foods and beverages and colorectal cancer risk: a systematic review and meta-analysis of cohort studies, an update of the evidence of the WCRF-AICR Continuous Update Project.* Ann Oncol, 2017. **28**(8): p. 1788-1802.

247. Kim, J.W., *Lactose Intolerance and Colorectal Cancer.* Ann Coloproctol, 2017. **33**(5): p. 157-158.

248. Campbell, B.J., et al., *Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic adenocarcinoma and ulcerative colitis mucin and its concealment in normal mucin.* J Clin Invest, 1995. **95**(2): p. 571-6.

249. Yang, W., et al., *Calcium intake and risk of colorectal cancer according to expression status of calcium-sensing receptor (CASR).* Gut, 2018. **67**(8): p. 1475-1483.

250. Kozu, T., et al., *Effect of Orally Administered Bovine Lactoferrin on the Growth of Adenomatous Colorectal Polyps in a Randomized, Placebo-Controlled Clinical TrialEffect of bLF on Colorectal Polyps.* Cancer Prevention Research, 2009. **2**(11): p. 975-983.

251. Abdelali, H., et al., *Effect of dairy products on initiation of precursor lesions of colon cancer in rats.* 1995.

252. Parodi, P., *Conjugated linoleic acid and other anticarcinogenic agents of bovine milk fat.* Journal of dairy science, 1999. **82**(6): p. 1339-1349.

253. Velázquez, O.C., et al., *In vivo crypt surface hyperproliferation is decreased by butyrate and increased by deoxycholate in normal rat colon: associated in vivo effects on c-fos and c-jun expression.* Journal of Parenteral and Enteral Nutrition, 1996. **20**(4): p. 243-250.

254. Sivan, A., et al., *Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–PD-L1 efficacy.* Science, 2015. **350**(6264): p. 1084-1089.

255. Burgess, S., N.M. Davies, and S.G. Thompson, *Bias due to participant overlap in two-sample Mendelian randomization.* Genet Epidemiol, 2016. **40**(7): p. 597-608.

256. Xie, Y.-H., Y.-X. Chen, and J.-Y. Fang, *Comprehensive review of targeted therapy for colorectal cancer.* Signal transduction and targeted therapy, 2020. **5**(1): p. 22.

257. Nagini, S., *Breast cancer: current molecular therapeutic targets and new players.* Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents), 2017. **17**(2): p. 152-163.

258. Ratta, R., et al., *PARP inhibitors as a new therapeutic option in metastatic prostate cancer: a systematic review.* Prostate Cancer and Prostatic Diseases, 2020. **23**(4): p. 549-560.

259. Yuan, M., et al., *The emerging treatment landscape of targeted therapy in non-small-cell lung cancer.* Signal transduction and targeted therapy, 2019. **4**(1): p. 61.

260. Chiorean, E.G. and A.L. Coveler, *Pancreatic cancer: optimizing treatment options, new, and emerging targeted therapies.* Drug design, development and therapy, 2015: p. 3529-3545.

261. Luo, M., et al., *Drug resistance in colorectal cancer: An epigenetic overview.* Biochim Biophys Acta Rev Cancer, 2021. **1876**(2): p. 188623.

262. Friedmann Angeli, J.P., D.V. Krysko, and M. Conrad, *Ferroptosis at the crossroads of cancer-acquired drug resistance and immune evasion.* Nat Rev Cancer, 2019. **19**(7): p. 405-414.

263. Suhre, K., et al., *Connecting genetic risk to disease end points through the human blood plasma proteome.* Nature communications, 2017. **8**(1): p. 14357.

264. Davies, M.P., et al., *Plasma protein biomarkers for early prediction of lung cancer.* EBioMedicine, 2023. **93**.

265. Sun, B.B., et al., *Genomic atlas of the human plasma proteome.* Nature, 2018. **558**(7708): p. 73-79.

266. Bonaterra, G.A., et al., *Increased density of growth differentiation factor-15+ immunoreactive M1/M2 macrophages in prostate cancer of different gleason scores compared with benign prostate hyperplasia.* Cancers, 2022. **14**(19): p. 4591.

267. Myte, R., et al., *A longitudinal study of prediagnostic metabolic biomarkers and the risk of molecular subtypes of colorectal cancer.* Scientific Reports, 2020. **10**(1): p. 5336.

268. Liu, H., et al., *Correlation of TBK1, AR, and other serum cancer-related biomarkers in breast cancer patients: An observational study.* Medicine, 2022. **101**(33): p. e29996.

269. Shimizu, J., et al., *Biomarker testing for personalized, first-line therapy in advanced nonsquamous non-small cell lung cancer patients in the real world setting in Japan: a retrospective, multicenter, observational study (the BRAVE study).* Therapeutic advances in medical oncology, 2020. **12**: p. 1758835920904522.

270. Ferkingstad, E., et al., *Large-scale integration of the plasma proteome with genetics and disease.* Nature genetics, 2021. **53**(12): p. 1712-1721.

271. Pietzner, M., et al., *Mapping the proteo-genomic convergence of human diseases.* Science, 2021. **374**(6569): p. eabj1541.

272. Sun, B.B., et al., *Plasma proteomic associations with genetics and health in the UK Biobank.* Nature, 2023. **622**(7982): p. 329-338.

273. Zhang, Y., et al., *Evaluating the causal effect of circulating proteome on the risk of osteoarthritis-related traits.* Annals of the Rheumatic Diseases, 2023. **82**(12): p. 1606-1617.

274. Kurki, M.I., et al., *FinnGen provides genetic insights from a well-phenotyped isolated population.* Nature, 2023. **613**(7944): p. 508-518.

275. Lawlor, D.A., et al., *Mendelian randomization: using genes as instruments for making causal inferences in epidemiology.* Statistics in medicine, 2008. **27**(8): p. 1133-1163.

276. Hemani, G., et al., *The MR-Base platform supports systematic causal inference across the human phenome.* elife, 2018. **7**: p. e34408.

277. Giambartolomei, C., et al., *Bayesian test for colocalisation between pairs of genetic association studies using summary statistics.* PLoS genetics, 2014. **10**(5): p. e1004383.

278. Hemani, G., K. Tilling, and G. Davey Smith, *Orienting the causal relationship between imprecisely measured traits using GWAS summary data.* PLoS genetics, 2017. **13**(11): p. e1007081.

279. Wu, T., et al., *clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.* The innovation, 2021. **2**(3).

280. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* Omics: a journal of integrative biology, 2012. **16**(5): p. 284-287.

281. Enattah, N.S., et al., *Identification of a variant associated with adult-type hypolactasia.* Nat Genet, 2002. **30**(2): p. 233-7.

282. Itan, Y., et al., *A worldwide correlation of lactase persistence phenotype and genotypes.* BMC Evol Biol, 2010. **10**: p. 36.

283. Li, R., et al., *Gremlin-1 promotes colorectal cancer cell metastasis by activating ATF6 and inhibiting ATF4 pathways.* Cells, 2022. **11**(14): p. 2136.

284. Karagiannis, G.S., et al., *Bone morphogenetic protein antagonist gremlin-1 regulates colon cancer progression.* Biological chemistry, 2015. **396**(2): p. 163-183.

285. Davies, G.C., et al. *Discovery of ginisortamab, a potent and novel anti-gremlin-1 antibody in clinical development for the treatment of cancer*. in *Mabs*. 2023. Taylor & Francis.

286. Han, S., et al., *Genetically determined circulating lactase/phlorizin hydrolase concentrations and risk of colorectal cancer: A two-sample mendelian randomization study.* Nutrients, 2024. **16**(6): p. 808.

287. Amiri, M., et al., *The Diverse Forms of Lactose Intolerance and the Putative Linkage to Several Cancers.* Nutrients, 2015. **7**(9): p. 7209-30.

288. Agarwal, M.m., et al., *Lactose intolerance in prostate cancer patients: incidence and associated factors.* Scandinavian journal of gastroenterology, 2008. **43**(3): p. 270-276.

289. Peak, T.C., et al., *The role of PDE5 inhibitors and the NO/cGMP pathway in cancer.* Sexual medicine reviews, 2016. **4**(1): p. 74-84.

290. Li, D., et al., *EPLIN-β is a novel substrate of ornithine decarboxylase antizyme 1 and mediates cellular migration.* Journal of Cell Science, 2023. **136**(12).

291. Zhang, S., et al., *EPLIN downregulation promotes epithelial–mesenchymal transition in prostate cancer cells and correlates with clinical lymph node metastasis.* Oncogene, 2011. **30**(50): p. 4941-4952.

292. American Cancer Society, *Colorectal cancer facts & figures 2023–2025.* 2023.