Cumulative Burden of Colorectal Cancer-Associated Genetic Variants is More Strongly Associated With Early-onset vs Late-onset Cancer

Alexi N Archambault[1], Yu-Ru Su[2], Jihyoun Jeon[3], Minta Thomas[2], Yi Lin[2], David V Conti[4], Aung Ko Win[5], Lori C Sakoda[2,6], Iris Lansdorp-Vogelaar[7], Elisabeth FP Peterse[7], Ann G Zauber[8], David Duggan[9], Andreana N Holowatyj[10], Jeroen R Huyghe[2], Hermann Brenner[11-13], Michelle Cotterchio[14], Stéphane Bézieau[15], Stephanie L Schmit[4,16], Christopher K Edlund[4], Melissa C Southey[17], Robert J MacInnis[5,18], Peter T Campbell[19], Jenny Chang-Claude[20,21], Martha L Slattery[22], Andrew T Chan[23-28], Amit D Joshi[25,27], Mingyang Song[29], Yin Cao[25,30], Michael O Woods[31], Emily White[2,32], Stephanie J Weinstein[33], Cornelia M Ulrich[34], Michael Hoffmeister[11], Stephanie A Bien[2], Tabitha A Harrison[2], Jochen Hampe[35], Christopher I Li[2], Clemens Schafmayer[36], Kenneth Offit[37,38], Paul D Pharoah[39], Victor Moreno[40-42], Annika Lindblom[43,44], Alicja Wolk[45], Anna H Wu[4], Li Li[46], Marc J Gunter[47], Andrea Gsur[48], Temitope O Keku[49], Rachel Pearlman[50], D Timothy Bishop[51], Sergi Castellví-Bel[52], Leticia Moreira[52], Pavel Vodicka[53-55], Ellen Kampman[56], Graham G Giles[5,16], Demetrius Albanes[32], John A Baron[57], Sonja I Berndt[32], Stefanie Brezina[48], Stephan Buch[34], Daniel D Buchanan[5,58-60], Antonia Trichopoulou[61], Gianluca Severi[62], María-Dolores Chirlaque[41,63], Maria-José Sánchez[64], Domenico Palli[65], Tilman Kühn[20], Neil Murphy[66], Amanda J Cross[67], Andrea N Burnett-Hartman[68], Stephen J Chanock[32], Albert de la Chapelle[69], Douglas F Easton[38], Faye Elliott[51], Dallas R English[5,16], Edith JM Feskens[56], Liesel M FitzGerald[16,70], Phyllis J Goodman[71], John L Hopper[5,72], Thomas J Hudson[73], David J Hunter[27,74], Eric J Jacobs[19], Corinne E Joshu[75], Sébastien Küry[18], Sanford D Markowitz[45], Roger L Milne[5,16], Elizabeth A Platz[75], Gad Rennert[76-78], Hedy S Rennert[76-78], Fredrick R Schumacher[79], Robert S Sandler[49], Daniela Seminara[80], Catherine M Tangen[71], Stephen N Thibodeau[81], Amanda E Toland[69], Franzel JB van Duijnhoven[56], Kala Visvanathan[75], Ludmila Vodickova[53-55], John D Potter[2], Satu Männistö[82], Korbinian Weigl[11,83], Jane Figueiredo[4,84], Vicente Martín[41,85], Susanna C Larsson[44], Patrick S Parfrey[86], Wen-Yi Huang[33], Heinz-Josef Lenz[87], Jose E Castelao[88], Manuela Gago-Dominguez[89,90], Victor Muñoz-Garzón[91], Christoph Mancao[92], Christopher A Haiman[4], Lynne R Wilkens[93], Erin Siegel[16], Elizabeth Barry[94], Ban Younghusband[30], Bethany Van Guelpen[95,96], Sophia Harlid[96], Anne Zeleniuch-Jacquotte[1], Peter S Liang[97], Mengmeng Du[8], Graham Casey[98], Noralane M Lindor[99], Loic Le Marchand[93], Steven J Gallinger[100], Mark A Jenkins[5], Polly A Newcomb[2,101], Stephen B Gruber[4], Robert E Schoen[102], Heather Hampel[50], Douglas A Corley[6]§, Li Hsu[2,103]§, Ulrike Peters[2,31]§, Richard B Hayes[1]§

§These authors jointly supervised this work.

1. Division of Epidemiology, Department of Population Health, New York University School of Medicine, New York, New York, USA.
2. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.
3. Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, USA.
4. Department of Preventive Medicine, USC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.

5. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia.
6. Division of Research, Kaiser Permanente Northern California, Oakland, California, USA.
7. Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, The Netherlands.
8. Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA.
9. Translational Genomics Research Institute - An Affiliate of City of Hope, Phoenix, Arizona, USA.
10. Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, Utah, USA.
11. Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.
12. Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany.
13. German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.
14. Population Health and Prevention, Cancer Care Ontario, Toronto, Ontario, Canada.
15. Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, Nantes, France.
16. Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA.
17. Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Australia.
18. Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia.
19. Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, Georgia, USA.
20. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
21. University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg

(UCCH), Hamburg, Germany.
22. Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA.
23. Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.
24. Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.
25. Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.
26. Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
27. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA.
28. Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA.
29. Department of Nutrition, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA.
30. Division of Public Health Sciences, Department of Surgery, Washington University in St. Louis, St. Louis, Missouri, USA.
31. Memorial University of Newfoundland, Discipline of Genetics, St. John's, Canada.
32. Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA.
33. Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA.
34. Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, Utah, USA.
35. Department of Medicine I, University Hospital Dresden, Technische Universität Dresden (TU Dresden), Dresden, Germany.
36. Department of General and Thoracic Surgery, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany.
37. Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA.
38. Department of Medicine, Weill Cornell Medical College, New York, New York, USA.

39. Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
40. Cancer Prevention and Control Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain.
41. CIBER in Epidemiology and Public Health (CIBERESP), Madrid, Spain
42. Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain.
43. Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden.
44. Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden.
45. Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden.
46. Department of Family Medicine, University of Virginia, Charlottesville, Virginia, USA.
47. Nutrition and Metabolism Section, International Agency for Research on Cancer, World Health Organization, Lyon, France.
48. Institute of Cancer Research, Department of Medicine I, Medical University of Vienna, Vienna, Austria.
49. Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, North Carolina, USA.
50. Division of Human Genetics, Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio, USA.
51. Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK.
52. Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Spain.
53. Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic.
54. Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic.
55. Faculty of Medicine and Biomedical Center in Pilsen, Charles University, Pilsen, Czech Republic.
56. Division of Human Nutrition and Health, Wageningen University & Research,

Wageningen, The Netherlands.
57. Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA.
58. Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Parkville, Victoria, Australia.
59. Genomic Medicine and Family Cancer Clinic, The Royal Melbourne Hospital, Parkville, Victoria, Australia.
60. University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Parkville, Victoria, Australia.
61. Hellenic Health Foundation, Athens, Greece.
62. Centre de Recherche en Épidémiologie et Santé des Populations (CESP, Inserm U1018), Facultés de Médecine, Université Paris-Saclay, Gustave Roussy, Villejuif, France.
63. Department of Epidemiology, Regional Health Council, IMIB-Arrixaca, Murcia University, Murcia, Spain.
64. Escuela Andaluza de Salud Pública, CIBER de Epidemiología y Salud Pública, Granada, Spain.
65. Cancer Risk Factors and Life-Style Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network - ISPRO, Florence, Italy.
66. Section of Nutrition and Metabolism, International Agency for Research on Cancer, Lyon, France.
67. School of Public Health, Imperial College London, London, UK.
68. Institute for Health Research, Kaiser Permanente Colorado, Aurora, Colorado, USA.
69. Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA.
70. Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia.
71. SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.
72. Department of Epidemiology, School of Public Health and Institute of Health and Environment, Seoul National University, Seoul, South Korea.
73. Ontario Institute for Cancer Research, Toronto, Ontario, Canada.

74. Nuffield Department of Population Health, University of Oxford, Oxford, UK.
75. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.
76. Department of Community Medicine and Epidemiology, Lady Davis Carmel Medical Center, Haifa, Israel.
77. Ruth and Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel.
78. Clalit National Cancer Control Center, Haifa, Israel.
79. Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA.
80. Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland, USA.
81. Division of Laboratory Genetics, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA.
82. Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland.
83. Medical Faculty, University of Heidelberg, Heidelberg, Germany.
84. Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.
85. Biomedicine Institute (IBIOMED), University of León, León, Spain.
86. The Clinical Epidemiology Unit, Memorial University Medical School, St. John's, Newfoundland, Canada.
87. Division of Medical Oncology, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
88. Instituto de Investigación Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Oncology and Genetics Unit, Vigo, Spain.
89. Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain.
90. Moores Cancer Center, University of California San Diego, La Jolla, CA, USA.
91. Radiotherapy Department, Complejo Hospitalario Universitario de Vigo, SERGAS, Vigo,

Spain.
92. Genentech, Inc., Basel, Switzerland.
93. Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA.
94. Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA.
95. Wallenberg Centre for Molecular Medicine, Umeå University, Umeå, Sweden.
96. Department of Radiation Sciences, Oncology Unit, Umeå University, Umeå, Sweden.
97. Department of Medicine, New York University School of Medicine, New York, New York, USA.
98. Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA.
99. Department of Health Science Research, Mayo Clinic, Scottsdale, Arizona, USA.
100. Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto,Toronto, Ontario, Canada.
101. School of Public Health, University of Washington, Seattle, Washington, USA.
102. Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA.
103. Department of Biostatistics, University of Washington, Seattle, Washington, USA.

Shared corresponding authors:

Richard B. Hayes

NYU Langone Health

180 Madison Ave, Room 415

New York, NY 10016

Ulrike Peters

Fred Hutchinson Cancer Research Center

1100 Fairview Avenue N, M4-B402

Seattle, WA 98109-1024

Grant Support: National Cancer Institute, National Institutes of Health (Grants: R03-CA215775-01A1, R01-CA206279-03).

Abstract

Background & Aims: Early-onset colorectal cancer (CRC, in persons younger than 50 years old) is increasing in incidence; yet, in the absence of a family history of CRC, this population lacks harmonized recommendations for prevention. We aimed to determine whether a polygenic risk score (PRS) developed from 95 CRC-associated common genetic risk variants was associated with risk for early-onset CRC.

Methods: We studied risk for CRC associated with a weighted PRS in 12,197 participants younger than 50 years old vs 95,865 participants 50 years or older. PRS was calculated based on single-nucleotide polymorphisms associated with CRC in a large-scale genome-wide association study as of January 2019. Participants were pooled from 3 large consortia that provided clinical and genotyping data: the Colon Cancer Family Registry, the Colorectal Transdisciplinary study, and the Genetics and Epidemiology of Colorectal Cancer Consortium and were all of genetically defined European descent. Findings were replicated in an independent cohort of 72,573 participants.

Results: Overall associations with CRC per standard deviation of PRS were significant for early-onset cancer, and were stronger compared with late-onset cancer (*P* for interaction=.01); when we compared the highest PRS quartile with the lowest, risk increased 3.7-fold for early-onset CRC (95% CI, 3.28–4.24) vs 2.9-fold for late-onset CRC (95% CI, 2.80–3.04). This association was strongest for participants without a first-degree family history of CRC (*P* for interaction=5.61×10^{-5}). When we compared the highest with the lowest quartiles in this group, risk increased 4.3-fold for early-onset CRC (95% CI, 3.61–5.01) vs 2.9-fold for late-onset CRC (95% CI, 2.70–3.00). Sensitivity analyses were consistent with these findings.

Conclusions: In an analysis of associations with CRC per standard deviation of PRS, we found the cumulative burden of CRC-associated common genetic variants to associate with early-onset cancer, and to be more strongly associated with early-onset than late-onset cancer—particularly in the absence of CRC family history. Analyses of PRS, along with environmental and lifestyle risk factors, might identify younger individuals who would benefit from preventative measures.

Introduction

Colorectal cancer (CRC) incidence and mortality have been declining in the U.S. over the last several decades.[1] These reductions are largely attributed to successes in CRC early detection, surveillance, and treatment for this disease.[2, 3] In contrast to these overall trends, the incidence of CRC in individuals less than 50 years of age (early-onset disease) has been increasing in the U.S. and elsewhere:[4] early-onset CRC incidence in the U.S. has increased by an average of 1.8% annually from 1992–2012, and is projected to account for 10% to 25% of newly-diagnosed CRC by 2030.[1, 5-10] Furthermore, early-onset CRC tends to present with higher pathologic grade, distant disease, and a greater incidence of recurrence and metastatic disease.[5] In response to this newly recognized disease burden, the US Preventative Services Task Force,[11] the American Cancer Society,[12] the U.S. Multi-Society Task Force on Colorectal Cancer[13] and other professional bodies[14] have initiated discussions on the merits of revising recent consensus CRC prevention guidelines to include early detection of average-risk individuals younger than 50 years of age. While the American Cancer Society recommends lowering the screening age to 45 years for individuals at average risk,[12] others recommend targeting only high-risk groups for early detection.[13, 15]

Weighing against the potential benefits of CRC early detection and prevention programs targeted to those aged younger than 50 years are concerns about adverse side effects and associated costs.[14, 16] New approaches to disease prevention in younger adults are warranted, and assessing germline genetic variants, along with other known risk factors, could facilitate tailored early detection of high risk individuals due to their genetic makeup and lifestyle. To date, genetic research on factors associated with early-onset CRC has been limited largely to the rare

monogenic, high-penetrance genetic syndromes associated with this disease in high-risk families, while the frequently occurring low-penetrance polymorphisms have been understudied.

Here, we report on CRC risks for early (<50 years of age) and late-onset disease (≥50 years of age) associated with a polygenic risk score (PRS) developed from 95 common genetic risk variants identified in previous CRC genome-wide association studies (GWAS). Our research provides the first substantive evidence that early-onset CRC exhibits differential genetic risks, compared with late-onset disease, due to low-penetrance, common genetic polymorphisms. The findings of our research may contribute to the identification of individuals susceptible to early-onset CRC for tailored early detection or other preventive interventions.

Methods

Study Participants

We studied 108,062 participants in the discovery dataset, including 50,023 CRC cases and 58,039 controls. Participants for this study were pooled from three large consortia that provided clinical and genotyping data: the Colon Cancer Family Registry (CCFR), the Colorectal Transdisciplinary (CORECT) Study, and the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) (Table 1 and Table S1) (for additional study information, see earlier publications[17-20]). All analyses were restricted to participants of genetically defined European descent. Family history of CRC was ascertained through self-report or interviewer-administered questionnaire, and defined as having one or more first-degree relatives with CRC. Participant recruitment across all studies occurred between the 1990's and the early 2010's. All study participants provided written informed consent and studies were approved by their respective Institutional Review Boards (see Supplementary Information).

Genotyping and SNP Selection

We included 95 CRC-risk-associated SNPs that reached genome-wide significance ($p \leq 5\times10^{-8}$), in large-scale GWAS, as of January, 2019. No new discovery of CRC-related SNPs was carried out here. Individual participant and genotype data for the 95 SNPs were extracted from GWAS and imputed to the Haplotype Reference Consortium panel, which provides high-quality, accurate imputation for variants with a minor allele frequency as low as 0.1%.[21] For details, see Huyghe et al.[17] Additional information on SNPs can be located in Table S2.

Statistical Analysis

For cases and controls, we compared baseline participant characteristics between individuals who had a reference age of <50 years to those with a reference age of ≥50 years of age. For cases, reference age was defined as the age of diagnosis of first primary CRC. For controls, reference age was defined as the age at selection.

Genotyped SNPs were coded as 0, 1, or 2 copies of the risk allele. Imputed SNPs were coded for the expected number of copies of the risk allele, as imputed dosages. Potential population substructure within the GECCO, CCFR, and CORECT studies was accounted for through adjustment by principal components of genetic ancestry. To develop the weighted PRS, we used log-odds ratios derived from the literature for 55 of the SNPs, and for the remaining 40 SNPs that were first identified within this discovery dataset, we computed log-odds ratios from a regression model fit with CRC as the outcome (1 vs. 0) and the following independent variables: 95 SNPs, age (in years), sex, principal components, and genotype platform. For the 40 SNPs identified within this discovery dataset, we then implemented a conservative winner's curse adjustment of the log-odds ratios from the risk model, using Zhong and Prentice's approach.[22] We then weighted the PRS for individuals, by multiplying the number of risk alleles for each

SNP by their adjusted log-odds ratios, summing and recoding as a percentile based on the distribution in the controls. The final PRS was modelled as a continuous variable per 1 standard deviation (SD), transformed to the standard normal distribution. Odds ratios and 95% confidence intervals were also estimated comparing quartiles of PRS.

We used unconditional logistic regression to assess the association between the PRS and CRC for those with a reference age <50 years and for those with a reference age ≥50 years. All models additionally included sex, reference age in years, principal components, and genotype platform. Further adjustment by study was not warranted as extensive genome-wide analyses with and without adjusting for study have been conducted, with the results being consistent.[17] To test for differences in associations across age, an interaction term was included for age category (<50, ≥50) and PRS (continuous). Models were also examined separately by first-degree family history of CRC. We evaluated the discriminatory accuracy of the risk prediction models by calculating the area under the receiver operating characteristic curve (AUC) for 5-year diagnostic age groups, adjusting for sex, PCs, and genotype platform, using the adjusted.ROC function from the R Package ROCt.

For the larger group with no first-degree family history of CRC, additional sub-group analyses were performed including estimation of CRC risk within specific reference-age groups (15-39, 40-49, 50-59, 60-69, and 70-79 years) and by disease site (proximal colon, distal colon, and rectum). The interaction term used to assess differences in associations across age categories consisted of age as a continuous variable and PRS (continuous). Multinomial logistic regression was used to assess risk differentials by disease site within age strata. Analyses were completed using the R statistical software program version 3.5.1.

Sensitivity Analyses

Replication accounting for cases with Lynch syndrome. Screening of colorectal cancer cases for the presence of Lynch syndrome was systematically carried out for CRC cases recruited through the Ohio State University Medical Center (OSUMC) (Table S1: HNPCC, OCCPI, and OSUMC) as described in detail elsewhere[23-25]. All cases were screened for MMR deficiency using immunohistochemical analysis. Cases with probable characteristics of Lynch syndrome were subjected to additional genetic testing for conclusively determining a diagnosis of Lynch syndrome based on the presence of one or more germline high penetrance mutations in DNA mismatch repair genes (MLH1, MSH2, MSH6, PMS2) or the EPCAM gene.

Using unconditional logistic regression in these studies, we evaluated the association between the PRS and CRC for those aged <50 years and for those ≥50 years of age, with consideration of Lynch syndrome status among cases. All models additionally included sex, reference age in years, and principal components. To test for differences in associations across age, an interaction term was included for age category (<50, ≥50) and PRS (continuous).

Replication in an independent cohort. To independently replicate the association of this PRS with younger and older-onset CRC, we studied all 72,573 participants of European ancestry who were genotyped in the Research Program on Genes, Environment and Health (RPGEH), a cohort comprised of Kaiser Permanente Northern California (KPNC) health plan members.[26, 27] This cohort was not included in the discovery of any of the 95 CRC genetic risk variants. Cancer history was determined from initiation of health plan membership by linkage to the KPNC Cancer Registry, which adheres to the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program standards.

Family history of CRC, defined as having one or more first-degree relatives with CRC, was ascertained through a baseline study questionnaire, electronic family history data in the medical records, and International Classification of Disease codes Z80.0 (Family history of malignant neoplasm of digestive organs) and V16.0 (Cancer family history, gastrointestinal tract). Analyses were restricted to participants of genetically defined European descent. All study participants provided written informed consent, and the study was approved by the Kaiser Permanente Northern California Institutional Review Board.

RPGEH biospecimens were genotyped using the Affymetrix Axiom platform. Details on the calling and quality control can be found elsewhere.[28] Consistent with genetic data in the discovery set, we imputed the genotyped data to the Haplotype Reference Consortium. To develop the PRS for this replication, we used 94 SNPs from the discovery dataset, as described above, and, for 1 unmatched SNP (rs755229494), we included the best available surrogate (rs112334046, R^2=0.40, MAF=0.0026).

For the longitudinal replication cohort, we employed Cox proportional hazards models to assess the association of PRS with CRC, which was not feasible for the discovery dataset since it included case-control data. The coefficients from the model fit with 95 SNPs in the discovery dataset were used to fit the PRS in the replication analysis, thereby reducing potential for overfitting. The observed time was defined from the age of initial KPNC enrollment to the earliest of age at CRC diagnosis, death or end of follow-up (the RPGEH cohort was followed until December 31, 2016). The replication models also included sex and principal components to account for potential population substructure. Estimates of absolute risk are inferred using Kaplan-Meier plots produced using RPGEH data.

Results

Early-onset CRC cases (N=5,479) had a mean age at diagnosis of 43.1 years, while the older-onset cases (N=44,544) had a mean age at diagnosis of 66.5 years (Table 1). Men and women were approximately equally represented across cases and controls. A first-degree family history of CRC, among those ascertained for family history, was reported for 17.2% of early-onset and 12.5% for late-onset CRC cases, and, respectively, for 8.6% of younger and 10.4% for older controls. Family history information was missing for >25% of participants; all of whom were from 9 studies that did not query participants on family history and therefore were not included in our family history-specific analyses. Younger onset cases tended to have fewer proximal colon tumors and a greater preponderance of tumors in the rectum. Both early-onset and late-onset CRC cases showed marked skewing toward higher PRS values compared with controls, when represented as quartiles (Table 1) and as a continuous score (Figure S1).

We found that associations with risk for CRC per SD of PRS were significant among participants <50 years of age, and were stronger compared with participants aged ≥50 years (*P* for interaction = 0.01). Contrasting the highest PRS quartile with the lowest, risks were 3.7-fold higher (OR: 3.73; 95% CI: 3.28, 4.24) for early-onset CRC and 2.9-fold higher (OR: 2.92; 95% CI: 2.80, 3.04) for late-onset disease (Table 2 and Figure 1A). For the larger group of participants who reported a negative first-degree family history of CRC, PRS-associated risks for CRC among participants aged <50 years were also stronger than those for individuals aged ≥50 years (*P* for interaction = 5.61×10^{-5}); risks comparing the highest with the lowest quartile of PRS were 4.3-fold (OR: 4.26; 95% CI: 3.61, 5.01) for early-onset CRC and 2.9-fold (OR: 2.85; 95% CI: 2.70, 3.00) for late-onset disease (Table 2 and Figure 1B). In contrast, for the smaller group of participants who reported a positive first-degree family history of CRC, risks per SD of PRS

tended to be greater for older individuals (*P* for interaction = 0.003); risks in the highest quartile for PRS were 1.7-fold (OR: 1.70; 95% CI: 1.17, 2.47) for early-onset CRC, and 2.5-fold (OR: 2.47; 95% CI: 2.18, 2.79) for late-onset disease (Table 2 and Figure 1C). The discriminatory capabilities for prediction (i.e., AUC) of these models across the entire age spectrum tended to be highest for early-onset individuals without a family history of CRC, ranging from 0.64 to 0.65 (Table S3).

As the PRS displayed the strongest association for early-onset CRC without a first-degree family history, we investigated whether certain subgroups could account for these strong effects. When stratified further by age at diagnosis, CRC risks were 1.7-fold (OR per SD of PRS: 1.74; 95% CI: 1.55, 1.96) for those diagnosed aged 15-39 years and 1.8-fold (OR per SD of PRS: 1.75; 95% CI: 1.64, 1.87) for those diagnosed aged 40-49 years of age. For participants diagnosed at ≥50 years of age, the related CRC risks were 1.6-fold (OR per SD of PRS: 1.60; 95% CI: 1.54, 1.67) for participants aged 50-59 years, 1.5-fold (OR per SD of PRS: 1.52; 95% CI: 1.48, 1.57) for individuals 60-69 years old, and 1.4-fold (OR per SD of PRS: 1.44; 95% CI: 1.39, 1.49) for those diagnosed between 70-79 years, with age and PRS exhibiting statistical interaction across the entire study age range (Table S4, *P* for interaction = 3.44×10^{-10}). Furthermore, as found for all cancer sites (Table 2 and Figure 1), the PRS was also more strongly associated with risks for early-onset, compared with late-onset, cancers of the proximal colon, distal colon and rectum (Table S5 and Figure S2), with the greatest risk differentials observed for cancers of the distal colon and rectum (Table S6).

Sensitivity Analyses

Replication accounting for cases with Lynch syndrome. A total of 37 Lynch cases <50 years of age (6.4%, among 574 cases) and 54 Lynch cases ≥50 years of age (2.1%, among 2525 cases)

were identified in the Ohio-based studies. Removing Lynch cases from the analysis demonstrated that the relatively small number of these cases did not substantially impact the relationship of PRS with CRC (Table 3). After exclusion of Lynch cases, risks for early-onset CRC per SD of PRS remained similarly increased in participants <50 years of age (OR per SD of PRS: 1.82; 95% CI: 1.61, 2.06) and were greater compared with participants aged ≥50 years (OR per SD of PRS: 1.49; 95% CI: 1.39, 1.60; *P* for interaction = 0.01). These trends held particularly for participants who reported a negative first-degree family history of CRC (aged <50 years, OR per SD of PRS: 1.83; 95% CI: 1.60, 2.09; aged ≥50 years, OR per SD of PRS: 1.46; 95% CI: 1.35,1.57; *P* for interaction = 0.01).

Replication in an independent cohort. In RPGEH, early-onset CRC cases (N=25) had a mean age of 45.2 years, while the older-onset cases (N=1,068) had a mean age of 73.7 years (Table 1). More women participated than men. A first-degree family history of CRC was reported for 28.0% of early-onset and 18.4% of late-onset CRC cases, compared to 9.6% for the cohort overall. Consistent with the discovery dataset, the distributions of PRS for both early and late-onset CRC cases were skewed towards higher PRS quartiles compared with controls. Right-censoring was due to either death (15%, N=11,165) or lost to follow-up (1%, N=735).

Hazard ratio estimates for PRS and CRC in the independent replication (Table 4) were consistent with findings from the discovery dataset (Table 2), overall (aged <50 years, HR per SD of PRS: 1.73; 95% CI: 1.17, 2.56; aged ≥50 years, HR per SD of PRS: 1.43; 95% CI: 1.34, 1.51) and for individuals who reported a negative first-degree family history of CRC (aged <50 years, HR per SD of PRS: 1.76; 95% CI: 1.11, 2.78; aged ≥50 years, HR per SD of PRS: 1.42; 95% CI: 1.33, 1.52). Although the effects seen for younger and older individuals were consistent with our primary analysis, the specific evaluation of whether these effects differ by age (<50 vs. age ≥50

years) was underpowered in RPGEH, due to the limited number of early-onset CRC cases in this cohort. Numbers of early-onset CRC among individuals with a first-degree family history of CRC in the replication dataset were too few for a meaningful interpretation of the analysis. Kaplan-Meier survival plots, stratified by family history, are displayed in Figure 2, consistent with the hypothesized PRS-related probability gradients across the full age range.

Discussion

Our study, including more than 50,000 CRC cases and 50,000 controls, demonstrated that a PRS, derived from common genetic variants, successfully identifies participants at increased risk for early-onset CRC, particularly among individuals without a family history of CRC; additionally, the PRS was more strongly associated with early-onset cancer compared with late-onset CRC. The PRS-associated risks were found for early-onset cancer of the proximal and distal colon, and the rectum, with a modest increased propensity for the non-proximal cancers. We confirmed the overall findings for early-onset CRC in a sub-study from Ohio, where Lynch syndrome cases were excluded from the analysis. The results from these case-control studies were also supported by a smaller, prospective study that showed increased PRS-associated risks for early-onset CRC, particularly in those negative for CRC family history. Our findings may have important clinical relevance, as they could contribute, along with other lifestyle and environmental risk factors, to tailored screening in people aged <50 years who are currently not targeted for early detection and for whom CRC rates have increased over the last decades.

The development of a PRS to evaluate the overall predictive power of common risk loci for CRC has previously been carried out;[29-31] however, few studies evaluated specifically for association of common polymorphisms with early-onset CRC.[32-36] These smaller studies, involving 10 to 33 SNPs, pointed to some individual loci differentially associated with early-onset CRC; however,

497 our much larger study, which included 95 loci identified from GWAS (Table S2), showed that
498 risks related to an individual's cumulative genetic risk profile for at-risk alleles, as reflected in
499 the PRS, were much greater than the contributions of individual SNPs. A caveat to using these
500 95 variants in a PRS intended for discriminating early-onset CRC risk is that they are produced
501 from GWAS analyses not specific to early-onset disease; adequately powered GWAS analyses
502 specific for early-onset CRC have yet to be performed. Therefore, although our PRS positively
503 identifies those at heightened risk for early-onset CRC, there is still room for improving its
504 discriminatory accuracy. Furthermore, combining a genetic PRS with lifestyle and environmental
505 risk factors could potentially contribute to even greater precision in identification of individuals
506 who may benefit from earlier onset CRC screening.[37]

507 Given that early-onset CRC is increasing in incidence and is commonly diagnosed at later stages,
508 which carries a poorer prognosis, recommendations have been made to lower the screening age
509 to 45 for individuals at average-risk.[12] Consideration of early detection for early-onset cancer is
510 dependent, however, on a number of factors, including differentials in CRC risk in absolute
511 terms, projected benefits, potential harms such as colonic perforation, and costs; therefore,
512 potentially tempering some enthusiasm for lowering the CRC screening age and calling for
513 identification of high-risk groups for more targeted early detection.[16, 38, 39] Our study highlights
514 the potential utility of a PRS in CRC risk stratification for people <50 years of age, which might
515 inform precision cancer screening in this population that currently lacks consistent early
516 detection recommendations, particularly for those without a family history of CRC.

517 This study is unique in the large size of the study population, particularly for those <50 years of
518 age, allowing for evaluation of PRS-related risks overall, and by family history, refined age
519 groups, and tumor site. Major results for association of the PRS with early-onset cancer were

also replicated in an independent community-based cohort, although the number of early-onset cases in that cohort was limited. Limitations of our study include the lack of CRC family history information on a substantial subset of study participants; however, missingness was defined by study and therefore unlikely to introduce bias. Also, our PRS was generated and validated in individuals of European ancestry, currently limiting its applicability for different ancestral groups, until a PRS is developed and validated in diverse populations. Another limitation is that we did not systematically take into account the genetic mutations related to Lynch and other rarer hereditary cancer syndromes;[23, 34, 40-42] however, our sensitivity analysis, in the Ohio investigations where this information was systematically assessed, indicated that risks associated with PRS remained very similar after the removal of Lynch cases from the analysis. Nevertheless, further research is needed on the combined utility for risk prediction of rare and common variants in those with or without a family history of CRC as it can be expected that accounting for both PRS and high penetrance genes will further improve risk stratification.[43, 44] There remains more to be discovered about the genetics of CRC, particularly for early-onset disease, as substantial heritability for CRC remains unexplained and genetic effects are typically stronger for early-onset diesase.[45, 46] As more risk loci will be discovered, the predictive power of the PRS is expected to further improve, and to be tested in clinical trials.

In conclusion, we demonstrated that a PRS, derived from common genetic variants, successfully stratifies individuals for early onset CRC based on genetic risk, particularly among individuals who report a negative first-degree family history of CRC. Furthermore, the associations between the PRS and CRC are greater for young-onset than for older-onset disease. The PRS may contribute, along with lifestyle and environmental risk profiling, toward prioritizing individuals at increased susceptibility to early-onset CRC for personalized screening regimens or other

intervention strategies. Early-onset CRC is increasing in the US and elsewhere; by selecting high-risk individuals <50 years of age, we can reduce the burden on early detection programs and potentially provide more individualized prevention approaches.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7-30.
2. Phillips KA, Liang SY, Ladabaum U, et al. Trends in colonoscopy for colorectal cancer screening. Med Care 2007;45:160-7.
3. Cress RD, Morris C, Ellison GL, et al. Secular changes in colorectal cancer incidence by subsite, stage at diagnosis, and race/ethnicity, 1992-2001. Cancer 2006;107:1142-52.
4. Siegel RL, Torre LA, Soerjomataram I, et al. Global patterns and trends in colorectal cancer incidence in young adults. Gut 2019:gutjnl-2019-319511.
5. Yeo H, Betel D, Abelson JS, et al. Early-onset colorectal cancer is distinct from traditional colorectal cancer. Clin Colorectal Cancer 2017;16:293-299.e6.
6. Bailey CE, Hu CY, You YN, et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. JAMA Surg 2015;150:17-22.
7. Murphy CC, Singal AG, Baron JA, et al. Decrease in incidence of young-onset colorectal cancer before recent increase. Gastroenterology 2018;155:1716-1719.e4.
8. Feletto E, Yu XQ, Lew J-B, et al. Trends in colon and rectal cancer incidence in australia from 1982 to 2014: analysis of data on over 375,000 cases. Cancer Epidemiology Biomarkers & Prevention 2018.
9. Brenner DR, Ruan Y, Shaw E, et al. Increasing colorectal cancer incidence trends among younger adults in Canada. Prev Med 2017;105:345-349.
10. Siegel RL, Jemal A, Ward EM. Increase in incidence of colorectal cancer among young men and women in the United States. Cancer Epidemiol Biomarkers Prev 2009;18:1695-8.
11. Knudsen AB, Zauber AG, Rutter CM, et al. Estimation of benefits, burden, and harms of colorectal cancer screening strategies: modeling study for the US Preventive Services Task Force. Jama 2016;315:2595-609.
12. Wolf AMD, Fontham ETH, Church TR, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA: A Cancer Journal for Clinicians 2018;68:250-281.
13. Rex DK, Boland CR, Dominitz JA, et al. Colorectal cancer screening: recommendations for physicians and patients from the U.S. Multi-Society Task Force on colorectal cancer. Gastroenterology 2017;153:307-323.
14. Corley DA, Peek RM, Jr. When should guidelines change? A clarion call for evidence regarding the benefits and risks of screening for colorectal cancer at earlier ages. Gastroenterology 2018;155:947-949.
15. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. JAMA 2016;315:2564-2575.
16. Liang PS, Allison J, Ladabaum U, et al. Potential intended and unintended consequences of recommending initiation of colorectal cancer screening at age 45 years. Gastroenterology 2018;155:950-954.
17. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nature Genetics 2018.
18. Schumacher FR, Schmit SL, Jiao S, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. Nat Commun 2015;6:7138.
19. Peters U, Jiao S, Schumacher FR, et al. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. Gastroenterology 2013;144:799-807.e24.

20. Schmit SL, Edlund CK, Schumacher FR, et al. Novel common genetic susceptibility loci for colorectal cancer. J Natl Cancer Inst 2018.
21. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48:1279-83.
22. Zhong H, Prentice RL. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. Biostatistics 2008;9:621-34.
23. Pearlman R, Frankel WL, Swanson B, et al. Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. JAMA Oncology 2017;3:464-471.
24. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N Engl J Med 2005;352:1851-60.
25. Hampel H, Frankel WL, Martin E, et al. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. J Clin Oncol 2008;26:5783-8.
26. Banda Y, Kvale MN, Hoffmann TJ, et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. Genetics 2015;200:1285-95.
27. Kvale MN, Hesselson S, Hoffmann TJ, et al. Genotyping informatics and quality control for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. Genetics 2015;200:1051-60.
28. Hoffmann TJ, Kvale MN, Hesselson SE, et al. Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. Genomics 2011;98:79-89.
29. Dunlop MG, Tenesa A, Farrington SM, et al. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. Gut 2013;62:871-81.
30. Jenkins MA, Makalic E, Dowty JG, et al. Quantifying the utility of single nucleotide polymorphisms to guide colorectal cancer screening. Future Oncol 2016;12:503-13.
31. Hsu L, Jeon J, Brenner H, et al. A model to determine colorectal cancer risk using common genetic susceptibility loci. Gastroenterology 2015;148:1330-9.e14.
32. He J, Wilkens LR, Stram DO, et al. Generalizability and epidemiologic characterization of eleven colorectal cancer GWAS hits in multiple populations. Cancer Epidemiol Biomarkers Prev 2011;20:70-81.
33. von Holst S, Picelli S, Edler D, et al. Association studies on 11 published colorectal cancer risk loci. Br J Cancer 2010;103:575-80.
34. Giráldez MD, López-Dóriga A, Bujanda L, et al. Susceptibility genetic variants associated with early-onset colorectal cancer. Carcinogenesis 2012;33:613-619.
35. Song N, Shin A, Park JW, et al. Common risk variants for colorectal cancer: an evaluation of associations with age at cancer onset. Sci Rep 2017;7.
36. Middeldorp A, Jagmohan-Changur S, van Eijk R, et al. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol Biomarkers Prev 2009;18:3062-7.
37. Jeon J, Du M, Schoen RE, et al. Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. Gastroenterology 2018;154:2152-2164.e19.

38. Murphy CC, Sanoff HK, Stitzenberg KB, et al. RE: Colorectal cancer incidence patterns in the United States, 1974–2013. JNCI: Journal of the National Cancer Institute 2017;109:djx104-djx104.
39. Warren JL, Klabunde CN, Mariotto AB, et al. Adverse events after outpatient colonoscopy in the Medicare population. Ann Intern Med 2009;150:849-57, w152.
40. Jasperson KW, Tuohy TM, Neklason DW, et al. Hereditary and familial colon cancer. Gastroenterology 2010;138:2044-58.
41. Pinto C, Veiga I, Pinheiro M, et al. MSH6 germline mutations in early-onset colorectal cancer patients without family history of the disease. Br J Cancer 2006;95:752-6.
42. de Voer RM, Hahn MM, Mensenkamp AR, et al. Deleterious germline BLM mutations and the risk for early-onset colorectal cancer. Sci Rep 2015;5:14060.
43. Whiffin N, Dobbins SE, Hosking FJ, et al. Deciphering the genetic architecture of low-penetrance susceptibility to colorectal cancer. Hum Mol Genet 2013;22:5075-82.
44. Wray NR, Purcell SM, Visscher PM. Synthetic associations created by rare variants do not explain most GWAS results. PLoS Biol 2011;9:e1000579.
45. Jiao S, Peters U, Berndt S, et al. Estimating the heritability of colorectal cancer. Hum Mol Genet 2014;23:3898-905.
46. Zaitlen N, Kraft P. Heritability in the genome-wide association era. Hum Genet 2012;131:1655-64.

TABLES

Table 1: Baseline study characteristics of the discovery and replication datasets

	Discovery dataset				Replication dataset			
	Cases (N=50,023)		Controls (N=58,039)		All participants		CRC Cases	
	<50 Years-Old	≥50 Years-Old	<50 Years-Old	≥50 Years-Old	Eligible cohort	CRC cases	<50 Years-Old	≥50 Years-Old
N	5479	44544	6718	51321	72573	1093	25	1068
Age, Mean (SD)	43.1 (5.6)	66.5 (8.7)	41.3 (7.2)	65.3 (8.3)	71.5 (13.1)	73.1 (10.8)	45.2 (3.3)	73.7 (10.1)
Sex, N (%)								
Male	2767 (50.5)	24145 (54.2)	3272 (48.7)	26886 (52.4)	30160 (41.6)	526 (48.1)	9 (36.0)	517 (48.4)
Female	2706 (49.4)	20336 (45.7)	3446 (51.3)	24435 (47.6)	42413 (58.4)	567 (51.9)	16 (64.0)	551 (51.6)
Missing	6 (0.1)	63 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Family History of CRC, N (%)								
Yes	944 (17.2)	5558 (12.5)	578 (8.6)	5330 (10.4)	6956 (9.6)	204 (18.7)	7 (28.0)	197 (18.4)
No	3159 (57.7)	24028 (53.9)	4130 (61.5)	28317 (55.2)	65617 (90.4)	889 (81.3)	18 (72.0)	871 (81.6)
Missing	1376 (25.1)	14958 (33.6)	2010 (29.9)	17674 (34.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Tumor Site, N (%)								
Proximal Colon	1231 (22.5)	12978 (29.1)	--	--	--	--	--	--
Distal Colon	1442 (26.3)	12036 (27.0)	--	--	--	--	--	--
Rectum	1920 (35.0)	12918 (29.0)	--	--	--	--	--	--
Missing	886 (16.2)	6612 (14.8)	--	--	--	--	--	--
PRS, N (%)								
Quartile 1	693 (12.6)	6227 (14.0)	1659 (24.7)	12863 (25.1)	18175 (25.0)	163 (14.9)	2 (8.0)	161 (15.1)
Quartile 2	1048 (19.1)	8824 (19.8)	1666 (24.8)	12848 (25.0)	18150 (25.0)	232 (21.2)	4 (16.0)	228 (21.3)
Quartile 3	1396 (25.5)	11877 (26.7)	1674 (24.9)	12824 (25.0)	18132 (25.0)	287 (26.3)	7 (28.0)	280 (26.2)
Quartile 4	2342 (42.7)	17616 (39.5)	1719 (25.6)	12786 (24.9)	18116 (25.0)	411 (37.6)	12 (48.0)	399 (37.4)

Table 2: Risk estimates for early-onset versus late-onset CRC associated with a 95-SNP PRS in the discovery dataset[a]

PRS	N (cases)	N (controls)	OR (95% CI)	*P* value	*P* value for interaction[b]
All Subjects					0.0137
<50 Years-Old					
per 1 SD	5479	6718	1.64 (1.57, 1.72)	6.00E-107	
Quartile 1 (ref)	693	1659	1.00		
Quartile 2	1048	1666	1.64 (1.43, 1.89)	2.07E-12	
Quartile 3	1396	1674	2.19 (1.91, 2.50)	2.17E-30	
Quartile 4	2342	1719	3.73 (3.28, 4.24)	1.13E-89	
≥50 Years-Old					
per 1 SD	44544	51321	1.52 (1.50, 1.54)	< 2.23E-308	
Quartile 1 (ref)	6227	12863	1.00		
Quartile 2	8824	12848	1.45 (1.39, 1.51)	8.55E-62	
Quartile 3	11877	12824	1.95 (1.87, 2.03)	1.37E-208	
Quartile 4	17616	12786	2.92 (2.80, 3.04)	< 2.23E-308	
Negative Family History					5.61E-05
<50 Years-Old					
per 1 SD	3159	4130	1.74 (1.65, 1.84)	1.33E-81	
Quartile 1 (ref)	388	1085	1.00		
Quartile 2	601	1025	1.66 (1.39, 1.98)	1.58E-08	
Quartile 3	820	1001	2.46 (2.07, 2.92)	3.37E-25	
Quartile 4	1350	1019	4.26 (3.61, 5.01)	3.65E-67	
≥50 Years-Old					
per 1 SD	24028	28317	1.50 (1.47, 1.53)	< 2.23E-308	
Quartile 1 (ref)	3529	7341	1.00		
Quartile 2	4869	7083	1.44 (1.36, 1.53)	1.85E-36	
Quartile 3	6494	7058	1.92 (1.82, 2.03)	6.17E-119	
Quartile 4	9136	6835	2.85 (2.70, 3.00)	< 2.23E-308	
Positive Family History					0.0028
<50 Years-Old					
per 1 SD	944	578	1.19 (1.05, 1.35)	0.0063	
Quartile 1 (ref)	133	105	1.00		
Quartile 2	203	133	1.58 (1.05, 2.36)	0.0265	
Quartile 3	208	152	1.22 (0.82, 1.83)	0.3277	
Quartile 4	400	188	1.70 (1.17, 2.47)	0.0052	
≥50 Years-Old					
per 1 SD	5558	5330	1.42 (1.36, 1.48)	7.02E-57	
Quartile 1 (ref)	690	1134	1.00		
Quartile 2	1037	1264	1.42 (1.24, 1.63)	5.85E-07	
Quartile 3	1478	1343	1.81 (1.59, 2.07)	8.44E-19	
Quartile 4	2353	1589	2.47 (2.18, 2.79)	2.70E-45	

[a]The logistic regression models include age, sex, principal components, genotype platform, and polygenic risk score.

[b]*P* value produced from interaction term with continuous PRS (per SD) and age (<50 versus ≥50 years).

Table 3: Risk estimates for early-onset versus late-onset CRC associated with a 95-SNP PRS among participants with and without Lynch Syndrome, in the Ohio cohort[a]

PRS per 1 SD	N (cases)	N (controls)	OR (95% CI)	*P* value	*P* value for interaction[b]
Including Lynch and Non-Lynch Cases					
All Subjects					0.0369
<50 Years-Old	574	979	1.73 (1.54, 1.95)	1.39E-19	
≥50 Years-Old	2525	1463	1.47 (1.37, 1.58)	1.77E-28	
Negative Family History					0.0106
<50 Years-Old	449	931	1.81 (1.59, 2.07)	9.64E-19	
≥50 Years-Old	1885	1271	1.45 (1.34, 1.56)	1.16E-21	
Positive Family History					0.1517
<50 Years-Old	106	48	1.28 (0.84, 1.97)	0.2530	
≥50 Years-Old	565	192	1.55 (1.30, 1.84)	1.12E-06	
Excluding Lynch Cases					
All Subjects					0.0149
<50 Years-Old	537	979	1.82 (1.61, 2.06)	2.63E-21	
≥50 Years-Old	2471	1463	1.49 (1.39, 1.60)	1.11E-29	
Negative Family History					0.0107
<50 Years-Old	438	931	1.83 (1.60, 2.09)	7.50E-19	
≥50 Years-Old	1856	1271	1.46 (1.35, 1.57)	4.30E-22	
Positive Family History					0.5627
<50 Years-Old	80	48	1.53 (0.98, 2.41)	0.0635	
≥50 Years-Old	540	192	1.61 (1.34, 1.92)	2.34E-07	

[a]The logistic regression models include age, sex, principal components, and polygenic risk score.
[b]*P* value produced from interaction term with continuous PRS (per SD) and age (<50 versus ≥50 years).

Table 4: Risk estimates for early-onset versus late-onset CRC associated with a 95-SNP PRS in the RPGEH replication cohort[a]

PRS	N in eligible cohort	N (cases)	HR (95% CI)	*P* value	*P* value for interaction[b]
All Subjects					0.3291
<50 Years-Old					
per 1 SD	26983	25	1.73 (1.17, 2.56)	0.0056	
≥50 Years-Old					
per 1 SD	67792	1068	1.43 (1.34, 1.51)	2.77E-31	
Negative Family History					0.3681
<50 Years-Old					
per 1 SD	24472	18	1.76 (1.11, 2.78)	0.0161	
≥50 Years-Old					
per 1 SD	61129	871	1.42 (1.33, 1.52)	2.85E-25	
Positive Family History					0.6920
<50 Years-Old					
per 1 SD	2511	7	1.56 (0.75, 3.26)	0.2334	
≥50 Years-Old					
per 1 SD	6668	202	1.34 (1.17, 1.54)	2.87E-05	

[a]The Cox models include sex, principal components, and polygenic risk score.
[b]*P* value produced from interaction term with continuous PRS (per SD) and age (<50 versus ≥50 years).

FIGURE LEGENDS

Figure 1: Risk estimates for early-onset versus late-onset CRC associated with a 95-SNP PRS in the discovery dataset. (A) Model includes all study participants regardless of first-degree family history of CRC. (B) Model includes study participants without a first-degree family history of CRC. (C) Model includes study participants with a first-degree family history of CRC. Models were adjusted for age, sex, principal components, genotype platform, and polygenic risk score. The interaction p-value reported was produced from a model including an interaction term with a continuous PRS (per SD) and age (<50 years versus ≥50 years).

Figure 2: Absolute risk estimates of being diagnosed with CRC across the age stratum by PRS percentile among individuals in the RPGEH cohort. (A) Among individuals with a first-degree relative with CRC. (B) Among individuals without a family history of CRC.

Figure S1: Distribution of the PRS across cases and controls. (A) Plot includes all cases and controls with a CRC diagnosis at <50 years of age. (B) Plot includes all cases and controls with a CRC diagnosis at ≥50 years of age.

Figure S2: Risk estimates for early-onset versus late-onset CRC associated with a 95-SNP PRS across disease site among participants with a negative family history of CRC in the discovery dataset. (A) Model includes all cases with CRC diagnosis within the proximal colon. (B) Model includes all cases with CRC diagnosis within the distal colon. (C) Model includes all cases with CRC diagnosis within the rectum. Models were adjusted for age, sex, principal components, genotype platform, and polygenic risk score. The interaction p-value reported was produced from a model including an interaction term with a continuous PRS (per SD) and age (<50 years versus ≥50 years).