DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Genetic, metabolomic, and lifestyle dynamics in colorectal cancer and chronic disease prevention

Citation

Bever, Alaina. 2024. Genetic, metabolomic, and lifestyle dynamics in colorectal cancer and chronic disease prevention. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37377927

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at
https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

HARVARD UNIVERSITY

Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Committee on Higher Degrees in Population Health Sciences,
have examined a dissertation, entitled,

**"Genetic, Metabolomic, and Lifestyle Dynamics
in Colorectal Cancer and Chronic Disease Prevention"**

presented by

ALAINA M. BEVER

candidate for the degree of Doctor of Philosophy
and hereby certify that it is worthy of acceptance.

Mingyang Song
Mingyang Song (Jan 17, 2024 20:13 EST)

Dr. Mingyang Song, M.D., Sc.D., Committee Chair, Harvard T.H. Chan School of Public Health

A. Heather Eliassen
A. Heather Eliassen (Jan 19, 2024 14:53 EST)

Dr. A. Heather Eliassen, Sc.D., Harvard T.H. Chan School of Public Health, Harvard Medical School

Liming Liang
Liming Liang (Jan 20, 2024 16:30 EST)

Dr. Liming Liang, Ph.D., Harvard T.H. Chan School of Public Health

Meir Stampfer
Meir Stampfer (Jan 20, 2024 16:41 EST)

Dr. Meir J. Stampfer, M.D., Dr.P.H., Harvard Medical School, Harvard T.H. Chan School of Public Health

Date: 19 January 2024

Genetic, metabolomic, and lifestyle dynamics in colorectal cancer and chronic disease prevention

Alaina M Bever

A dissertation presented to
the Department of Population Health Sciences
and
the Department of Epidemiology
in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Population Health Sciences (Epidemiology)

Harvard University
Cambridge, Massachusetts

January 2024

© 2024 Alaina M Bever

All rights reserved.

Dissertation Advisor: Professor Mingyang Song Alaina M Bever

Genetic, metabolomic, and lifestyle dynamics in colorectal cancer and chronic disease prevention

Abstract

Noncommunicable chronic disease is a group of conditions that includes type 2 diabetes, cardiovascular disease, and cancer, and accounts for the majority of total deaths in the United States. Many of these diseases are associated with overweight and obese body mass index, however, the pathways linking genetics, modifiable factors, and physiologic characteristics such as inflammation and insulin resistance to chronic disease development are not well understood. A more nuanced understanding of how excess adiposity and related risk factors influence chronic disease is needed to develop improved prevention strategies. We leveraged data from three prospective cohort studies, the Nurses' Health Study, Nurses' Health Study II, and Health Professionals Follow-up Study, to explore these risk factors in relation to chronic disease.

In **Chapter 1**, we evaluated the association of genetically predicted BMI with different colorectal cancer precursor lesions: conventional adenomas and serrated polyps. We found similar associations between measured or genetic BMI and colorectal cancer precursors and observed a stronger association of both measured and genetic BMI with serrated polyps and synchronous serrated polyp and conventional adenoma, compared to conventional adenoma alone.

In **Chapter 2**, we developed metabolomic signatures of inflammation and metabolic disturbance and evaluated the association of the signatures with colorectal cancer. We found that both signatures predicted colorectal cancer risk among men, but not women, and we

identified promising individual metabolites linking the states of inflammation and metabolic dysregulation to colorectal cancer development.

In **Chapter 3**, we estimated the proportion and number of cases of chronic disease attributable to poor quality diet and inadequate physical activity in the United States, accounting for the effects of these risk factors that is mediated by changes in body weight. We found that 18 percent of chronic disease in 2018 was attributable to poor quality diet and physical activity, with 13 percent not mediated by effects on change in body weight.

TABLE OF CONTENTS

Acknowledgements

Thank you to the 290,000 participants in the Nurses Health Studies and Health Professionals Follow-up Study – without your contributions, none of this work would have been possible. Thank you to my thesis advisors; Mingyang Song, Meir Stampfer, Heather Eliassen, and Liming Liang; and to all my research mentors and co-authors who contributed to this work and who taught and continue to teach me invaluable skills for research and life. Thank you to the Harvard TH Chan School of Public Health and Harvard-MIT MD-PhD program administrative teams for all of your support and guidance. Thank you to Loren Walensky, Deborah Blacker, and Richard Mitchell for your mentorship, advice, and support. Thank you to all the incredible instructors who taught me the epidemiology and statistical content used to produce this work. Thank you to Frank Shih, Ian Suydam, Kim Woodrow, Katherine Grantz, Enrique Schisterman, and Sunni Mumford for introducing me to research, equipping me with research skills, and encouraging me to pursue a career in academia. Thank you to my family, to my classmates and friends, and especially to Enrique Toloza for all of your support.

CHAPTER ONE

Genetic obesity variants and risk of conventional adenomas and serrated polyps

Alaina M Bever, Dong Hang, Xiaosheng He, Amit D Joshi, Ming Ding, Kana Wu, Andrew T Chang, Edward L Giovannucci, and Mingyang Song

ABSTRACT

Background: Higher body mass index (BMI) is associated with increased risk of colorectal cancer. How genetically predicted BMI may be associated with colorectal cancer precursors is unknown.

Aims: Our objective was to quantify the association of genetically predicted and measured BMI with risk of colorectal cancer precursors.

Methods: We evaluated the association of genetically predicted and measured BMI with risk of conventional adenomas, serrated polyps, and synchronous polyps among 27,426 participants who had undergone at least one lower gastrointestinal endoscopy in the Nurses' Health Study, Nurses' Health Study II and Health Professionals Follow-up Study. Genetic risk score was derived from 97 BMI-related single nucleotide polymorphisms. Multivariable logistic regression evaluated each polyp subtype compared to non-polyps.

Results: For conventional adenomas, the OR per 2-kg/m^2 increase was 1.03 (95% CI, 1.01-1.04) for measured BMI and 0.98 (95% CI, 0.88-1.10) for genetically predicted BMI; for serrated polyps, the OR was 1.06 (95% CI, 1.04-1.08) and 1.04 (95% CI, 0.90-1.20), respectively; for synchronous polyps, the OR was 1.10 (95% CI, 1.07-1.13) and 1.09 (95% CI, 0.89-1.34), respectively. Genetically predicted BMI was associated with synchronous polyps in women (OR=1.37, 95% CI: 1.05-1.79).

Conclusion: Genetically predicted BMI was not associated with colorectal cancer precursor lesions. The confidence intervals were wide and encompassed those for measured BMI, indicating that null findings may be due to insufficient power.

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer in the United States and is associated with modifiable risk factors including obesity, dietary factors, and smoking.[1] Although the observed association between increased body mass index (BMI) and CRC risk has been fairly consistent in prospective cohort studies, the association may vary among CRC subtypes and by biological sex. For example, in observational studies, BMI appears to be more strongly associated with colorectal cancer risk in men than women.[1–3] Recently, distinct subtypes of CRC have been established, and these subtypes are thought to arise from distinct precursor lesions: conventional adenomas and serrated polyps. In one study, for example, BMI showed a much stronger association with serrated polyps than with conventional adenomas.[4] Other studies suggest that BMI is more strongly associated with microsatellite-instable CRC (compared to microsatellite-stable), which is more likely to develop from serrated polyps than conventional adenomas.[5]

Evidence for the association between increased BMI and polyp subtypes is limited to observational studies. Mendelian randomization provides insight on disentangling the effects of predisposed obesity versus acquired obesity on CRC risk. We previously reported that genetic risk of obesity, derived from 97 established adult BMI-associated variants, correlated with BMI across all ages.[6] Genetic risk of BMI has been associated with risk of CRC using Mendelian randomization.[7] When stratified by sex, genetic risk of BMI was associated with CRC in women only, in contrast to studies of measured BMI.[7] The association between genetic risk of adulthood obesity and CRC precursor lesions has not yet been assessed and has the potential to improve the understanding of the influence of increased adiposity on heterogeneous CRC pathways.

The objective of this study was to evaluate the relationship between genetically predicted BMI and two distinct CRC precursor lesions (conventional adenomas and serrated polyps). We created a genetic risk score (GRS) using 97 previously identified SNPs that were associated with adult BMI. The association between GRS and risk of conventional adenomas or

serrated polyps was evaluated among participants with genetic data in three large cohort studies: Nurses' Health Study (NHS), Nurses' Health Study II (NHS2), and Health Professionals Follow-up Study (HPFS). For comparison, we also assessed the associations for measured BMI in the same set of samples. To test for potential sex difference, we also performed the analysis in men and women separately.

MATERIALS AND METHODS

Study population

The NHS enrolled 121,700 registered female nurses, aged 30-55 (at enrollment), in 1976; the NHS2 enrolled 116,686 registered female nurses, aged 25-42, in 1989; and the HPFS enrolled 51,529 male health professionals, aged 40-75, in 1986. A subset of participants provided blood specimens: 32,826 NHS participants, between 1989 and 1990; 29,611 NHS2 participants, between 1996 and 1999; and 18,225 HPFS participants, between 1993 and 1995. Blood specimens were returned on ice packs by overnight courier.

The present study was limited to 37,661 participants with genetic information available from previous, nested genome-wide association studies (NHS, 18,498; NHS2, 8,274; HPFS, 10,889). After excluding participants who were of non-European origin; had a history of cancer (except non-melanoma skin cancer), colorectal polyps, or inflammatory bowel disease; or who had no reported endoscopies of the lower GI tract; a total of 27,436 participants (NHS, 13,103; NHS2, 6,494; HPFS, 7,839) were included in the present analysis.

The collection of detailed histological information on colorectal polyps began in 1992 for the NHS and HPFS, and 1991 for the NHS2; these years were used as baseline for the present study. Eligible participants were followed until first colorectal polyp diagnosis, death, or end of follow-up for the present study: June 1, 2012 for NHS, June 1, 2011 for NHS2, and January 1, 2010 for the HPFS.

Computation of GRS

Genotype data were derived from various nested studies within the NHS, NHS2, and HPFS studies. Details of genotyping and imputation of SNPs included in the GRS have been described elsewhere in detail.[8] Weighted GRS was calculated from 97 SNPs and relative effect size (β coefficient) identified as associated with adult BMI in the most recent genome-wide association study (GWAS).[9] GRS was calculated using established methods, with an assigned value of 0, 1, or 2 for the number of risk alleles and the following equation: GRS=$\left(\sum_{i=1}^{97}\beta_i * SNP_i\right) * \left(97/\sum_{i=1}^{97}\beta_i\right)$, where β_i is the regression coefficient identified in the GWAS. Sex-specific GRS was calculated using the sex-specific β_i.[9]

Assessment of exposure variables

Measured BMI was calculated as the cumulative average of BMI based on self-reported height at baseline and self-reported weight from enrollment to polyp diagnosis or end of follow-up. Weight reported on the NHS questionnaires was validated among 140 NHS participants; self-reported and measured weights were highly correlated (r = 0.97), indicating that self-reported weight measurements are reasonably valid. Lifestyle characteristics were assessed by questionnaire at baseline and biennially via follow-up questionnaires. Food frequency questionnaires were administered every four years to assess dietary risk factors. Missing data for covariates during follow-up questionnaires was carried forward from the most recent available data.

Assessment of outcomes

Every two years via follow-up questionnaire, participants reported whether they had undergone a colonoscopy or sigmoidoscopy, and whether any colorectal polyps had been diagnosed, during that two-year period. If a polyp diagnosis was reported, we requested permission to

obtain the endoscopy and pathology reports. Investigators who were blinded to participants' exposures and genetic data reviewed the medical records and confirmed polyp diagnoses. Relevant clinical and pathological data were also extracted from the medical records.[4]

In the present study, conventional adenomas included tubular, tubulovillous, and villous adenomas and adenomas with high-grade dysplasia. Serrated polyps included hyperplastic polyps and mixed/serrated adenomas. Mixed/serrated adenomas consisted of both mixed polyps (those with both adenomatous and hyperplastic changes in histology) and polyps with any serrated diagnosis (e.g. serrated adenomas, serrated polyps, and sessile serrated adenoma/polyp). Diagnosis of serrated polyp and conventional adenoma at the same endoscopy was considered a synchronous polyp.

Statistical analyses

All analyses were conducted using three pooled cohorts and repeated separately in women (NHS, NHS2) and men (HPFS) to assess potential differences in the relationship between BMI and CRC risk by biologic sex. We evaluated measured BMI and genetically predicted BMI in relation to risk of polyp type (non-polyps, conventional adenoma only, serrated polyp only, and synchronous polyps). GRS was regressed on measured BMI to derive the change in GRS associated with a 2 kg/m^2 change in measured BMI.

Multivariable logistic regression was used to evaluate the risk of conventional, serrated, or synchronous polyps in relation to genetically predicted and measured BMI, compared to non-polyps. We also modeled the risk of conventional adenoma subtypes (non-advanced or advanced) and serrated polyp subtypes (small, <10 mm; large, ≥10 mm) compared to non-polyps. Advanced conventional adenomas were defined as having at least 1 conventional adenoma of 10 mm or greater in diameter or with advanced histology (tubulovillous/villous histologic features or high-grade or severe dysplasia).

Effect modification by biologic sex (i.e., study cohort) was evaluated using Wald test for the product term. Heterogeneity among the polyp subtypes was assessed using multivariate regression in case-only analyses. To account for multiple records per participant and to handle time-varying covariates efficiently, we used an Andersen-Gill data structure with a new record for each 2-year follow-up period during which a participant underwent an endoscopy.

In a secondary analysis, the models were repeated with conventional adenomas, serrated polyps, and their subtypes with the inclusion of synchronous polyps in each subtype grouping. We also modeled the risk of polyp location (proximal colon, distal colon, or rectum) for conventional adenomas and serrated polyps. Polyps were classified as proximal if they were removed from the cecum to the transverse colon, distal if they were removed from the splenic flexure to the sigmoid colon, and rectal if they were removed from the rectosigmoid junction to the anal canal (excluding anal squamous cell carcinoma). Groupings by polyp location were not mutually exclusive; for example, one participant could have contributed both a proximal and distal conventional adenoma, or a proximal conventional adenoma and distal serrated polyp. In another secondary analysis, we assessed the polyp associations for each of the 97 individual SNPs included in the GRS. Bonferroni correction for multiple testing (adjusted $\alpha = 0.05/97 = 5.2 \times 10^{-4}$) was applied in this analysis.

Genetically predicted BMI models were adjusted for the following characteristics: age, study cohort (NHS, NHS2, HPFS; in pooled models only), time period of endoscopy (in 2-year intervals), reason for endoscopy (screening or symptoms), number of previous endoscopies, time in years since most recent endoscopy, and top three principal components for population structure (in order to account for systematic ancestry differences).[10] Measured BMI models were adjusted for age, study cohort, time period of endoscopy, reason for endoscopy, number of previous endoscopies, time in years since most recent endoscopy, family history of colorectal cancer (yes or no), height, alcohol intake, regular aspirin use (yes or no), and physical activity. All covariates were treated as continuous variables unless otherwise noted.

RESULTS

Among 27,436 participants with genetic data and outcomes available, there were a total of 2,952 participants with conventional adenomas only, 1,589 with serrated polyps only, and 790 with synchronous polyps. Genetically predicted BMI, measured BMI, and endoscopy-related characteristics are shown in **Table 1.1**. Participants with serrated polyps were more likely to be women. Serrated polyps and synchronous polyps both tracked with higher measured BMI. Participants with no polyps had a higher average number of endoscopies, longer time since most recent previous period endoscopy, and were less likely to report symptoms as the reason for endoscopy.

The median measured BMI was 1.7 kg/m^2 greater at the highest vs. lowest quintile of BMI genetic risk score (26.3 kg/m^2 compared to 24.6 kg/m^2), with a wide interval of measured BMI at each quintile (**Figure 1.1**). An 18-unit increase in GRS was correlated with 2 kg/m^2 increase in measured BMI (R-squared= 0.02).

Measured BMI was significantly associated with risk of conventional adenomas (OR per 2 kg/m^2 increase=1.03, 95% CI: 1.01-1.04), serrated polyps (OR=1.06, 95% CI: 1.04-1.08), and synchronous polyps (OR=1.10, 95% CI: 1.07-1.13) (**Figure 1.2 and Appendix Table 1.1**). Among conventional adenoma subtypes, measured BMI was associated with increased risk of non-advanced adenomas (OR=1.03, 95% CI: 1.01-1.05) and advanced adenomas (OR=1.02, 95% CI: 0.99-1.04). Among serrated polyp subtypes, measured BMI was associated with increased risk of small polyps (OR=1.06, 95% CI: 1.04-1.08) and large polyps (OR=1.03, 95% CI: 0.96-1.11), although the large serrated polyps group had a smaller sample size and confidence interval included unity (**Appendix Table 1.1**).

Table 1.1. BMI and endoscopy characteristics of pooled study participants from three cohort studies (NHS, NHS2, HPFS) by polyp diagnosis

	Polyp Type			
	Non-Polyps (n=22 105)	**Conventional Adenoma Only (n=2952)**	**Serrated Polyp Only (n=1589)**	**Synchronous Conventional Adenoma and Serrated Polyp (n=790)**
Adult BMI, genetic risk score	87.8 (6.2)[a]	87.8 (6.2)	87.9 (6.2)	88.0 (6.1)
Adult BMI, measured, kg/m^2	26.1 (4.7)[b]	26.3 (4.3)	26.7 (4.8)	27.1 (4.8)
Number of prior endoscopies	3.2 (2.0)	2.2 (1.6)	2.1 (1.5)	2.1 (1.6)
Time since most recent previous period endoscopy, years	3.0 (3.2)	2.6 (3.7)	2.4 (3.3)	2.2 (3.4)
Reason for endoscopy is symptoms, %	26.0	31.2	32.4	30.5
Female, %	73.8	56.8	72.8	57.5

BMI: body mass index, NHS: Nurses' Health Study, NHS2: Nurses' Health Study II, HPFS: Health Professionals Follow-up Study

[a] Values are mean (SD) for continuous variables, percentages for categorical variables.
[b] n=20 participants missing BMI, all non-polyps

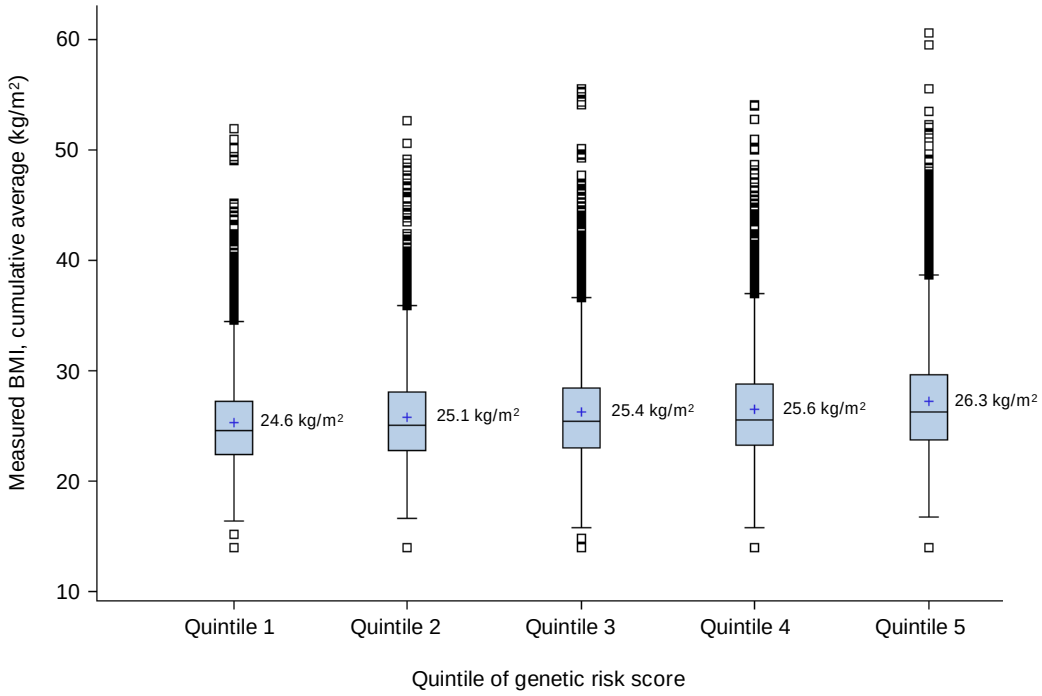


Figure 1.1. Distribution of measured BMI according to quintiles of genetic risk score in three cohorts (NHS, NHS2, HPFS). Median value of measured BMI for each quintile is shown.

Abbreviations: BMI, body mass index; NHS, Nurses' Health Study; NHS2, Nurses' Health Study II; HPFS, Health Professionals Follow-up Study

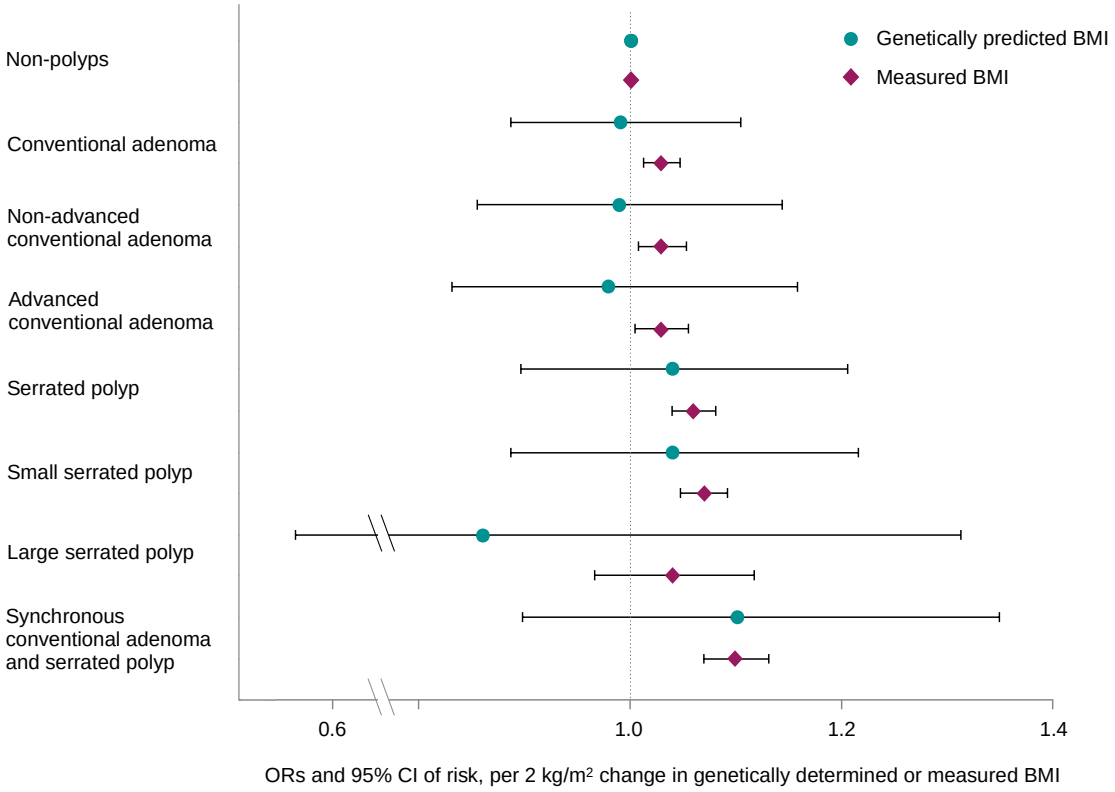


Figure 1.2. Association between genetically predicted [a,b] or measured BMI [c] and risk of colorectal polyp subtypes in three cohorts (NHS, NHS2, HPFS)

BMI: body mass index, NHS: Nurses' Health Study, NHS2: Nurses' Health Study II, HPFS: Health Professionals Follow-up Study

[a] An 18-unit change in genetic risk score is equivalent to a 2 kg/m^2 change in measured BMI, per regression analysis

[b] Genetically predicted BMI multivariable logistic regression model adjusted for age, study cohort (NHS, NHS2, HPFS), time period of endoscopy (in 2-year intervals), reason for endoscopy (screening or symptoms), number of previous endoscopies, time in years since most recent endoscopy, and top three principal components for population structure. All covariates were treated as continuous variables unless otherwise noted.

The association between genetically predicted BMI and polyp risk in the pooled cohorts was the same as measured BMI for serrated polyps (OR per 18 unit [equivalent to 2 kg/m^2] increase=1.04, 95% CI: 0.90, 1.20) and synchronous polyps (OR=1.09, 95% CI: 0.89, 1.34), although confidence intervals were wide and included unity (**Figure 1.2 and Appendix Table 1.1).** Test for heterogeneity showed that measured BMI was associated with a significantly higher risk of serrated polyps compared to conventional adenomas and significantly higher risk of synchronous polyps compared to conventional adenomas or serrated polyps alone. All other tests for heterogeneity were null. Including synchronous polyps in the conventional adenoma and serrated polyp groups did not change the association with genetically predicted or measured BMI (**Appendix Table 1.2**).

Table 1.2 shows the association between genetically predicted BMI or measured BMI and risk of colorectal polyp subtypes by polyp location, in the three pooled cohorts (NHS, NHS2, HPFS). Measured BMI was associated with increased risk of conventional adenomas and serrated polyps at all locations. Of the subtypes and locations, measured BMI was associated with serrated polyps of the proximal colon (OR=1.07, CI: 1.04-1.10), distal colon (OR per 2 kg/m^2 increase=1.07, 95% CI: 1.05-1.10), and rectum (OR=1.08, 95% CI: 1.05-1.11). Genetically predicted BMI was also associated with increased risk of serrated polyps of the distal colon (OR per 18 unit [equivalent to 2 kg/m^2] increase=1.07, CI: 0.90-1.28) and rectum (OR=1.10, 95% CI: 0.91-1.34), although the confidence intervals were wide and included unity. All tests for heterogeneity were null.

Table 1.2. Association between genetically predicted or measured BMI and risk of colorectal polyps, by location, in three cohort studies (NHS, NHS2, HPFS)

	Genetically predicted BMI, per 2 kg/m^2 increase [a,b]	Measured BMI, per 2 kg/m^2 increase [c]
Conventional Adenoma		
Proximal Colon		
n	1921	1921
Mean (GRS or BMI)	87.8	26.6
OR (95% CI)	1.00 (0.88-1.15)	1.05 (1.03-1.07)
Distal Colon		
n	1843	1843
Mean (GRS or BMI)	87.7	26.5
OR (95% CI)	0.96 (0.84-1.11)	1.05 (1.03-1.07)
Rectum		
n	668	668
Mean (GRS or BMI)	87.9	26.4
OR (95% CI)	1.04 (0.83-1.30)	1.04 (1.01-1.08)
Serrated Polyp		
Proximal Colon		
n	762	762
Mean (GRS or BMI)	87.7	26.8
OR (95% CI)	0.95 (0.77-1.18)	1.07 (1.04-1.10)
Distal Colon		
n	1086	1086
Mean (GRS or BMI)	88.0	26.9
OR (95% CI)	1.07 (0.90-1.28)	1.07 (1.05-1.10)
Rectum		
n	855	855
Mean (GRS or BMI)	88.0	27.0
OR (95% CI)	1.10 (0.91-1.34)	1.08 (1.05-1.11)

BMI: body mass index, NHS: Nurses' Health Study, NHS2: Nurses' Health Study II, HPFS: Health Professionals Follow-up Study, GRS: genetic risk score,

[a] An 18-unit change in GRS is equivalent to a 2 kg/m^2 change in measured BMI, per regression analysis

[b] Genetically predicted BMI multivariable logistic regression model adjusted for age, study cohort (NHS, NHS2, HPFS), time period of endoscopy (in 2-year intervals), reason for endoscopy (screening or symptoms), number of previous endoscopies, time in years since most recent endoscopy, and top three principal components for population structure. All covariates were treated as continuous variables unless otherwise noted.

[c] Measured BMI multivariable logistic regression model adjusted for age, study cohort (NHS, NHS2, HPFS), time period of endoscopy (in 2-year intervals), reason for endoscopy (screening or symptoms), number of previous endoscopies, time in years since most recent endoscopy, family history of colorectal cancer (yes or no), height, alcohol intake, regular aspirin use (yes or no), and physical activity. All covariates were treated as continuous variables unless otherwise noted.

No differences were found between men and women, with the exception of genetically predicted BMI and risk of synchronous polyps (P value for interaction = 0.04). Both genetically predicted BMI and measured BMI were associated with increased risk of synchronous polyps in women (genetically predicted BMI, OR per 18 unit [equivalent to 2 kg/m^2] increase=1.37, 95% CI: 1.05-1.79; measured BMI, OR=1.11, 95% CI: 1.07-1.14). Measured BMI, but not genetically predicted BMI, was associated with increased risk of synchronous polyps in men (genetically predicted BMI, OR=0.88, 95% CI: 0.65-1.20; measured BMI, OR=1.09, 95% CI: 1.03-1.16).

Tests for association between individual SNPs and risk of polyp subtypes at the adjusted α level revealed two SNPs that were inversely associated with synchronous polyp: rs11727676 (OR per 1-allele=0.72, 95% CI: 0.61-0.84) and rs10132280 (OR=0.83, 95% CI: 0.74-0.92). No individual SNPs were significantly associated with risk of conventional adenomas or serrated polyps. There were no significant associations between individual SNPs and polyps in men or women separately.

Discussion

In three cohorts of men and women, genetically predicted BMI was correlated with measured adulthood BMI. Measured BMI showed a positive association with risk of conventional adenoma, serrated polyp, and synchronous polyps. These associations tended to be stronger for serrated polyp and synchronous polyps compared to conventional adenoma. Genetically predicted BMI showed similar positive associations with risk of serrated and synchronous polyps, however, confidence intervals for genetically predicted BMI were wide and included unity. When stratified by sex, genetically predicted BMI was associated with increased risk of synchronous polyps in women only.

This is the first study to our knowledge to assess risk of CRC polyp subtypes in relation to BMI genetic risk score, thus it is difficult to put these findings in the context of the literature. Several studies analyzed genetic BMI in relation to CRC diagnosis: all three found that genetic

BMI was associated with increased risk of CRC overall[11,12] and in women only, when stratified by sex.[7] Our finding that genetically predicted BMI was not significantly associated with CRC precursor lesions could be due to limited power in the present study, given that these other studies investigating CRC risk included 9,254 to 51,537 cases. It is not possible to draw conclusions from the present study regarding whether genetically predicted BMI is more strongly associated with CRC than CRC precursor lesions.

Our finding that measured BMI was associated most strongly with risk of serrated polyp and synchronous polyps is consistent with one other large study which analyzed measured BMI in relation to polyp subtypes in the same three cohorts (NHS, NHS2, HPFS), and found that BMI was most strongly associated with synchronous polyps compared to serrated polyp or conventional adenoma [4]. Prior studies have demonstrated an association between BMI and conventional adenoma and serrated polyp independently, but no other studies to our knowledge have compared the association between BMI and various polyp subtypes..[13,14] The association between increased BMI and serrated polyp risk is reasonable given that serrated polyps tend to be associated with other environmental factors related to inflammation, including smoking and alcohol consumption.[4] Inflammatory cytokines, which are increased in a state of excess adiposity, are hypothesized to contribute to the development of microsatellite-instable colorectal cancer.[15,16] These microsatellite-instable cancers are more commonly developed in serrated polyps and more commonly found in women; thus the inflammatory state of obesity may explain the association of increased adiposity with serrated as well as synchronous polyps, especially in women.[5]

The significance of the relationship between BMI and synchronous polyps is unclear. Synchronous polyps were not associated with greater risk of progression to CRC, compared to conventional adenoma only, in a previous study, suggesting that synchronous polyp diagnosis is not necessarily predictive of a more advanced lesion.[17] In contrast to observational studies, which found increased CRC risk in association with BMI in men only, but consistent with

Mendelian randomization studies, which found increased CRC risk in association with BMI in women only, we observed an association between genetically predicted BMI and synchronous polyps in women only.[3,12] Genetically predicted BMI has been associated with adiposity across the lifespan, particularly in early life, and therefore may capture the influence of early-life adiposity.[6,18] Interestingly, early-life body fatness has been associated with higher CRC risk in women but not in men.[19]

Our analysis of individual obesity-related SNPs showed no statistically significant associations of individual SNPs and conventional adenoma or serrated polyp. Another study evaluated the association between five individual *FTO* gene polymorphisms and colorectal neoplasia, and found no significant associations in Caucasians, consistent with our findings for individual BMI alleles in relation to serrated polyp or conventional adenoma.[20] Interestingly, we identified two SNPs which were inversely related to synchronous polyp risk: rs11727676 and rs10132280. SNP rs11727676 is associated with the *HHIP* (Hedgehog interacting protein) gene, which is hypothesized to be associated with increased subcutaneous fat with more favorable metabolic markers.[9] SNP rs10132280 is related to the *STXBP6* (syntaxin-binding protein 6) gene, which has previously been shown to be negatively correlated with lung adenocarcinoma.[21] No other studies to our knowledge have investigated individual obesity-related SNPs in relation to CRC precursor lesions or CRC risk.

The strengths of this study include the large cohort with prospective assessment of lifestyle factors, polyp documentation, and detailed histopathological information, which allowed for stratification of CRC precursor lesions by histology and location. We used a genetic risk score derived from 97 SNPs that have been associated with adult BMI to date. Despite the large sample size, the number of polyp cases was still relatively small for a Mendelian randomization study, and the non-significant findings for genetically predicted BMI and polyp subtypes may have been due to limited power.

In conclusion, in three large prospective cohorts of men and women, we found similar associations between measured or genetic BMI and CRC precursor lesions. Associations tended to be stronger for serrated and synchronous polyps, compared to conventional adenoma, in congruence with previous studies. Larger cohorts and meta-analyses would be helpful in elucidating the relationship between genetic risk of BMI and distinct CRC precursor lesions.

Ethics approval

The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The study protocol was approved by the institutional review boards of Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Metabolomic signatures of inflammation and metabolic dysregulation in relation to colorectal cancer risk

Alaina M Bever, Dong Hang, Dong Hoon Lee, Fred K Tabung,

Tomotaka Ugai, Shuji Ogino, Jeffrey A Meyerhardt, Andrew T Chan,

A Heather Eliassen, Liming Liang, Meir J Stampfer, and Mingyang Song

ABSTRACT

Background: Inflammation and metabolic dysregulation are associated with increased risk of colorectal cancer (CRC); the underlying mechanisms are not fully understood. We characterized metabolomic signatures of inflammation and metabolic dysregulation and evaluated the association of the signatures and individual metabolites with CRC risk.

Methods: Among 684 incident CRC cases and 684 age-matched controls in the Nurses' Health Study (n=818 women) and Health Professionals Follow-up Study (n=550 men), we applied reduced rank and elastic net regression to 277 metabolites for markers of inflammation (CRP, IL6, TNFRSF1B, and GDF15) or metabolic dysregulation (body mass index, waist circumference, C-peptide, and adiponectin) to derive metabolomic signatures. We evaluated the association of the signatures and individual metabolites with CRC using multivariable conditional logistic regression. All statistical tests were 2-sided.

Results: We derived a signature of 100 metabolites that explained 24% of variation in markers of inflammation and a signature of 73 metabolites that explained 27% of variation in markers of metabolic dysregulation. Among men, both signatures were associated with CRC (odds ratio per 1-standard deviation increase, inflammation = 1.34, 95% confidence interval 1.07 to 1.68; metabolic dysregulation = 1.25, 1.00 to 1.55); neither signature was associated with CRC in women. Eleven metabolites were individually associated with CRC and biomarkers of inflammation or metabolic dysregulation among either men or women.

Conclusion: We derived metabolomic signatures and identified individual metabolites associated with inflammation, metabolic dysregulation, and CRC, highlighting several

metabolites as promising candidates involved in the inflammatory and metabolic dysregulation pathways for CRC incidence.

INTRODUCTION

Chronic inflammation and obesity-related metabolic dysregulation are key mechanisms in colorectal cancer (CRC). Plasma biomarkers including higher levels of CRP, IL6, TNFSRF1B, GDF15, C-peptide (reflecting insulin production), and lower adiponectin are associated with higher risk of CRC, albeit with some inconsistency.[1–6] Waist circumference, a proxy for visceral adiposity, is associated with increased CRC risk in men and women.[7] Associations of individual markers of inflammation or metabolic dysregulation tend to be stronger for colon cancer than rectal cancer and stronger and more consistent for men than women. Empirical dietary inflammatory and insulinemic pattern scores derived from these biomarkers have also been associated with increased CRC risk with stronger associations among men.[8,9] These insights motivate questions regarding the mechanisms by which inflammation and metabolic dysregulation contribute to CRC development.

Expanding our understanding of the complex interactions among inflammation, metabolism, and CRC warrants high throughput metabolomic analysis. Metabolites may influence CRC incidence through signaling pathways or by modulating the nutrient milieu that fuels tumor proliferation. Several prior studies have explored the prospective relation between plasma metabolites and CRC risk, including one evaluating the association of metabolomic signatures of anthropometrics with CRC.[10–14] However, few metabolites have been consistently associated with CRC risk. Moreover, no studies have yet investigated inflammatory and metabolic mechanism-specific metabolites in relation to CRC risk. We therefore derived metabolomic signatures of inflammation and metabolic dysregulation in two prospective cohort studies, the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS); evaluated the signatures and individual metabolites in relation to incident CRC; and investigated

possible differences by biologic sex, tumor location, and tumor molecular profile. Identifying these key metabolomic pathways could provide insights for devising targeted interventions for CRC interception and offer biomarkers that can be applied to other conditions related to inflammation and metabolic dysregulation.

MATERIALS AND METHODS

Study population and blood sample collection

The NHS enrolled 121,700 female nurses, aged 30–55 years, in 1976. The HPFS enrolled 51,529 male health professionals, aged 40–75 years, in 1986. Participants returned self-administered biennial questionnaires regarding medical conditions, family history, and lifestyle, with food frequency questionnaires every four years. A subset provided blood specimens: 32,826 NHS participants 1989–1990, and 18,225 HPFS participants 1993–1995. Specimens were returned on ice packs by overnight courier and at least 95% arrived within 26 hours of blood collection.[15,16] Individuals who provided blood samples had similar characteristics to those who did not.[17]

The study protocol was approved by the institutional review boards of Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating cancer registries.

Identification of CRC cases and matched controls

CRC diagnoses were self-reported on biennial questionnaires and confirmed, with permission, via review of medical records and pathologic reports. A study physician, blinded to exposure information, reviewed records to confirm CRC diagnosis and extracted information on anatomic location, stage, and histology. None of the cases had known CRC at the time of blood draw and were diagnosed between one month and 23 years after the blood collection (cases diagnosed within five years of blood collection were excluded in a sensitivity analysis). Risk-set sampling

was used to randomly select one control for each CRC case—matched on age at blood collection, date of blood collection, and self-reported fasting at blood collection—from participants with archived blood samples who were alive and free of CRC at the time of case diagnosis. Details of participant inclusion are shown in **Appendix Figure 2.1.**

Biomarker assays

Metabolites were assayed in 2015 from archived plasma samples among confirmed CRC cases and matched controls; metabolomics data were generated at the Broad Institute using liquid chromatography tandem mass spectrometry;[18] details of biomarker processing are described in **Appendix Methods 2.1**. Following exclusions (as described in the supplement), 277 metabolites from over 30 different subclasses were included in the main analyses (a comprehensive list of the metabolites and their associated subclasses is provided in **Appendix Table 2.2**). Plasma biomarkers of inflammation (CRP, IL6, TNFRSF1B, GDF15) and metabolic dysregulation (C-peptide and adiponectin) were assayed between 2003–2010 as part of nested CRC case-control studies that included most of the same cases and controls selected for metabolomic assay.[4,19,20] We used plasma C-peptide to assess hyperinsulinemia because C-peptide has a slower clearance and is a better measure of insulin secretion compared to insulin.[21,22] Other common markers of metabolic dysregulation, such as triglycerides, HDL-C, and glucose, were not assayed.

Non-biomarker exposure and covariates

Body mass index (BMI) was calculated using self-reported height on the baseline questionnaire and cumulative average weight from biennial questionnaires returned prior to blood collection (on average, 14 years for NHS and eight for HPFS). Waist circumference was assessed at a single time prior to blood collection in 1986 (NHS) or 1988 (HPFS). BMI and waist

circumference were included in the metabolic dysregulation signature along with C-peptide and adiponectin.

The following potential confounders were selected based on established association with CRC: age, height, physical activity (metabolic equivalent of task score (METS)-hours/week), alcohol consumption, pack-years of cigarette smoking, history of CRC in a parent or sibling, history of prior endoscopy for screening, regular aspirin use, multivitamin use, alternative healthy eating index (AHEI) score, and in women, menopausal status and postmenopausal hormone use. Details of confounder assessment and imputation are provided in **Appendix Methods 2.1**. Race was self-reported in 1986 (HPFS) or 1992 (NHS); we reported the distribution of self-reported race among cases and controls but did not include race as a confounder due to the limited diversity of the study population (> 95% White).

Statistical analysis

Participants were split into ten sets with one set as a testing (hold-out) set and the other nine sets as a training set; we derived the metabolomic signatures using a two-step process that was repeated in each of the ten iterations of training data (**Figure 2.1)**. First, we used reduced rank regression (RRR) with the four markers of inflammation (CRP, IL6, TNFRSF1B, GDF15) or metabolic dysregulation (BMI, waist circumference, C-peptide, adiponectin) as response variables and all 277 available metabolites as predictor variables.[23,24] We extracted the first factor, a linear combination of the response variables for which maximal variation is explained by the metabolites, and used it as a summary score for inflammation or metabolic dysregulation. This is analogous to extracting the first component derived via principal components analysis of the markers of inflammation or metabolic dysregulation, with the added benefit of RRR specifically maximizing the variation in the response variables explained by the predictor

variables (metabolites). Next, we used each summary score (one score for inflammation and one for metabolic dysregulation) derived from RRR as the dependent variable in elastic net regression, again with all 277 metabolites as predictor variables. Elastic net regression is a regularization method that combines the penalties of ridge regression and least absolute shrinkage and selection operator (LASSO) to perform both variable selection and shrinkage to reduce overfitting.[25–27] Gaussian elastic net regression models were constructed using the "glmnet" package in R, and adjusted for age at blood collection, date of blood collection, fasting status, and sex. We performed 10-fold cross validation to select a lambda parameter that minimized mean squared error and selected 0.5 as the alpha parameter a priori.

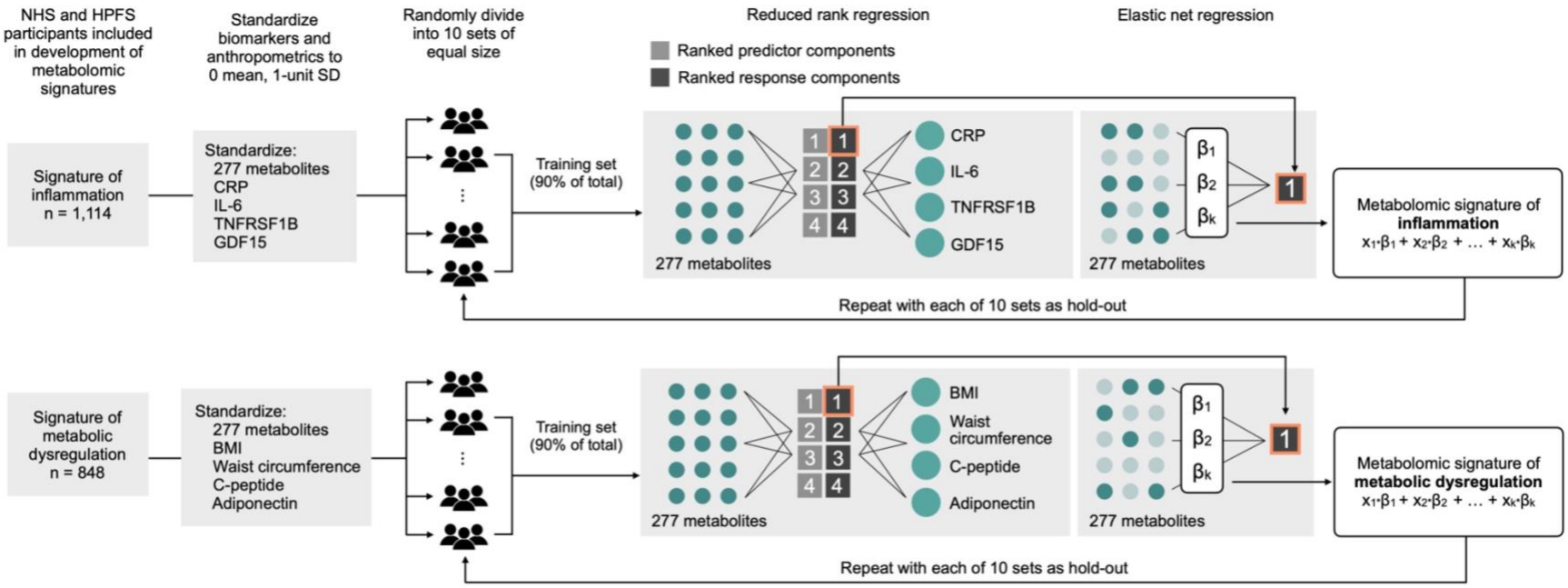


Figure 2.1 Statistical methods used to derive the metabolomic signatures of inflammation and metabolic dysregulation.

Within each of 10 training sets derived from the same population of women (NHS) and men (HPFS), we applied reduced rank regression and elastic net regression to 277 unique metabolites and 4 intermediate markers of inflammation (CRP, IL6, TNFRSF1B, GDF15) or metabolic dysregulation (BMI, waist circumference, C-peptide, adiponectin

Abbreviations: BMI, body mass index; CRP, C-reactive protein; GDF15, growth differentiation factor 15; HPFS, Health Professionals Follow-up Study; IL6, interleukin-6; NHS, Nurses' Health Study; TNFRSF1B, tumor necrosis factor receptor super family 1B.

To test the robustness of the signatures, we compared our a priori selected model, the combination of RRR and ENR, to four alternative models: 1) combined principal components analysis (PCA) and ENR, 2) sparse RRR, 3) RRR alone, and 4) combined RRR and stepwise linear regression. These four alternative methods cover different combinations of processes for summarizing the markers of inflammation or metabolic dysregulation and deriving parsimonious models of metabolites to predict those markers, with details provided in the **Appendix Methods 2.1**. We compared models in terms of correlation with the RRR/ENR model, proportion of variance explained (R-squared) in inflammation and metabolic dysregulation markers in each of the 10 hold-out test sets, and overlapping metabolites retained in all 10 iterations of each model. We also modeled the association of 1-SD increase in each individual marker of inflammation (CRP, IL-6, TNFRSF1B, GDF15) and metabolic dysregulation (BMI, waist circumference, C-peptide, adiponectin) with CRC.

We categorized individuals into quartiles based on the distribution of metabolomic signatures among controls (to reflect the distribution among the source population) and calculated the odds ratio (OR) of CRC per 1-SD increase in each signature using multivariable conditional logistic regression. P-trend was derived from the p-value in the 1-SD regression model. We tested for effect measure modification of the inflammation signature by BMI group (lean, BMI < 25 kg/m^2 vs. overweight and obese, BMI ≥ 25 kg/m^2) and by the signature of metabolic dysregulation. We also conducted multivariable linear or conditional logistic regression to assess the association of individual metabolites with individual markers of inflammation, metabolic dysregulation, and CRC, with statistical significance reported both before and after controlling the false discovery rate (FDR) at level $\alpha = 0.05$ using the Benjamini–Hochberg procedure.[28] We identified key metabolites as those retained in all ten cross-validations of either metabolomic signature and individually associated with both CRC and markers of inflammation or metabolic dysregulation. Given the established sex difference in

CRC risk related to metabolic and inflammatory factors, we assessed all CRC associations among women and men separately.[6,29–31]

In the secondary analyses, we assessed the association of the metabolomic signatures according to anatomic location of CRC (proximal colon, distal colon, or rectal), common molecular subtypes (conventional, serrated, alternate, or other), and *KRAS* mutation status (mutation or wildtype).[32–34] Details of the molecular subtypes are described in the **Appendix Methods 2.1**. We used inverse probability weighting to reduce selection bias due to tissue availability for the molecular subtype analyses. To assess reverse causality due to preclinical disease, we evaluated the metabolomic association with CRC after excluding cases diagnosed within five years following blood collection. We derived sex specific signatures among women and men separately and compared the sex-specific signatures to those used in the main analysis (derived among men and women combined). All analyses were conducted using SAS (version 9.4, SAS Institute Inc, Cary, NC) or R (version 4.2.0, available at https://www.r-project.org/), and all tests for statistical significance were 2-sided with $\alpha = 0.05$.

RESULTS

Participants were predominantly White with mean age of 59 (NHS) or 66 (HPFS) years at blood collection (**Table 2.1**). Compared to controls, cases had more family history of CRC, fewer prior screening endoscopies, and less aspirin use. Cases also had higher BMI, waist circumference, and plasma C-peptide, CRP, IL6, TNFRSF1B, and GDF15 and lower plasma adiponectin than controls, with the differences more pronounced in men than women.

Through RRR and elastic net regression, 100 metabolites were consistently selected for the signature of inflammation and 73 for metabolic dysregulation across training sets (**Appendix Table 2.1**). As shown in **Figure 2.2a–c,** 13 common metabolites were positively loaded and 13 negatively loaded in both signatures. The inflammation signature was positively loaded for fatty acids, glycinated bile acids, and purines, and negatively loaded for bilirubins, steroids, and

pyrimidines; the metabolic dysregulation signature was positively loaded for purines and ceramides and negatively loaded for steroids (**Figure 2.3a–b**).

Table 2.1. Baseline characteristics of study participants in the NHS and HPFS [a,b]

	Women (NHS)		Men (HPFS)	
	Controls	Cases	Controls	Cases
Baseline characteristics	**N = 409**	**N = 409**	**N = 275**	**N = 275**
Age at blood collection,[c] years	58.7 (6.8)[d]	58.7 (6.8)	65.7 (8.0)	65.7 (8.0)
Date of blood collection	February 1990 (4 mos)	February 1990 (4 mos)	February 1994 (5 mos)	February 1994 (5 mos)
Fasting at blood collection, %	68.3	68.0	57.7	56.3
Race,[e] %				
Asian	0	0.5	0	0
Black	0.2	0.7	0.2	0.2
Indigenous American	0.2	0	—	—
White	99.5	98.8	97.2	93.5
Other	0	0	2.6	6.3
Postmenopausal, %	87.8	86.4	—	—
Current use of hormone therapy,[f] %	46.4	43.5	—	—
Colorectal cancer in a sibling or parent, %	14.0	14.9	17.3	21.0
History of previous endoscopy, %	10.4	8.9	50.0	44.2
Regular aspirin use (2 or more tablets/week), %	45.3	36.3	48.4	45.1
Smoker, <40 pack-years, %	46.8	43.9	42.3	41.2
Smoker, 40+ pack-years, %	10.2	11.8	7.6	10.4
Alcohol, daily intake, g/day	6.8 (9.2)	7.1 (10.3)	11.7 (12.2)	12.1 (13)
Alternative Healthy Eating Index (AHEI) score, 0–100-pt scale	47.1 (8.7)	46.1 (8.5)	48.6 (9.5)	48.6 (9.8)
Physical activity, METS-hr/week	15.9 (17.4)	15.0 (14.6)	29.4 (19.4)	28.2 (18.9)
Body mass index (BMI), kg/m^2	24.5 (4)	24.9 (4)	25.2 (2.4)	26.1 (2.5)

Table 2.1. (Continued) Baseline characteristics of study participants in the NHS and HPFS [a,b]

Waist circumference, inches	31.1 (4.1)	31.8 (4.2)	37.3 (3.1)	38.4 (3.1)
C-peptide, ng/mL, median (IQR)	1.80 (1.27–2.68)	1.91 (1.35–2.74)	2.20 (1.51–3.36)	2.40 (1.74–3.63)
Adiponectin, µg/mL median (IQR)	7.92 (5.19–10.17)	7.96 (5.51–10.87)	6.62 (4.92–7.78)	6.52 (4.72–7.75)
CRP, mg/L, median (IQR)	1.62 (0.78–3.26)	1.52 (0.66–3.18)	1.02 (0.56–2.06)	1.39 (0.68–2.71)
IL6, pg/mL, median (IQR)	1.04 (0.71–1.62)	1.04 (0.74–1.63)	1.36 (0.93–2.26)	1.58 (1.00–2.68)
TNFRSF1B, ng/mL, median (IQR)	2.64 (2.22–3.07)	2.71 (2.30–3.16)	2.69 (2.3–3.29)	2.79 (2.35–3.2)
GDF15, pg/mL, median (IQR)	749 (570–966)	755 (612–919)	834 (627–1,060)	930 (669–1,150)

Abbreviations: BMI, body mass index; CRC, colorectal cancer; CRP, C-reactive protein; GDF15, growth differentiation factor 15; HPFS, Health Professionals Follow-up Study; IL6, interleukin 6; NHS, Nurses' Health Study; TNFRSF1B, tumor necrosis factor receptor superfamily member 1B

[a] Lifestyle factor data were collected every two years and dietary data were collected every four years starting in 1986 (HPFS) or 1976 (NHS). For the following variables, the cumulative average of all data preceding blood collection was used: regular aspirin use, daily alcohol intake, alternative healthy eating index score, physical activity, and body mass index. Waist circumference was assessed once in 1986 (NHS) or 1987 (HPFS).

[b] Missing values that could not be imputed with cumulative individual data were imputed with median imputation (or most frequent category, for categorical variables); % missing was less than 5% for all confounders. Missing values for variables used in the metabolomic signature derivation were not imputed (complete case analysis): BMI (NHS, n=1, <1%; HPFS, n=0, 0%), waist circumference (NHS, n=249, 30%; HPFS, n=107, 13%), C-peptide (NHS, n=169, 21%; HPFS, n=46, 6%), adiponectin (NHS, n=171, 21%; HPFS, n=42, 5%), CRP (NHS, n=168, 21%; HPFS, n=47, 6%), IL6 (NHS, n=171, 21%; HPFS, n=46, 6%), TNFRSF1B (NHS, n=173, 21%; HPFS, n=41, 5%); MIC-1 (NHS, n=145, 18%; HPFS, n=44, 8%).

[c] Value is not age adjusted

[d] Values are means (SD) and are standardized to the age distribution of the study population, unless noted otherwise

[e] Race was self-reported by study participants as follows: NHS, Asian, Black, Indigenous American, White (includes Hispanic and non-Hispanic ethnicity), or other (not specified); HPFS, Asian, Black, White (includes Hispanic and non-Hispanic ethnicity), or other (not specified).

[f] Percent is among postmenopausal women (n = 696)

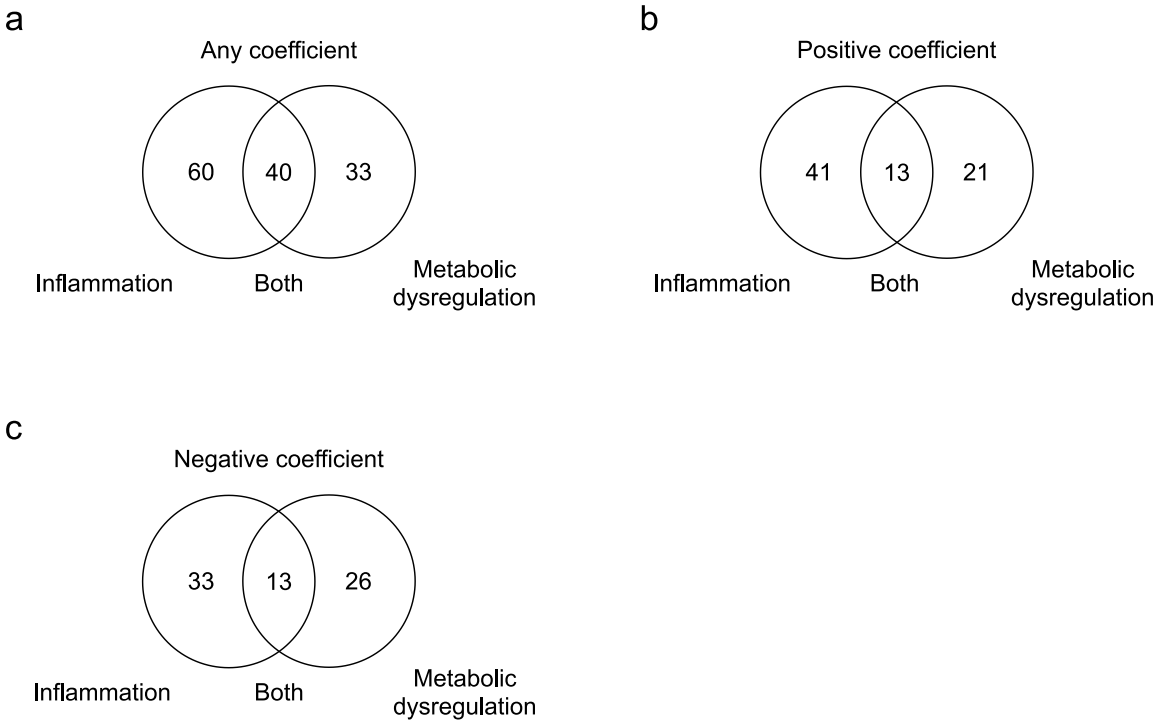


Figure 2.2. Overlap in metabolites selected for the derived signatures of inflammation and metabolic dysregulation. a–c, 100 metabolites were retained in all 10 training sets for the signature of inflammation and 73 metabolites were retained in all 10 training sets for the signature of metabolic dysregulation; the common and unique metabolites retained in all 10 iterations of both signatures **a)** overall, **b**) with positive coefficient, and **c)** with negative coefficient are shown.

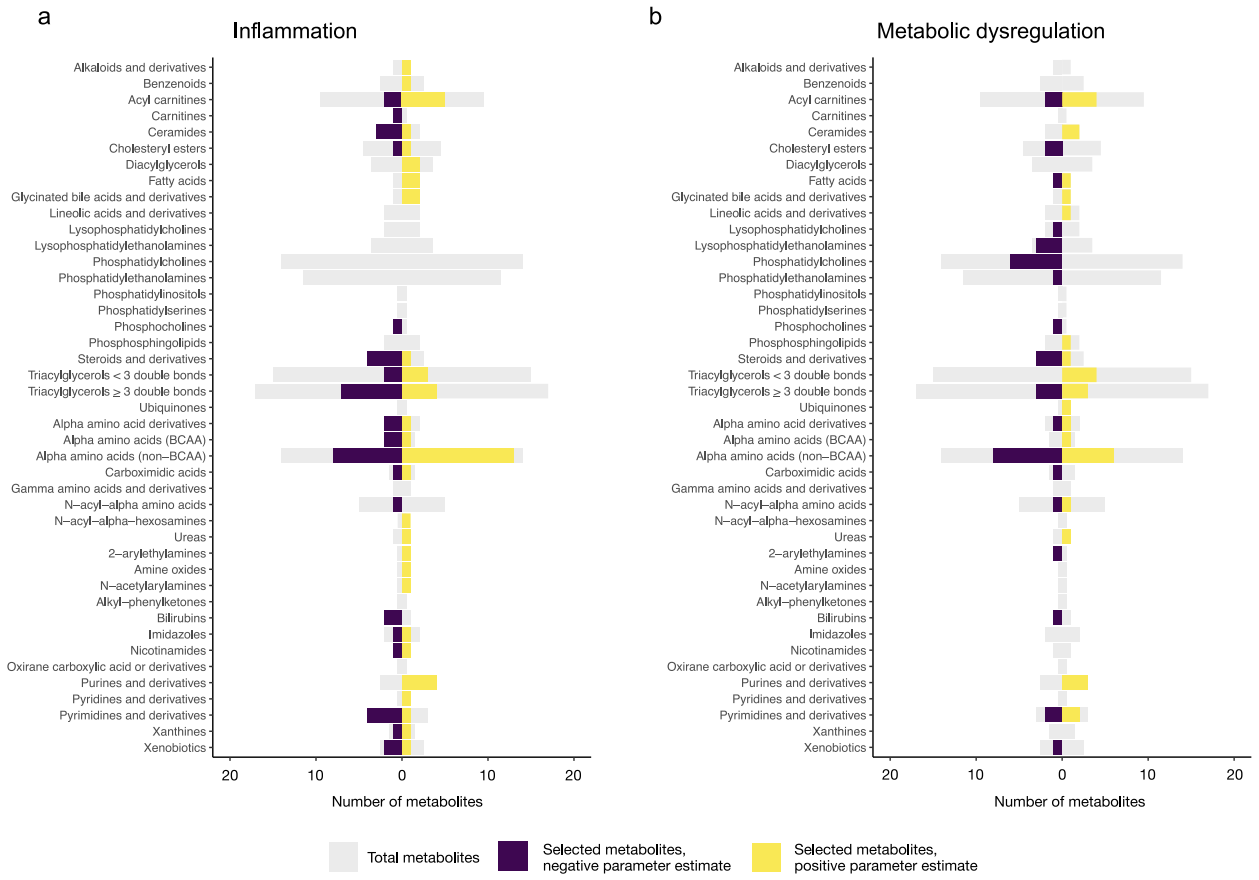


Figure 2.3 Metabolite subclass enrichment of the derived signatures of a) inflammation and b) metabolic dysregulation. The number of metabolites with a positive parameter estimate (yellow bars, directly related to inflammation or metabolic dysregulation) vs. a negative parameter estimate (purple bars, inversely related to inflammation or metabolic dysregulation) retained in all 10 iterations is overlayed on the total number of available metabolites (gray, centered on 0) for a given metabolite subclass.

The inflammation and metabolic dysregulation signatures explained, on average, 24% of variation in the markers of inflammation and 27% for metabolic dysregulation; the explained percent variation was comparable among women and men (**Appendix Figure 2.3**). Among men, the inflammation signature was associated with increased odds of CRC (OR, 95% CI comparing extreme quartiles = 2.19, 1.23 to 3.92; OR, 95% CI per 1-SD increase = 1.34, 1.07 to 1.68, **Table 2.2**). Similarly, the metabolic dysregulation signature was associated with increased odds of CRC among men (OR, 95% CI comparing extreme quartiles = 1.31, 0.76 to 2.25; OR, 95% CI per 1-SD increase = 1.25, 1.00 to 1.55). Neither signature was statistically significantly associated with CRC among women. The individual markers of inflammation and metabolic dysregulation showed similar associations with CRC as the metabolomic signatures (**Appendix Figure 2.4**). No differential associations were found by tumor subsite, molecular subtype, or *KRAS* mutation status (**Appendix Table 2.3**). There was no statistically significant effect modification of inflammation signature by the metabolic dysregulation signature or by BMI group (data not shown).

Multiple logistic regression of all 277 individual metabolites with CRC revealed 55 metabolites significantly associated with CRC in either men or women (**Figure 2.4**). Only one metabolite, glycocholate, was associated with increased odds of CRC in both sexes. Among men, all but one of the 19 metabolites associated with increased odds of CRC were associated with higher biomarker levels of inflammation or metabolic dysregulation; among women, these same metabolites were similarly associated with inflammation and metabolic dysregulation but not associated with CRC. Most metabolites associated with CRC among women were associated with at least one marker of inflammation or metabolic dysregulation in a direction paradoxical to what was expected, e.g., associated with decreased odds of CRC and increased BMI, waist circumference, and C-peptide and decreased adiponectin.

Table 2.2. Multivariable adjusted [a] odds ratio of CRC according to quartiles and 1-SD increase of metabolomic signatures of inflammation and metabolic dysregulation, in women (NHS) and men (HPFS).

Metabolomic signature of inflammation
Combination of metabolites predictive of CRP, IL6, TNFRSF1, and GDF15

	Q1 [b]	Q2	Q3	Q4	1-SD increase	P_{trend} [c]
Women (NHS)						
Cases	90	115	74	130	409	0.80
OR (95% CI)	1.00 (Ref)	1.23 (0.80 to 1.91)	0.66 (0.42 to 1.06)	1.33 (0.85 to 2.08)	1.02 (0.8 to 1.19)	
Men (HPFS)						
Cases	59	62	54	102	225	0.01
OR (95% CI)	1.00 (Ref)	1.14 (0.67 to 1.94)	1.05 (0.59 to 1.88)	2.19 (1.23 to 3.92)	1.34 (1.07 to 1.68)	

Metabolomic signature of metabolic dysregulation
Combination of metabolites BMI, waist circumference, C-peptide, and adiponectin

	Q1	Q2	Q3	Q4	1-SD increase	P_{trend}
Women (NHS)						
Cases	131	79	74	125	409	0.62
OR (95% CI)	1.00 (Ref)	0.71 (0.47 to 1.07)	0.57 (0.37 to 0.88)	0.93 (0.61 to 1.42)	0.96 (0.83 to 1.12)	
Men (HPFS)						
Cases	59	63	72	83	225	0.05
OR (95% CI)	1.00 (Ref)	0.96 (0.58 to 1.58)	1.24 (0.71 to 2.16)	1.31 (0.76 to 2.25)	1.25 (1.00 to 1.55)	

Table 2.2 (Continued) Multivariable adjusted [a] odds ratio of CRC according to quartiles and 1-SD increase of metabolomic signatures of inflammation and metabolic dysregulation, in women (NHS) and men (HPFS).

Abbreviations: BMI, body mass index; CRC, colorectal cancer; CRP, C-reactive protein; GDF15, growth differentiation factor 15; HPFS, Health Professionals Follow-up Study; IL6, interleukin-6; METS, metabolic equivalent of task score; NHS, Nurses' Health Study; SD, standard deviation; TNFRSF1B, tumor necrosis factor receptor superfamily member 1B

[a] Adjusted for age at blood collection, date of blood collection, fasting status (yes or no), height, physical activity (METS-hours/week), alcohol intake (0, 0.1-5 g/day, 5-10 g/day, 10-15 g/day, 15+ g/day), pack-years of smoking, history of CRC in parent or sibling (yes or no), history of prior endoscopy (yes or no), regular aspirin use (>2 tablets/week, yes or no), alternative healthy eating index (AHEI) score. Models in women additionally adjusted for menopausal status (pre- or postmenopausal) and postmenopausal hormone use (never, past, or current).

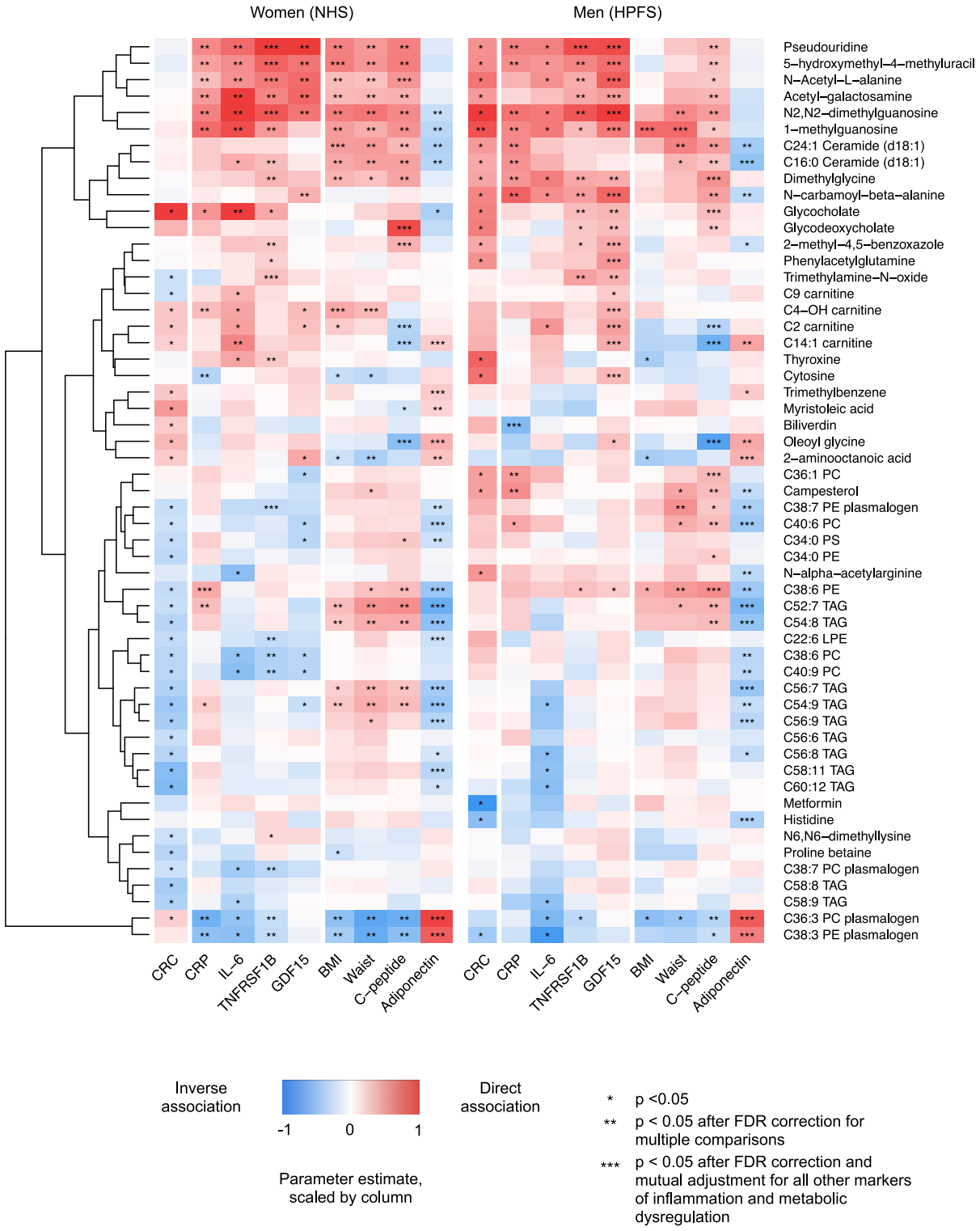
[b] Quartile cut points were determined based on IQR among controls

[c] P_{trend} is derived from 1-SD multivariable adjusted linear regression model

Figure 2.4. Individual metabolites associated with incident CRC in either men or women, shown beside association with markers of inflammation and markers of metabolic dysregulation using hierarchical clustering among men (HPFS) and women (NHS). Colors show the magnitude and direction of association for individual metabolites from multivariable adjusted logistic regression (CRC) or linear regression (markers of inflammation or metabolic dysregulation), with darker blue denoting a stronger inverse association and darker red a stronger direct association. All models were adjusted for age at blood collection, date of blood collection, fasting status (yes or no), height, physical activity (METS-hours/week), alcohol intake (0, 0.1-5 g/day, 5-10 g/day, 10-15 g/day, 15+ g/day), pack-years of smoking, history of CRC in parent or sibling (yes or no), history of prior endoscopy (yes or no), regular aspirin use (>2 tablets/week, yes or no), alternative healthy eating index (AHEI) score. Models in women additionally adjusted for menopausal status (pre- or postmenopausal) and postmenopausal hormone use (never, past, or current).

***** indicates significance at $p<0.05$ before FDR correction for multiple testing, ****** indicates significance at $p<0.05$ after FDR correction, ******* indicates significance at $p<0.05$ after FDR correction and mutual adjustment for the other 7 markers of inflammation and metabolic dysregulation (BMI was not mutually adjusted for waist circumference and waist circumference not mutually adjusted for BMI, as this materially changes the interpretation of these variables).

Abbreviations: BMI, body mass index; CRP, C-reactive protein; Health Professionals Follow-up Study; IL6, interleukin-6; GDF15, growth differentiation factor 15; METS, metabolic equivalent of task score; NHS, Nurses' Health Study; TNFRSF1B, tumor necrosis factor receptor super family 1B


Women (NHS)
Men (HPFS)
Pseudouridine
5–hydroxymethyl–4–methyluracil
N–Acetyl–L–alanine
Acetyl–galactosamine
N2,N2–dimethylguanosine
1–methylguanosine
C24:1 Ceramide (d18:1)
C16:0 Ceramide (d18:1)
Dimethylglycine
N–carbamoyl–beta–alanine
Glycocholate
Glycodeoxycholate
2–methyl–4,5–benzoxazole
Phenylacetylglutamine
Trimethylamine–N–oxide
C9 carnitine
C4–OH carnitine
C2 carnitine
C14:1 carnitine
Thyroxine
Cytosine
Trimethylbenzene
Myristoleic acid
Biliverdin
Oleoyl glycine
2–aminooctanoic acid
C36:1 PC
Campesterol
C38:7 PE plasmalogen
C40:6 PC
C34:0 PS
C34:0 PE
N–alpha–acetylarginine
C38:6 PE
C52:7 TAG
C54:8 TAG
C22:6 LPE
C38:6 PC
C40:9 PC
C56:7 TAG
C54:9 TAG
C56:9 TAG
C56:6 TAG
C56:8 TAG
C58:11 TAG
C60:12 TAG
Metformin
Histidine
N6,N6–dimethyllysine
Proline betaine
C38:7 PC plasmalogen
C58:8 TAG
C58:9 TAG
C36:3 PC plasmalogen
C38:3 PE plasmalogen
CRC
CRP
IL–6
TNFRSF1B
GDF15
BMI
Waist
C–peptide
Adiponectin
Inverse association
Direct association
-1
0
1
Parameter estimate, scaled by column
* p <0.05
** p < 0.05 after FDR correction for multiple comparisons
*** p < 0.05 after FDR correction and mutual adjustment for all other markers of inflammation and metabolic dysregulation

We identified 11 most promising metabolites as potential mediators of the inflammatory and metabolic dysregulation pathways of CRC development: C16:0 ceramide (d18:1), C24:1 ceramide (d18:1), bile acids (glycocholate and glycodeoxycholate), N-acetyl-L-alanine, acetyl-galactosamine, methylated purine derivatives (1-methylguanosine and N2,N2-dimethylguanosine), pyrimidine derivatives (pseudouridine and 5-hydroxymethyl-4-methyluracil), and N-carbamoyl-beta-alanine (**Table 2.3)**.

The CRC associations of individual metabolites and signatures remained essentially unchanged after excluding cases diagnosed within five years following blood collection (data not shown). There was significant overlap in the metabolites selected for each signature among the five different methods and the signatures derived using alternative methods showed almost identical association with CRC as the signatures in the main analysis (**Appendix Figures 2.5** and **2.6**). The association of the signatures with CRC did not change with the use of sex-specific signatures (data not shown).

Table 2.3. Summary of select metabolites and metabolite subclasses identified as strongest potential mediators of the association of metabolic dysregulation and inflammation with CRC (retained in at least one of the metabolomic signatures, and individually associated with both CRC and markers of inflammation or metabolic dysregulation).

Subclass	Metabolite	Selected in signature	Association with CRC		Association with inflammation markers		Association with metabolic markers	
			Women (NHS)	Men (HPFS)	Women (NHS)	Men (HPFS)	Women (NHS)	Men (HPFS)
Ceramides	C16:0 Ceramide (d18:1)	Both		+	IL6 + TNFRSF1B ++	CRP ++	BMI ++ Waist ++ C-peptide ++ Adiponectin --	Waist + C-peptide ++ Adiponectin - -
	C24:1 Ceramide (d18:1)	Metabolic dysregulation		+	CRP + IL6 ++ TNFRSF1B +	CRP ++	BMI ++ Waist ++ C-peptide ++ Adiponectin --	Waist ++ C-peptide ++ Adiponectin - -
Glycinated bile acids and derivatives	Glycocholate [a]	Both	+	+	CRP + IL6 ++ TNFRSF1B +	TNFRSF1B ++ GDF15 ++	Adiponectin -	C-peptide +++
	Glycodeoxycholate [a]	Inflammation		+		TNFRSF1B + GDF15 ++	C-peptide +++	C-peptide ++
N-acyl-alpha amino acids	N-Acetyl-L-alanine [a]	Metabolic dysregulation		+	CRP ++ IL6 ++ TNFRSF1B +++ GDF15 ++	IL6 + TNFRSF1B ++ GDF15 +++	BMI ++ Waist ++ C-peptide +++	C-peptide +

Table 2.3 (Continued) Summary of select metabolites and metabolite subclasses identified as strongest potential mediators of the association of metabolic dysregulation and inflammation with CRC (retained in at least one of the metabolomic signatures, and individually associated with both CRC and markers of inflammation or metabolic dysregulation).

N-acyl-alpha-hexosamines	Acetyl-galactosamine	Inflammation	+	CRP ++ IL6 ++ TNFRSF1B ++ GDF15 ++	TNFRSF1B ++ GDF15 +++	BMI ++ Waist ++ C-peptide ++	C-peptide ++
Purines and derivatives	1-methylguanosine [a]	Both	+	CRP ++ IL6 ++ TNFRSF1B ++	CRP ++ IL6 + TNFRSF1B + GDF15 +++	BMI ++ Waist ++ C-peptide ++ Adiponectin --	BMI +++ Waist +++ C-peptide +
	N2,N2-dimethylguanosine [a]	Both	+	CRP ++ IL6 ++ TNFRSF1B +++ GDF15 ++	CRP ++ IL6 + TNFRSF1B ++ GDF15 +++	BMI ++ Waist ++ C-peptide ++ Adiponectin --	Waist ++ C-peptide ++
Pyrimidines and derivatives	Pseudouridine	Inflammation	+	CRP ++ IL6 ++ TNFRSF1B +++ GDF15 ++	CRP ++ IL6 + TNFRSF1B +++ GDF15 +++	BMI ++ Waist ++ C-peptide ++	C-peptide ++
	5-hydroxymethyl-4-methyluracil [a]	Metabolic dysregulation	+	CRP ++ IL6 ++ TNFRSF1B +++ GDF15 ++	CRP ++ IL6 + TNFRSF1B ++ GDF15 +++	BMI +++ Waist ++ C-peptide ++	C-peptide ++
Ureas	N-carbamoyl-beta-alanine	Inflammation	+	GDF15 ++	CRP ++ IL6 + TNFRSF1B ++ GDF15 +++		C-peptide ++ Adiponectin -- -

Table 2.3 (Continued) Summary of select metabolites and metabolite subclasses identified as strongest potential mediators of the association of metabolic dysregulation and inflammation with CRC (retained in at least one of the metabolomic signatures, and individually associated with both CRC and markers of inflammation or metabolic dysregulation).

The select metabolites shown were: selected by elastic net regression for either the metabolomic signature of inflammation, metabolic dysregulation, or both in all 10 training sets; significantly associated with CRC ($p<0.05$) in multivariable-adjusted logistic regression among men; significantly associated with one or more of the markers of metabolic dysregulation (BMI, waist circumference, C-peptide, adiponectin) or inflammation (CRP, IL6, TNFRSF1B, GDF15) ($p<0.05$) in multivariable-adjusted linear regression in a direction consistent with the expected association of inflammation or metabolic dysregulation with increased CRC incidence.

'+' indicates significant, direct association at $p<0.05$ before FDR correction; '++' indicates significant, direct association at $p<0.05$ after FDR correction; '+++' indicates significant, direct association at $p<0.05$ after FDR correction and mutual adjustment for the other 7 markers of inflammation and metabolic dysregulation (BMI was not mutually adjusted for waist circumference and waist circumference not mutually adjusted for BMI, as this materially changes the interpretation of these variables)

[a] Retained statistically significant association ($p<0.05$) with CRC after limiting to cases diagnosed at least five years after blood collection.

Abbreviations: BMI, body mass index; CRP, C-reactive protein; GDF15, growth differentiation factor 15, Health Professionals Follow-up Study; IL6, interleukin-6; NHS, Nurses' Health Study; TNFRSF1B, tumor necrosis factor receptor super family 1B

DISCUSSION

We derived metabolomic signatures of inflammation and metabolic dysregulation among women and men in two large cohorts, and both signatures explained meaningful variation in markers of inflammation or metabolic dysregulation. Both signatures were associated with increased risk of incident CRC among men but not among women. We identified key metabolites likely to be involved in the inflammatory and metabolic pathways of CRC development.

Sex differences in the relation of inflammatory markers with CRC were not driven by the ability of the signatures to explain the inflammatory markers, which were similar in women and men. Prior studies showed associations of CRP, IL6, and GDF15 with incident CRC among men, but not women.[4,6,20,35] In one study, the associations of CRP and IL6 with CRC were no longer significant after limiting to cases diagnosed at least two years after blood collection, however, we found that the signatures retained an association with CRC even after excluding cases diagnosed within five years.[6] Among men, GDF15 was the inflammatory marker most consistently associated with individual CRC-related metabolites after mutual adjustment for other inflammatory and metabolic biomarkers, consistent with prior findings that GDF15 was associated with almost 3-fold risk of incident CRC among men.[20] GDF15 expression is associated with age-related diseases and may serve as a marker of mitochondrial DNA mutations, oxidative stress, reactive oxygen species production, and other processes associated with biological age.[36] The reason for the observed differences between women and men is not clear; some evidence suggests that estrogen has anti-inflammatory effects, but studies of endogenous estrogen and CRC have been inconsistent,[37] and whether estrogen and other sex hormones modify the association of inflammation with CRC is unknown.[38]

As with the signature of inflammation, the signature of metabolic dysregulation was associated with CRC among men, but not among women, despite explaining similar variation in markers of metabolic dysregulation among both men and women. In previous studies, BMI was associated with increased risk of CRC in both men and women, but less strongly among women

for both measured and genetically predicted BMI.[39,40] Studies of waist circumference and C-peptide with CRC show comparable results among both men and women, and the empirical dietary index for hyperinsulinemia was associated with increased risk of CRC among both men and women.[7,41–44] On the other hand, adiponectin has been associated with decreased risk of CRC among men, with inconsistent results among women.[3,45,46] Adiponectin is more strongly associated with *KRAS*-mutated CRC, and we observed that the metabolic dysregulation signature was nonsignificantly more strongly associated with *KRAS*-mutated CRC than *KRAS*-wild type CRC among men (p, heterogeneity 0.17).[34] Several individual metabolites associated with both CRC and inflammation were also associated with C-peptide or other markers of metabolic dysregulation among men.

Eleven metabolites were retained in either the signature of inflammation or the signature of metabolic dysregulation (or both), and individually associated with CRC, and individually associated with at least one marker of inflammation or metabolic dysregulation. These include two ceramides, a class of metabolites associated with decreased insulin sensitivity, inflammation, red meat consumption, and increased risk of CRC precursors; evidence suggests that upregulated ceramide synthesis is part of a key pathway in intestinal epithelial cells that modulates stem cell proliferation and potentially carcinogenesis.[47–52] Bile acids glycocholate and glycodeoxycholate, previously associated with red meat consumption and unhealthy dietary patterns, have been shown to promote inflammation in mouse models, and were associated with incident CRC in other cohorts.[52–55] Pseudouridine, 1-methylguanosine, and N2,N2-dimethylguanosine levels were elevated in individuals with prevalent CRC, and we found that 1-methylguanosine and N2,N2-dimethylguanosine were associated with incident CRC more than five years after blood collection, indicating that these metabolites may contribute to the development of CRC.[56,57] Although N2,N2-dimethylguanosine levels have not been directly linked to inflammation, they are associated with unhealthy dietary patterns, particularly the consumption of ultra-processed foods and sugar-sweetened beverages, which have been

implicated in promotion of inflammation.[58,59] Finally, 5-hydroxymethyluracil is a byproduct of oxidative damage and is elevated in individuals with prevalent CRC.[60,61] Less is known about the other metabolites identified and further investigation is warranted.

Strengths of the present study include the prospective design, integrated biomarker and metabolomic data, detailed confounder data, the assessment of CRC subtypes, and several robustness checks of the signature development methods and identification of key metabolites. The consistency among methods for both selected metabolites and association with CRC provides strong evidence that the metabolomic signatures are valid predictors of inflammation, metabolic dysregulation, and CRC.

Several limitations are worth noting: the study population was primarily White, limiting generalizability. Additionally, although well-powered for the primary outcome, we lacked the power to detect effect measure modification by tumor molecular subtype and by pre- and postmenopausal status, factors that might possibly explain the differential associations between women and men. Finally, C-peptide, adiponectin, inflammatory biomarkers, and metabolites were characterized from a single blood collection, restricting our ability to draw any conclusions about whether inflammation and metabolic dysregulation precede the observed metabolomic profiles, or vice versa.

In conclusion, we derived metabolomic signatures of inflammation and metabolic dysregulation and observed increased risk of CRC associated with both signatures among men. Based on our findings, the metabolomic pathways linking inflammation and metabolic dysregulation to CRC include ceramides, glycinated bile acids and derivatives, and modified nucleosides. Additional studies to explain the observed differences between men and women and validate the effects of the signatures among other racial and ethnic populations are warranted. The identification of these signatures contributes to improved understanding of CRC-related metabolomic pathways and could potentially inform the development of targeted

interventions for CRC and other conditions affected by chronic inflammation and metabolic dysregulation.

Proportion of major chronic disease attributable to poor quality diet and inadequate physical activity in the United States, independent of adult body weight

Alaina M Bever, Edward Giovannucci, A Heather Eliassen
Liming Liang, Meir Stampfer, and Mingyang Song

ABSTRACT

Background: The distinct effects of lifestyle intervention independent of adiposity are not well quantified. We estimated the proportion and number of incident chronic disease cases in the United States (US) attributable to poor quality diet and inadequate physical activity, before and after accounting for body weight change.

Methods: We followed 199,814 individuals in the Nurses' Health Study, Nurses' Health Study II, and Health Professionals Follow-up Study who provided diet and lifestyle data via biennial questionnaires. Diet quality was evaluated using the Alternative Healthy Eating Index 2010 (AHEI, modified to exclude alcohol and trans fat components) and physical activity was defined as total metabolic equivalent of task (MET)-hours of recreational activity per week. New diagnoses of type 2 diabetes, cardiovascular disease (CVD), and cancer were self-reported and confirmed via medical record. Multivariable-adjusted pooled logistic regression estimated risk ratios by joint effect of diet quality and physical activity, first independent of baseline BMI and then additionally independent of change in body weight over follow-up. Population attributable fractions (PAFs) were estimated using risk factor distribution and case incidence in the US.

Results: We documented 64,181 chronic disease cases over 3.7 million person-years of follow-up (16,667 type 2 diabetes, 16,240 CVD, 31,274 cancer). The combination of high dietary quality (AHEI score > 50 vs. <30) and physical activity (> 28 vs. <7 MET-hours/week) was associated with 23% lower risk of chronic disease in multivariable models (95% CI 18–38%); this estimate was reduced to 17% after adjustment for change in body weight (95% CI 12–32%). Associations varied substantially by outcome, ranging from 0.50 (95% CI 0.44–0.56) for diabetes to 0.99 for cancer (95% CI 0.91–1.08). In the US, 18% (95% CI 5–26%) of chronic

disease was attributable to poor quality diet and low physical activity, with 13% (95% CI 0–23%) not mediated by effects on body weight.

Conclusion: A significant fraction of US chronic disease was attributable to poor quality diet and inadequate physical activity in adulthood, independent of starting BMI and weight change.

INTRODUCTION

Population attributable fraction (PAF) estimates the proportion of cases attributable to modifiable risk factors and can be used to inform public health interventions. Current PAF estimates for major chronic diseases in the United States (US) typically combine body mass index (BMI), diet, and physical activity, or use risk ratios of diet and physical activity that are not adjusted for BMI, making it difficult to disentangle the distinct effects of excess adiposity and obesity-related risk factors.[1–3] In 2024, 82 percent of US adults have overweight or obese BMI, and considering the limited success of lifestyle interventions for long-term weight loss, there is a need to evaluate the impact of diet and physical activity on health independent of their effects on body weight. Moreover, the emerging efficacy of pharmaceutical weight loss interventions raises new questions about the weight-independent benefits of diet quality and physical activity.

We evaluated the effects of diet quality and physical activity on risk of three prevalent noncommunicable diseases—type 2 diabetes, cardiovascular disease (CVD), and cancer—independent of baseline BMI and changes in body weight using longitudinal data from the Nurses' Health Study (NHS), Nurses' Health Study II (NHSII), and Health Professionals Follow-up Study (HPFS). We used nationally representative data to estimate the PAF of each condition due to poor quality diet and inadequate physical activity in the United States in 2018.

METHODS

Risk ratio estimates

Study population

The NHS enrolled 121,700 female registered nurses aged 30–55 years in 1976, NHSII enrolled 116,686 nurses 25–42 years in 1989, and HPFS enrolled 51,529 male health professionals 40–75 years in 1986. Participants completed questionnaires with details regarding medical history and lifestyle factors at enrollment and every two years thereafter; dietary questionnaires were administered in 1984 (NHS), 1991 (NHS2) or 1986 (HPFS), and every four years thereafter.

This study was approved by the institutional review boards at the Harvard T. H. Chan School of Public Health and Brigham and Women's Hospital and participating registries as required.

Baseline and follow-up definition

To emulate a randomized trial, models were adjusted for baseline diet, lifestyle, and covariates, therefore, we derived initial exposure from the second questionnaire cycle with available data on diet and physical activity. Follow-up began two years after the initial exposure to reduce the potential for reverse causation and ended with the first case diagnosis, death, or end of follow-up (NHS follow-up, 1990–2016; NHSII, 1997–2017; HPFS, 1992–2016).

Inclusion criteria

We included participants who were alive and free of the outcome of interest at baseline. We excluded participants below the 1st or above the 99th percentile of reported values for body mass index (BMI), change in body weight, and daily energy intake and participants above the 99th percentile of reported values for physical activity, based on the assumption that these data represented outliers or measurement error.

Physical activity assessment

Participants reported average time per week spent engaging in various recreational activities including walking, running, and other aerobic sports. Self-reported minutes per week were converted to MET-hours per week using previously derived activity-specific estimates.[4] Correlations between self-reported physical activity from questionnaires and physical activity recorded in four single-week diaries administered across four different seasons ranged from 0.62–0.65,[5,6] and the correlation between self-reported physical activity from questionnaires and self-reported resting heart rate among men was -0.21.[7]

Diet quality assessment

Participants reported usual diet through a validated semi-quantitative food frequency questionnaire (FFQ) evaluating average intake of over 130 items. The Alternative Healthy Eating Index-2010 (AHEI) was developed as an evidence-based alternative to the Healthy Eating Index and has been associated with reduced risk of chronic disease,[8,9] and we derived a modified version of the AHEI score to capture diet quality using the traditional method of assigning 0–10 points for each component (with 10 points indicating optimal and 0 points indicating least optimal) but excluding two of the 11 components: alcohol intake and trans fat. We excluded the alcohol component because an optimal AHEI score requires moderate alcohol consumption and we aimed to make the results generalizable to the entire population, including individuals who cannot or choose not to consume alcohol. We excluded the trans fat component because the majority of trans fat was removed from the US food supply by 2018, due to the FDA's determination that partially hydrogenated oils are no longer "Generally Recognized as Safe."[10] FFQ-derived AHEI scores have shown good reproducibility (energy-adjusted intraclass correlation for AHEI derived from two FFQs taken one year apart = 0.81–0.84) and validity (energy-adjusted and deattenuated Spearman rank correlation coefficient for AHEI derived from FFQ and AHEI from 7-day diet records = 0.77–0.80).[11]

Exposure definition

For the main exposure we used cumulative average physical activity and diet in the 18 years preceding a diagnosis of chronic disease or end of follow-up. For missing data on AHEI or physical activity, data from the most recent prior questionnaire were carried forward up to 8 years and were otherwise set to missing. AHEI and physical activity were classified into four intervals based on the distribution in the cohorts and considering previous estimates required for an observable effect.

Covariate assessment

Multivariable models included the following covariates: height; body mass index (BMI); cardiometabolic risk (high or low, based on self-reported use of anti-hypertensive, cholesterol-lowering, or glycemic control prescription drugs or history of high cholesterol, hypertension, or angina); regular aspirin use; alcohol consumption; total energy intake; menopausal status; post-menopausal hormone use (current, former, never); pack-years of smoking and smoking status (current, former with > 10 years since quitting, former with < 10 years since quitting, never smoker); family history of type 2 diabetes, CVD, or cancer; estimated median income based on zip code; and race (White or other). Covariates were modeled as baseline only (height, BMI, cardiometabolic risk, race), baseline and cumulative average over follow-up (aspirin, alcohol, energy intake, income), or baseline and most recent update (menopausal status, post-menopausal hormone use, pack-years of smoking, smoking status, family history). Missing covariate data were first imputed using last observation carried forward and, if missingness remained, we conducted imputation using cohort-specific median value . Risk ratios were estimated before and after adjustment for change in body weight over follow-up, modeled as continuous change in body weight (from baseline to exposure period) as percent of baseline body weight. Models were adjusted for baseline prevalence of the other outcomes (e.g., cancer models were adjusted for baseline type 2 diabetes and CVD).

Case assessment

Diagnoses of type 2 diabetes, CVD, and cancer were self-reported on biennial questionnaires; with permission, study physicians blinded to risk factors reviewed medical records to confirm diagnoses and extract relevant details. Type 2 diabetes (ICD-8 code 250) was confirmed according to the National Diabetes Data Group criteria (before 1988) or the American Diabetes Association criteria (after 1988) via a supplementary questionnaire.[12,13] CVD was defined as the first occurrence of coronary heart disease (fatal or nonfatal myocardial infarction, ICD-8 codes

410, 412) or stroke (fatal or nonfatal, ICD-8 codes 430–438). Nonfatal myocardial infarction was confirmed according to the World Health Organization criteria and nonfatal stroke was confirmed according to the National Survey of Stroke criteria.[14,15] Fatal coronary heart disease and fatal stroke were confirmed by death certificate and evidence from medical records. Cancer was defined as the first occurrence of any cancer type (ICD-8 codes 140–209); due to the very high prevalence of non-melanoma skin cancer and prostate cancer, the majority of which do not cause significant morbidity, we excluded non-melanoma skin cancer and non-lethal prostate cancer (with lethal defined as metastatic or fatal). Unlike type 2 diabetes and CVD, only some cancer types are associated with obesity; we presented "obesity-related cancer" as a separate outcome defined as the first occurrence of esophageal cancer (ICD-8 code 150), gastric cancer (ICD-8 code 151), colorectal cancer (ICD-8 codes 153-154), liver cancer (ICD-8 code 155), gallbladder cancer (ICD-8 code 156), pancreatic cancer (ICD-8 code 157), postmenopausal breast cancer (ICD-8 code 174), endometrial cancer (ICD-8 code 182), ovarian cancer (ICD-8 code 183), kidney cancer (ICD-8 code 189), thyroid cancer (ICD-8 code 193), and multiple myeloma (ICD-8 code 203). Total chronic disease was defined as the first occurrence of incident type 2 diabetes, CVD, or total cancer.

Statistical analysis: Risk ratios

We used pooled logistic regression to approximate multivariable adjusted hazard ratios under the assumption of low disease incidence within each two-year interval. We tested for effect modification of diet quality by physical activity and of joint diet quality and physical activity by age (≤ 60 years vs. > 60 years), sex, baseline BMI (< 25 kg/m^2 vs. ≥ 25 kg/m^2), and smoking status (never or former smoker > 10 years vs. current or recent smoker ≤ 10 years since quit). A significance threshold of 0.15 was used to detect effect modification.

PAF estimates

Estimation of United States exposure prevalence and chronic disease incidence

We used nationally representative data from the National Health and Nutrition Examination Survey (NHANES) 2017-2018 to estimate the distribution of diet quality and physical activity among US adults. AHEI food scores were derived from Food Patterns Equivalents Database (FPED) food groups converted to match the serving size definitions in the NHS and HPFS. Physical activity was derived by applying 4-METS to moderate and 8-METs to vigorous recreational activity. We then estimated the individual and joint frequency of AHEI and physical activity categories by age group (30-60 years and >60 years). Incident diabetes cases in 2018 by age group (30–60 years old and >60 years old) were obtained from the CDC US Diabetes Surveillance System (USDSS); we approximated incident type 2 diabetes cases by assuming that 2% of all cases were non-type 2 diabetes.[16,17] Incident MI and stroke statistics are not directly available and were approximated using age-specific myocardial infarction incidence rates from Atherosclerosis Risk in Communities (ARIC) Community Surveillance Study data collected from 2004–2009, stroke incidence rates from the Greater Cincinnati/Northern Kentucky Stroke Study in 2015, and US census data.[18,19] Age-specific incident cancer cases were obtained from the National Cancer Institute (NCI) Surveillance, Epidemiology, and End Results Program (SEER).[20] Chronic disease was defined as the sum of incident type 2 diabetes, CVD, and cancer.

Statistical analysis: Population attributable fraction

We estimated disease-specific PAF using the expression introduced by Miettinen and modified for a multicategory risk factor:

$$PAF = \frac{\sum_{x=1}^{x=L} \pi(x) RR_U(x) \frac{RR_C(x) - 1}{RR_C(x)}}{\pi(0) + \sum_{x=1}^{x=L} \pi(x) RR_U(x)}$$

with each exposure category x_i ($x(0)$ is the prevalence of the exposure category hypothesized to minimize disease risk), US exposure prevalence ($\pi(x)$), and unadjusted risk ratio ($RR_U(x)$) and causal risk ratio ($RR_C(x)$) estimates derived from the NHS, NHSII, and HPFS.[21,22] To account for effect modification by age, we estimated age-specific risk ratios and PAFs among individuals aged 30-60 years and > 60 years and estimated the aggregate PAF using on the number of cases in each age stratum, as demonstrated by Hanley:

$$PAF_{aggregate} \frac{\sum_{i=1}^{i=N} cases_i * PAF_i}{\sum_{i=1}^{i=N} cases_i}$$

where N = the total number of strata. The Miettinen's PAF does not depend on the absence of confounding to produce a valid PAF.[21] PAF derived from multivariable-adjusted models can be interpreted as the fraction of cases attributable to poor quality diet and low physical activity; PAF derived from models additionally adjusted for change in body weight can be interpreted as the fraction of cases attributable to poor quality diet and low physical activity not mediated by body weight.

RESULTS

NHS, NHSII, and HPFS cohort participants with highest quality diet (AHEI score > 50 out of 0–90 possible points) and maximal physical activity (> 28 MET-hours/week) were slightly older with lower average BMI and higher income than individuals with lowest quality diet (AHEI score < 30) and minimal physical activity (< 7 MET-hours/week, **Table 3.1**). Compared to the cohorts, the US population had a greater number of individuals with minimal physical activity (US 51%, cohorts 33%) and joint lowest quality diet and minimal physical activity (US 17%, cohorts 9%; **Appendix Figure 3.1**). Over 3.7 million person-years of follow-up, we documented 16,667 incident cases of type 2 diabetes, 16,240 incident cases of CVD, and 31274 incident cases of

cancer. Combined chronic disease, which was limited to the first incident diagnosis of any of the three disease, consisted of 55,085 cases (26% type 2 diabetes, 23% CVD, 51% cancer). CVD consisted of 61% CHD, and breast (28%), colorectal (9%), and lung (9%) were the most common cancer types.

Table 3.1. Participant characteristics at baseline according to joint level of physical activity and diet quality, NHS, NHS2, and HPFS

	Overall	Lowest quality diet (AHEI score < 30) [a] + Minimal physical activity (< 7 MET-hours/week)	Highest quality diet (AHEI score > 50) [a] + Maximal physical activity (>28 MET-hours/week)
Number of individuals	194,329	17,423 (9%)	8,820 (5%)
Age, years	49 (10)	46 (9)	53 (11)
AHEI score (0-90 points)	39 (10)	25 (4)	56 (5)
Activity, MET-hr/week	20 (22)	3 (2)	54 (25)
BMI, kg/m2	25 (5)	26 (5)	24 (4)
BMI category			
Lean BMI (<25 kg/m2), %	54	50	65
Overweight BMI (25-30 kg/m2), %	31	29	29
Obese BMI (>30 kg/m2), %	14	21	6
Postmenopausal, %	28	26	25
Postmenopausal hormone use			
Never PMH use, %	44	48	42
Past PMH use, %	18	17	18
Current PMH use, %	38	35	39
Total energy intake, calories/day	1,810 (530)	1,920 (510)	1,770 (520)
Alcohol, g/day	6 (10)	5 (11)	7 (10)
Height, cm	167 (9)	166 (8)	170 (9)
Non-Hispanic White, %	96	96	95
Median income, USD	45,200 (17,100)	42,100 (14,300)	48,600 (20,300)
Regular multivitamin use, %	43	36	51
Regular aspirin use (>2 tabs/week), %	24	22	26
Pack-years of smoking	9 (15)	10 (17)	9 (14)
Smoking status			
Never smoked, %	55	56	51
Past smoker, quit over 10 years ago, %	22	14	31
Past smoker, quit less than 10 years ago, %	10	9	11
Current smoker, %	13	21	7
Family history of type 2 diabetes, %	19	19	18
Family history of cardiovascular disease, %	29	27	32
Family history of cancer, %	21	19	23

Table 3.1 (Continued) Participant characteristics at baseline according to joint level of physical activity and diet quality, NHS, NHS2, and HPFS

Abbreviations: AHEI, alternative healthy eating index 2010; BMI, body mass index; HPFS, Health Professionals Follow-up Study; MET, metabolic equivalent of task; NHS, Nurses' Health Study; PMH, post-menopausal hormone therapy; USD, United States dollar

[a] Range of possible points for the modified AHEI score (excluding alcohol and trans fat components) = 0–90

Risk ratios

In the multivariable adjusted models, risk of chronic disease was lower for every combination of diet quality and physical activity compared to individuals with lowest quality diet and minimal physical activity, and the lowest risk was observed for individuals with combined highest quality diet and maximal physical activity (RR = 0.77, 95% CI 0.72–0.82; **Table 3.2**). The results were attenuated after further adjustment for change in body weight (RR = 0.83, 95% CI 0.78–0.88). There was no statistically significant effect modification of diet quality by physical activity (p = 0.26).

The joint association of diet quality and physical activity varied by individual outcome (**Figure 3.1**): in the multivariable models, the strongest association was observed for type 2 diabetes (RR comparing highest quality diet and maximal physical activity to lowest quality diet and minimal physical activity = 0.50, 95% CI 0.44–0.56) followed by CVD (RR = 0.70, 95% CI 0.62–0.78). The risk ratio for type 2 diabetes was attenuated by additional adjustment for percent change in body weight over follow-up (RR = 0.65, 95% CI 0.57–0.73), whereas the risk ratio for CVD was unchanged. Joint diet quality and physical activity was not associated with reduced risk of all cancer (RR = 0.99, 95% CI 0.91–1.08) or obesity-related cancer (RR = 0.95, 95% CI 0.85–1.07).

Table 3.2. Risk ratios for chronic disease (type 2 diabetes, major cardiovascular disease, all cancer) by joint effect of physical activity and diet quality

		Diet quality			
		AHEI score < 30 points [a]	**AHEI score 30 – 40 points**	**AHEI score 40 – 50 points**	**AHEI score > 50 points**
Physical activity		No. or RR (95% CI)	No. or RR (95% CI)	No. or RR (95% CI)	No. or RR (95% CI)
< 7 MET-hours/week	Cases	2,954	6,220	4,165	1,148
	Person-years	189,064	370,914	243,400	71,168
	Age-adjusted	Ref	0.91 (0.87 – 0.95)	0.84 (0.80 – 0.88)	0.75 (0.70 – 0.80)
	Multivariable-adjusted (MV) [b]	Ref	0.94 (0.90 – 0.98)	0.90 (0.86 – 0.95)	0.84 (0.78 – 0.91)
	MV + baseline BMI	Ref	0.94 (0.89 – 0.98)	0.90 (0.86 – 0.95)	0.86 (0.80 – 0.93)
	MV + baseline BMI + change in body weight	Ref	0.94 (0.90 – 0.99)	0.92 (0.87 – 0.97)	0.89 (0.82 – 0.96)
7 – 14 MET-hours/week	Cases	1,553	4,781	4,165	1,515
	Person-years	116,056	310,800	263,672	95,344
	Age-adjusted	0.85 (0.80 – 0.91)	0.82 (0.78 – 0.86)	0.76 (0.72 – 0.80)	0.71 (0.66 – 0.75)
	MV	0.88 (0.83 – 0.94)	0.87 (0.83 – 0.92)	0.84 (0.80 – 0.89)	0.83 (0.78 – 0.89)
	MV + baseline BMI	0.92 (0.86 – 0.98)	0.91 (0.87 – 0.96)	0.88 (0.83 – 0.93)	0.89 (0.83 – 0.96)
	MV + baseline BMI + change in body weight	0.92 (0.87 – 0.98)	0.93 (0.88 – 0.97)	0.90 (0.86 – 0.95)	0.93 (0.87 – 1.00)
7 – 28 MET-hours/week	Cases	1,465	5,013	5,689	2,414
	Person-years	111,244	352,274	370,978	174,938
	Age-adjusted	0.81 (0.76 – 0.86)	0.74 (0.70 – 0.77)	0.72 (0.69 – 0.75)	0.61 (0.58 – 0.64)
	MV	0.86 (0.80 – 0.91)	0.80 (0.76 – 0.84)	0.81 (0.77 – 0.86)	0.73 (0.69 – 0.78)
	MV + baseline BMI	0.91 (0.85 – 0.97)	0.86 (0.82 – 0.90)	0.88 (0.83 – 0.92)	0.81 (0.76 – 0.86)
	MV + baseline BMI + change in body weight	0.92 (0.87 – 0.99)	0.88 (0.84 – 0.93)	0.91 (0.87 – 0.96)	0.85 (0.80 – 0.91)

Table 3.2 (Continued) Risk ratios for chronic disease (type 2 diabetes, major cardiovascular disease, all cancer) by joint effect of physical activity and diet quality

> 28 MET-hours/week	Cases	985	3,731	5,046	3,090
	Person-years	78,912	282,460	368,208	242,012
	Age-adjusted	0.72 (0.67 – 0.78)	0.65 (0.62 – 0.68)	0.61 (0.58 – 0.64)	0.54 (0.51 – 0.57)
	MV	0.79 (0.74 – 0.86)	0.73 (0.69 – 0.77)	0.72 (0.68 – 0.76)	0.67 (0.63 – 0.71)
	MV + baseline BMI	0.85 (0.79 – 0.92)	0.80 (0.76 – 0.84)	0.79 (0.75 – 0.84)	0.77 (0.72 – 0.82)
	MV + baseline BMI + change in body weight	0.88 (0.82 – 0.95)	0.83 (0.79 – 0.88)	0.84 (0.79 – 0.89)	0.83 (0.78 – 0.88)

Abbreviations: BMI, body mass index; AHEI, alternative healthy eating index; HPFS, Health Professionals Follow-up Study; MET, metabolic equivalent of task; MV, multivariable; NHS, Nurses' Health Study

[a] Range of possible points for the modified AHEI score (excluding alcohol and trans fat components) = 0–90

[b] Multivariable models adjusted for cohort (NHS, NHSII, HPFS); age (linear and quadratic); height; baseline metabolic risk (high-risk or low-risk, based on self-reported use of anti-hypertensive, cholesterol-lowering, or glycemic control prescription drugs or history of high cholesterol, hypertension, or angina); baseline physical activity; baseline diet score; baseline and follow-up total energy intake; baseline and follow-up alcohol intake; baseline and follow-up multivitamin use (yes or no); baseline and follow-up regular aspirin use (yes or no); baseline and follow-up menopausal status; baseline and follow-up post-menopausal hormone use (current, former, never); baseline and follow-up pack-years of smoking and smoking status (current, former, never smoker); baseline and follow-up reported family history of type 2 diabetes, cardiovascular disease, or cancer; and baseline and follow-up median income (from census data and zip code). All confounders modeled as linear variables unless otherwise noted.

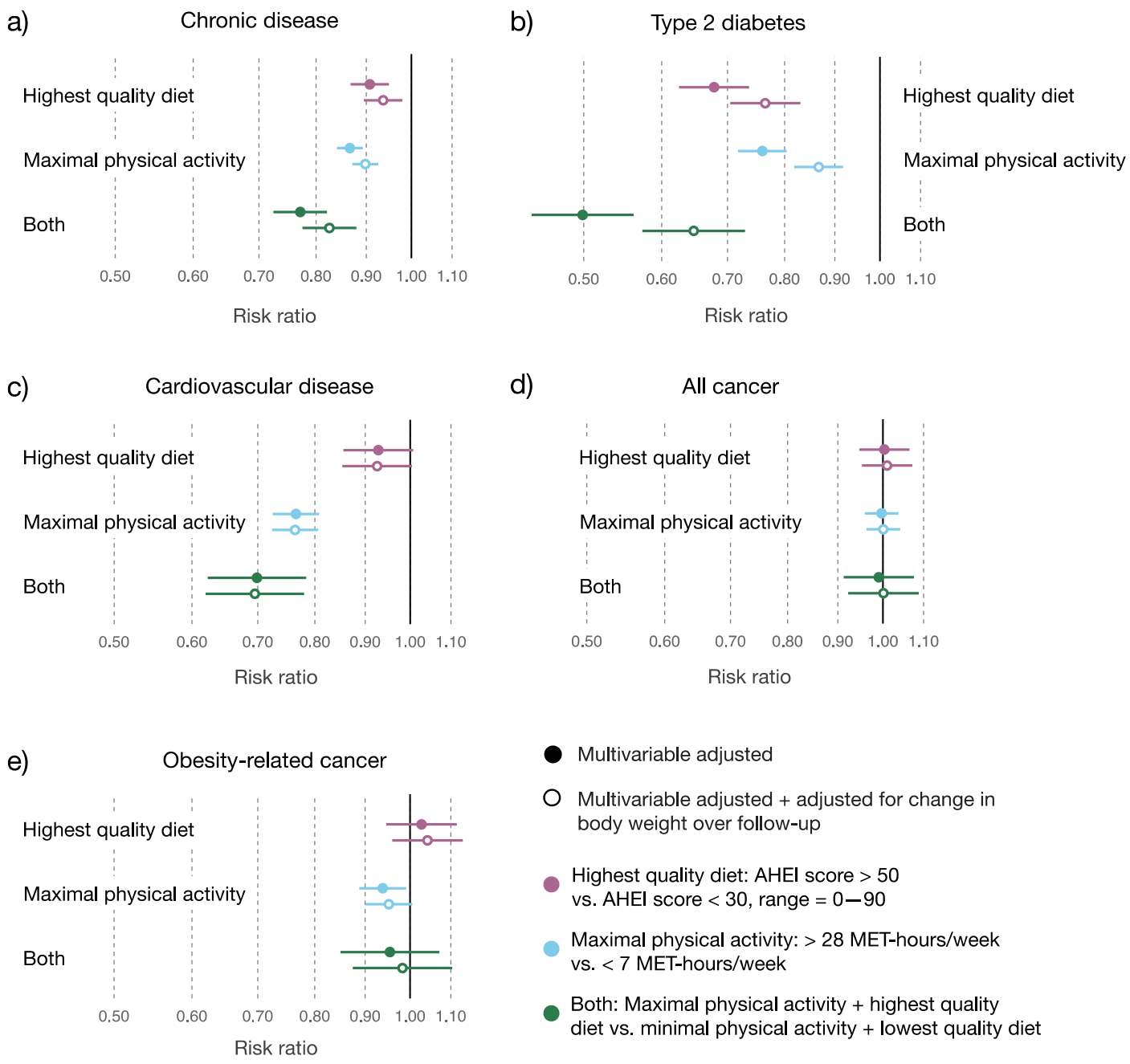


Figure 3.1. Risk ratios for combined and individual chronic disease by highest vs. lowest quality diet (AHEI score > 50 vs. < 30), maximal vs. minimal physical activity (> 28 MET-hours/week vs. < 7 MET-hours/week), and joint highest quality diet and maximal physical activity vs. lowest quality diet and minimal physical activity

Abbreviations: AHEI, alternative healthy eating index; MET, metabolic equivalent of task

For chronic disease, the association of physical activity was slightly stronger than diet quality, and the confidence intervals for the two estimates overlapped (diet RR = 0.91, 95% CI 0.87–1.95; activity RR = 0.87, 95% CI 0.84–0.89). The association of highest compared to lowest diet quality was slightly stronger than the effect of maximal compared to minimal physical activity for type 2 diabetes (diet RR = 0.68, 95% CI 0.62–0.74; activity RR = 0.76, 95% CI 0.72–0.80), whereas the association of physical activity was stronger than that of diet quality for CVD (diet RR = 0.93, 95% CI 0.86–1.01; activity RR = 0.77, 95% CI 0.72–0.81). Although the joint association was not significant for obesity-related cancer, there was a statistically significant association effect of physical activity for obesity-related cancer (RR = 0.94, 95%CI 0.89–0.99) that was virtually unchanged after adjustment for change in body weight.

We observed significant effect modification for the joint association of diet quality and physical activity on chronic disease by age ($p = 0.08$) and sex ($p = 0.003$), on type 2 diabetes by age ($p = 0.01$) and sex ($p = 0.12$), on CVD by sex ($p = 0.11$), and on cancer by baseline BMI ($p = 0.07$). In stratified models, the joint effect on chronic disease was stronger among individuals ≤ 60 years old compared to individuals > 60 years old and among men compared to women (**Figure 3.2)**. No effect modification was observed by smoking status for combined or individual chronic disease.

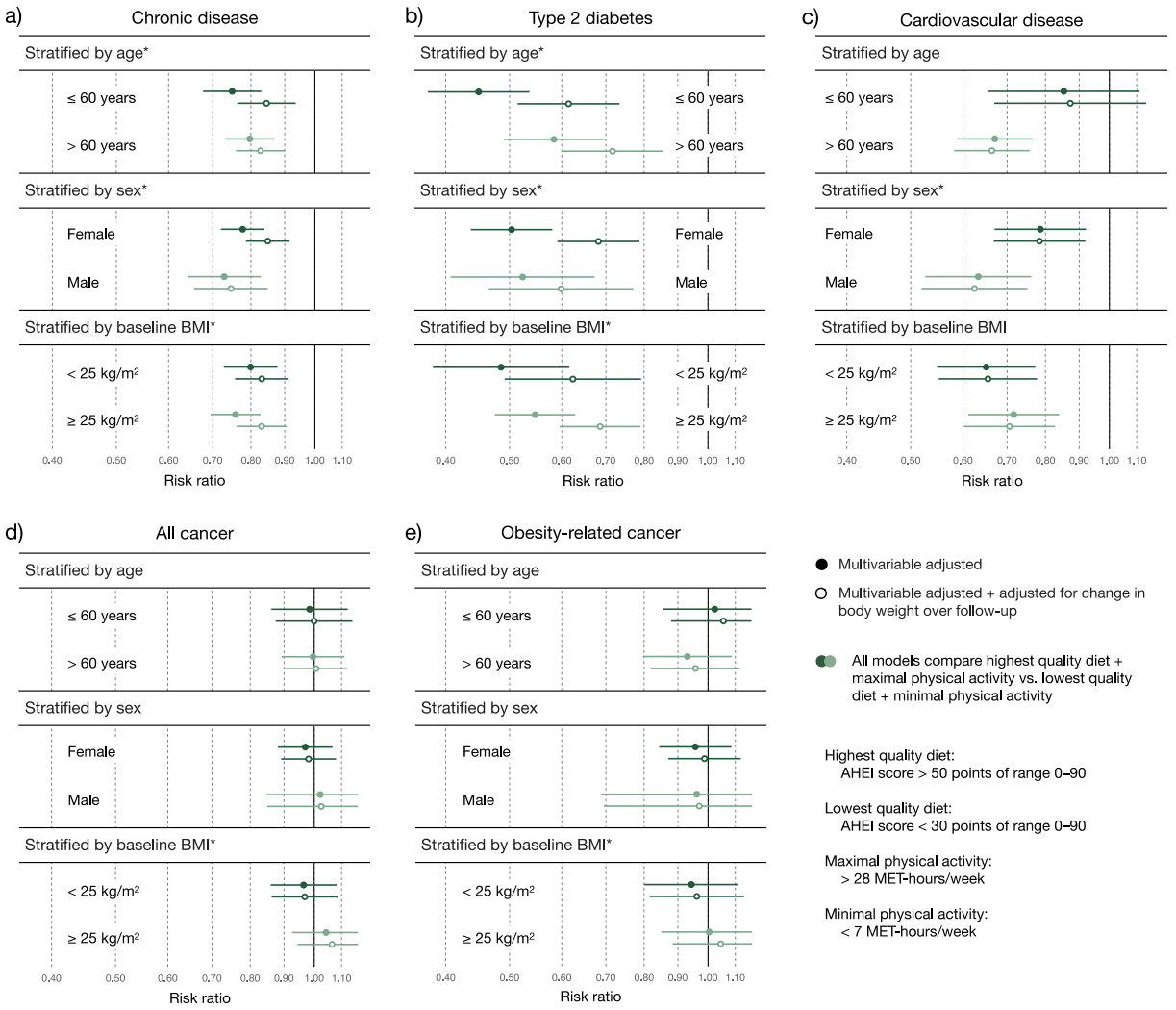


Figure 3.2. Risk ratios for combined and individual chronic disease) by joint highest quality diet and maximal physical activity vs. lowest quality diet and minimal physical activity, stratified by baseline age, baseline BMI, and sex.

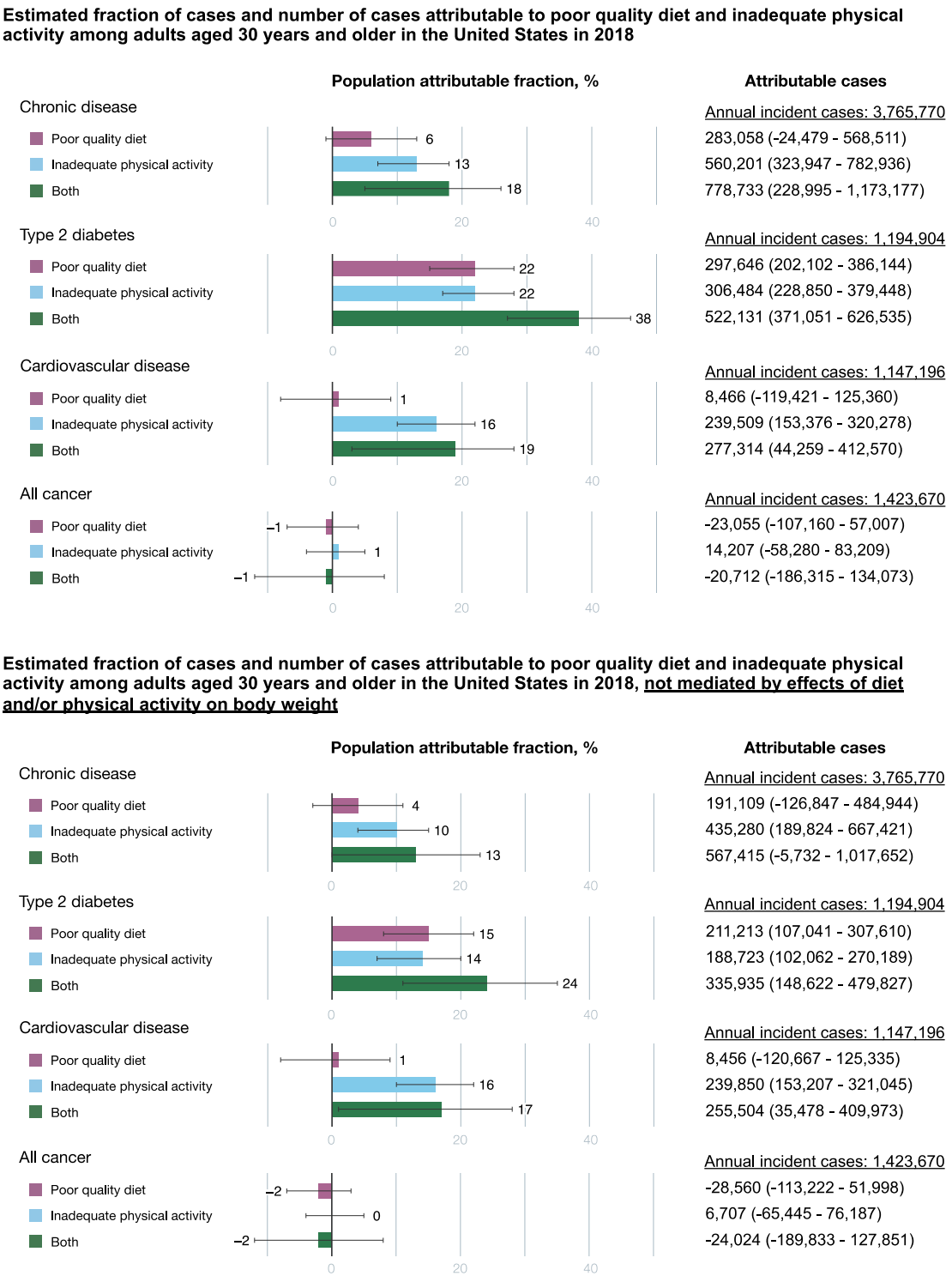
Abbreviations: AHEI, alternative healthy eating index; MET, metabolic equivalent of task

Population attributable fraction

The estimated fraction of cases attributable to combined poor quality diet and inadequate physical activity in the US was 18% for chronic disease (95% CI 5–26%, **Figure 3.3**). Although risk ratios for individual effects of diet quality and physical activity with chronic disease were similar, a higher fraction of chronic disease cases was attributable to inadequate physical activity (13%, 95% CI 7–18%) than poor quality diet (6%, 95% CI -1–13) due to the substantial prevalence of low physical activity in the US population. The PAF was 38% for type 2 diabetes (95% CI 27–46) and 19% for CVD (95% CI 3–28%). The estimated fraction of cases attributable to combined poor quality diet and inadequate physical activity not mediated by body weight was 13% for chronic disease (95% CI 0–23%), 24% for type 2 diabetes (95% CI 11–35%), and 17% for CVD (95% CI 1–28%).

Figure 3.3. Population attributable fraction, incident cases of combined and individual chronic disease in the United States, 2018. PAF is derived from the estimated number of cases given the prevalence of poor quality diet (AHEI score < 30, 30–40, and 40–50 with possible range 0–90) and inadequate physical activity (< 7, 7-14, and 14-28 MET-hours/week) among US adults in 2018 compared to the estimated number of cases had all US adults met the threshold for high quality diet (AHEI score > 50 with possible range 0–90) and/or adequate physical activity (> 28 MET-hours/week).

Figure 3.3 (Continued) Population attributable fraction, incident cases of combined and individual chronic disease in the United States, 2018.



Abbreviations: AHEI, alternative healthy eating index; MET, metabolic equivalent of task

DISCUSSION

In a prospective cohort study with 64,181 cases of chronic disease, we found that high quality diet as assessed by AHEI score and high physical activity were associated with decreased risk of type 2 diabetes and CVD independent of both baseline BMI and change in body weight over follow-up. Using the derived risk ratios, we estimated that 18% or 778,733 cases of chronic disease in the US in 2018 were attributable to poor quality diet and inadequate physical activity in adulthood. A larger fraction of cases in the US were attributable to physical activity than poor diet quality, driven by the substantial fraction of US adults obtaining less than 7 MET-hours per week of recreational physical activity (7 MET-hours is the equivalent of 60-90 minutes of walking).

Our findings are consistent with existing research on the impact of diet quality and physical activity on chronic disease risk. In a study covering multiple risk factors for chronic disease, higher fruit and vegetable consumption (a proxy for diet quality) was associated with a 15% decrease in chronic disease risk for women and 9% for men; higher physical activity was associated with a 8% decrease in risk for women and 22% for men.[23] Our PAF estimations for type 2 diabetes are also consistent with prior findings, for example among individuals in the Multiethnic Cohort study, it was estimated that 29% of type 2 diabetes cases in women and 13% in men were attributable to low physical activity and 23% of cases in women and 27% in men attributable to processed red meat intake, independent of BMI.[24]

Our PAF estimates for CVD and physical activity are consistent with a prior study in the NHS that estimated that 19% of CHD were attributable to low physical activity, however, the same study estimated that 21% of CHD cases were attributable to low AHEI score, which is much higher than our estimate of 1% of CHD cases attributable to poor quality diet.[25] This discrepancy is most likely driven by our modification of the AHEI score to exclude alcohol and

trans fat, two factors that are strongly associated with CHD (moderate alcohol consumption with decreased CHD risk and trans fat with increased CHD risk).

We found that no cancer cases were attributable to poor quality diet or inadequate physical activity independent of baseline BMI; one prior study found that the fraction of cancer cases in the United States due to poor diet was 4%—however, that PAF was derived by combining risk ratios for individual dietary factors from various meta-analyses, whereas the use of a dietary pattern in our study takes into account correlation among dietary factors.[26] Another study found that 3% of all cancer cases in the US were attributable to physical inactivity; this PAF was derived from risk ratios not adjusted for body mass index, likely explaining the higher attributable proportion compared to the present study.

The longitudinal design of our study allowed us to evaluate the effects of diet and physical activity independent of starting BMI and future changes in body weight. By including numerous baseline confounders, we aimed to derive risk ratios that closely approximate causal estimates. Another major strength of this study is the direct comparison of the effects of diet and physical activity on three distinct chronic diseases, which revealed intriguing differences. Hypotheses proposing common etiologic pathways for type 2 diabetes and obesity-related cancer, particularly insulin resistance, imply overlapping preventive lifestyle factors for the two conditions. However, we observed a dramatic effect of diet and lifestyle for the prevention of type 2 diabetes and, surprisingly, no effect for obesity-related or total cancer. The lack of effect on obesity-related cancer among adults of all ages, both before and after accounting for changes in body weight, lends support to the notion that pre-adult body size may play a significant role in obesity-related cancers. Estimates for CVD also differed from type 2 diabetes in that they were unaffected by adjustment for change in body weight. This finding highlights the importance of non-weight-related factors in CVD prevention. The observed difference for CVD could also be influenced by the older age of CVD onset compared to diabetes, which is more likely to occur after the peak time of adult weight gain.

Despite large sample size and high number of cases, there were sources of uncertainty in our PAF results and although we were able to estimate the PAF using age group stratification, the data were not powered to further stratify by other potential effect modifiers including gender and baseline BMI. Our risk ratio estimates were derived among a cohort with a higher proportion female and lower proportion overweight/obese compared to the US population; because effect modification estimates indicated a stronger effect of diet and physical activity among men and individuals with overweight/obese BMI, bias resulting from our failure to account for effect modification in the PAR estimates is likely to result in an underestimated PAF. Another major limitation is that although the maximum possible modified AHEI score was 90, less than 1 percent of the NHS/HPFS population obtained a score greater than 60. Thus, our estimates for the fraction of chronic disease attributable to poor diet were limited by the maximum healthfulness of the diets consumed by individuals in the cohorts. For both diet and physical activity, measurement error and variation individual response to diet and physical activity may have contributed to error. Future studies incorporating measures of response to lifestyle interventions, such as VO_2 max, visceral adiposity, lipids, and insulin sensitivity, may elucidate more effective diet and physical activity interventions and an even greater fraction of chronic disease considered attributable to diet and physical activity independent of body mass.

BIBLIOGRAPHY

Chapter 1 References

1. Johnson CM, Wei C, Ensor JE, et al. Meta-analyses of Colorectal Cancer Risk Factors. *Cancer Causes Control*. 2013;24(6):1207-1222. doi:10.1007/s10552-013-0201-5

2. Kobiela J, Wieszczy P, Reguła J, Kamiński MF. Association of obesity with colonic findings in screening colonoscopy in a large population-based study. *United European Gastroenterology Journal*. 2018;6(10):1538-1546. doi:10.1177/2050640618803825

3. Frezza EE, Wachtel MS, Chiriva-Internati M. Influence of obesity on the risk of developing colon cancer. *Gut*. 2006;55(2):285-291. doi:10.1136/gut.2005.073163

4. He X, Wu K, Ogino S, Giovannucci EL, Chan AT, Song M. Association Between Risk Factors for Colorectal Cancer and Risk of Serrated Polyps and Conventional Adenomas. *Gastroenterology*. 2018;155(2):355-373.e18. doi:10.1053/j.gastro.2018.04.019

5. Hoffmeister M, Blaker H, Kloor M, et al. Body Mass Index and Microsatellite Instability in Colorectal Cancer: A Population-based Study. *Cancer Epidemiology Biomarkers & Prevention*. 2013;22(12):2303-2311. doi:10.1158/1055-9965.EPI-13-0239

6. Song M, Zheng Y, Qi L, Hu FB, Chan AT, Giovannucci EL. Longitudinal Analysis of Genetic Susceptibility and BMI Throughout Adult Life. *Diabetes*. 2018;67(2):248-255. doi:10.2337/db17-1156

7. Thrift AP, Gong J, Peters U, et al. Mendelian Randomization Study of Body Mass Index and Colorectal Cancer Risk. *Cancer Epidemiology Biomarkers & Prevention*. 2015;24(7):1024-1031. doi:10.1158/1055-9965.EPI-14-1309

8. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007;39(7):870-874. doi:10.1038/ng2075

9. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206. doi:10.1038/nature14177

10. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006;38(8):904-909. doi:10.1038/ng1847

11. Gao C, Patel CJ, Michailidou K, et al. Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. *Int J Epidemiol*. 2016;45(3):896-908. doi:10.1093/ije/dyw129

12. Jarvis D, Mitchell JS, Law PJ, et al. Mendelian randomisation analysis strongly implicates adiposity with risk of developing colorectal cancer. *British Journal of Cancer*. 2016;115(2):266-272. doi:10.1038/bjc.2016.188

13. Ben Q, An W, Jiang Y, et al. Body Mass Index Increases Risk for Colorectal Adenomas Based on Meta-analysis. *Gastroenterology*. 2012;142(4):762-772. doi:10.1053/j.gastro.2011.12.050

14. Bailie L, Loughrey MB, Coleman HG. Lifestyle Risk Factors for Serrated Colorectal Polyps: A Systematic Review and Meta-analysis. *Gastroenterology*. 2017;152(1):92-104. doi:10.1053/j.gastro.2016.09.003

15. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis*. 2009;30(7):1073-1081. doi:10.1093/carcin/bgp127

16. Schmidt FM, Weschenfelder J, Sander C, et al. Inflammatory Cytokines in General and Central Obesity and Modulating Effects of Physical Activity. *PLoS One*. 2015;10(3). doi:10.1371/journal.pone.0121971

17. He X, Hang D, Wu K, et al. Long-term Risk of Colorectal Cancer After Removal of Conventional Adenomas and Serrated Polyps. *Gastroenterology*. Published online July 2019. doi:10.1053/j.gastro.2019.06.039

18. Garcia H, Song M. Early-life obesity and adulthood colorectal cancer risk: a meta-analysis. *Rev Panam Salud Publica*. 2019;43. doi:10.26633/RPSP.2019.3

19. Zhang X, Wu K, Giovannucci EL, et al. Early Life Body Fatness and Risk of Colorectal Cancer in U.S. Women and Men--Results from Two Large Cohort Studies. *Cancer Epidemiology Biomarkers & Prevention*. 2015;24(4):690-697. doi:10.1158/1055-9965.EPI-14-0909-T

20. Nock NL, Plummer SJ, Thompson CL, Casey G, Li L. FTO polymorphisms are associated with adult body mass index (BMI) and colorectal adenomas in African-Americans. *Carcinogenesis*. 2011;32(5):748-756. doi:10.1093/carcin/bgr026

21. Lenka G, Tsai MH, Lin HC, et al. Identification of Methylation-Driven, Differentially Expressed STXBP6 as a Novel Biomarker in Lung Adenocarcinoma. *Sci Rep*. 2017;7. doi:10.1038/srep42573

Chapter 2 References

1. Takasu C, Shimada M, Kurita N, et al. Impact of C-reactive protein on prognosis of patients with colorectal carcinoma. *Hepatogastroenterology*. 2013;60(123):507-511. doi:10.5754/hge11425

2. Yamamoto T, Tsunedomi R, Nakajima M, et al. IL-6 Levels Correlate with Prognosis and Immunosuppressive Stromal Cells in Patients with Colorectal Cancer. *Ann Surg Oncol*. 2023;30(8):5267-5277. doi:10.1245/s10434-023-13527-y

3. Song M, Zhang X, Wu K, et al. Plasma Adiponectin and Soluble Leptin Receptor and Risk of Colorectal Cancer: A Prospective Study. *Cancer Prevention Research*. 2013;6(9):875-885. doi:10.1158/1940-6207.CAPR-13-0169

4. Song M, Mehta RS, Wu K, et al. Plasma Inflammatory Markers and Risk of Advanced Colorectal Adenoma in Women. *Cancer Prev Res (Phila)*. 2016;9(1):27-34. doi:10.1158/1940-6207.CAPR-15-0307

5. Otani K, Ishihara S, Yamaguchi H, et al. Adiponectin and colorectal cancer. *Surg Today*. 2017;47(2):151-158. doi:10.1007/s00595-016-1334-4

6. Song M, Wu K, Ogino S, Fuchs CS, Giovannucci EL, Chan AT. A prospective study of plasma inflammatory markers and risk of colorectal cancer in men. *Br J Cancer*. 2013;108(9):1891-1898. doi:10.1038/bjc.2013.172

7. Dai Z, Xu YC, Niu L. Obesity and colorectal cancer risk: A meta-analysis of cohort studies. *World J Gastroenterol*. 2007;13(31):4199-4206. doi:10.3748/wjg.v13.i31.4199

8. Tabung FK, Wang W, Fung TT, et al. Association of dietary insulinemic potential and colorectal cancer risk in men and women. *Am J Clin Nutr*. 2018;108(2):363-370. doi:10.1093/ajcn/nqy093

9. Tabung FK, Liu L, Wang W, et al. Association of Dietary Inflammatory Potential With Colorectal Cancer Risk in Men and Women. *JAMA Oncology*. 2018;4(3):366. doi:10.1001/jamaoncol.2017.4844

10. McCullough ML, Hodge RA, Campbell PT, Stevens VL, Wang Y. Pre-Diagnostic Circulating Metabolites and Colorectal Cancer Risk in the Cancer Prevention Study-II Nutrition Cohort. *Metabolites*. 2021;11(3):156. doi:10.3390/metabo11030156

11. Vidman L, Zheng R, Bodén S, et al. Untargeted plasma metabolomics and risk of colorectal cancer—an analysis nested within a large-scale prospective cohort. *Cancer Metab*. 2023;11:17. doi:10.1186/s40170-023-00319-x

12. Rothwell JA, Bešević J, Dimou N, et al. Circulating amino acid levels and colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition and UK Biobank cohorts. *BMC Med*. 2023;21(1):80. doi:10.1186/s12916-023-02739-4

13. Breeur M, Ferrari P, Dossus L, et al. Pan-cancer analysis of pre-diagnostic blood metabolite concentrations in the European Prospective Investigation into Cancer and Nutrition. *BMC Med*. 2022;20(1):351. doi:10.1186/s12916-022-02553-4

14. Kliemann N, Viallon V, Murphy N, et al. Metabolic signatures of greater body size and their associations with risk of colorectal and endometrial cancers in the European Prospective Investigation into Cancer and Nutrition. *BMC Med*. 2021;19:101. doi:10.1186/s12916-021-01970-1

15. Hankinson SE, Willett WC, Manson JE, et al. Plasma Sex Steroid Hormone Levels and Risk of Breast Cancer in Postmenopausal Women. *JNCI: Journal of the National Cancer Institute*. 1998;90(17):1292-1299. doi:10.1093/jnci/90.17.1292

16. Chu NF, Spiegelman D, Yu J, Rifai N, Hotamisligil GS, Rimm EB. Plasma leptin concentrations and four-year weight gain among US men. *Int J Obes Relat Metab Disord*. 2001;25(3):346-353. doi:10.1038/sj.ijo.0801549

17. Hunter DJ, Hankinson SE, Hough H, et al. A prospective study of NAT2 acetylation genotype, cigarette smoking, and risk of breast cancer. *Carcinogenesis*. 1997;18(11):2127-2132. doi:10.1093/carcin/18.11.2127

18. Townsend MK, Clish CB, Kraft P, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin Chem*. 2013;59(11):1657-1667. doi:10.1373/clinchem.2012.199133

19. Song M, Zhang X, Wu K, et al. Plasma Adiponectin and Soluble Leptin Receptor and Risk of Colorectal Cancer: A Prospective Study. *Cancer Prevention Research*. 2013;6(9):875-885. doi:10.1158/1940-6207.CAPR-13-0169

20. Mehta RS, Song M, Bezawada N, et al. A Prospective Study of Macrophage Inhibitory Cytokine-1 (MIC-1/GDF15) and Risk of Colorectal Cancer. *J Natl Cancer Inst*. 2014;106(4):dju016. doi:10.1093/jnci/dju016

21. Bonser AM, Garcia-Webb P. C-peptide measurement: methods and clinical utility. *Crit Rev Clin Lab Sci*. 1984;19(4):297-352. doi:10.3109/10408368409165766

22. Tabung FK, Wang W, Fung TT, et al. Development and validation of empirical indices to assess the insulinemic potential of diet and lifestyle. *Br J Nutr*. Published online November 8, 2016:1-12. doi:10.1017/S0007114516003755

23. Izenman AJ. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*. 1975;5(2):248-264. doi:10.1016/0047-259X(75)90042-1

24. Hoffmann K. Application of a New Statistical Method to Derive Dietary Patterns in Nutritional Epidemiology. *American Journal of Epidemiology*. 2004;159(10):935-944. doi:10.1093/aje/kwh134

25. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Statistical Soc B*. 2005;67(2):301-320. doi:10.1111/j.1467-9868.2005.00503.x

26. Li J, Guasch-Ferré M, Chung W, et al. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. *European Heart Journal*. Published online May 14, 2020. doi:10.1093/eurheartj/ehaa209

27. Zou H, Hastie T. Regression shrinkage and selection via the elastic net, with applications to microarrays. Published online January 5, 2004.

28. Benjamini Y, Hochberg Y. 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):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

29. Chan AT, Ogino S, Giovannucci EL, Fuchs CS. Inflammatory markers are associated with risk of colorectal cancer and chemopreventive response to anti-inflammatory drugs. *Gastroenterology*. 2011;140(3):799-808, quiz e11. doi:10.1053/j.gastro.2010.11.041

30. Song M, Hu FB, Spiegelman D, 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):871-883. doi:10.1093/ije/dyv177

31. Fang Z, Hang D, Wang K, et al. Risk prediction models for colorectal cancer: Evaluating the discrimination due to added biomarkers. *Int J Cancer*. 2021;149(5):1021-1030. doi:10.1002/ijc.33621

32. Wang L, Lo CH, He X, et al. Risk Factor Profiles Differ for Cancers of Different Regions of the Colorectum. *Gastroenterology*. 2020;159(1):241-256.e13. doi:10.1053/j.gastro.2020.03.054

33. Wang L, He X, Ugai T, et al. Risk Factors and Incidence of Colorectal Cancer According to Major Molecular Subtypes. *JNCI Cancer Spectrum*. 2021;5(1). doi:10.1093/jncics/pkaa089

34. Inamura K, Song M, Jung S, et al. Prediagnosis Plasma Adiponectin in Relation to Colorectal Cancer Risk According to KRAS Mutation Status. *J Natl Cancer Inst*. 2016;108(4):djv363. doi:10.1093/jnci/djv363

35. Wang SW, Sun YM. The IL-6/JAK/STAT3 pathway: Potential therapeutic strategies in treating colorectal cancer (Review). *International Journal of Oncology*. 2014;44(4):1032-1040. doi:10.3892/ijo.2014.2259

36. Conte M, Giuliani C, Chiariello A, Iannuzzi V, Franceschi C, Salvioli S. GDF15, an emerging key player in human aging. *Ageing Res Rev*. 2022;75:101569. doi:10.1016/j.arr.2022.101569

37. Bouras E, Papandreou C, Tzoulaki I, Tsilidis KK. Endogenous sex steroid hormones and colorectal cancer risk: a systematic review and meta-analysis. *Discov Oncol*. 2021;12:8. doi:10.1007/s12672-021-00402-z

38. Straub RH. The Complex Role of Estrogens in Inflammation. *Endocrine Reviews*. 2007;28(5):521-574. doi:10.1210/er.2007-0001

39. Bull CJ, Bell JA, Murphy N, et al. Adiposity, metabolites, and colorectal cancer risk: Mendelian randomization study. *BMC Med*. 2020;18:396. doi:10.1186/s12916-020-01855-9

40. Jochem C, Leitzmann M. Obesity and Colorectal Cancer. *Recent Results Cancer Res*. 2016;208:17-41. doi:10.1007/978-3-319-42542-9_2

41. Wei EK. A Prospective Study of C-Peptide, Insulin-like Growth Factor-I, Insulin-like Growth Factor Binding Protein-1, and the Risk of Colorectal Cancer in Women. *Cancer Epidemiology Biomarkers & Prevention*. 2005;14(4):850-855. doi:10.1158/1055-9965.EPI-04-0661

42. Kaaks R, Toniolo P, Akhmedkhanov A, et al. Serum C-Peptide, Insulin-Like Growth Factor (IGF)-I, IGF-Binding Proteins, and Colorectal Cancer Risk in Women. *J Natl Cancer Inst*. 2000;92(19):1592-1600. doi:10.1093/jnci/92.19.1592

43. Ma J, Giovannucci E, Pollak M, et al. A prospective study of plasma C-peptide and colorectal cancer risk in men. *J Natl Cancer Inst*. 2004;96(7):546-553. doi:10.1093/jnci/djh082

44. Pischon T, Lahmann PH, Boeing H, et al. Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst*. 2006;98(13):920-931. doi:10.1093/jnci/djj246

45. Aleksandrova K, Drogan D, Boeing H, et al. Adiposity, mediating biomarkers and risk of colon cancer in the European prospective investigation into cancer and nutrition study. *International Journal of Cancer*. 2014;134(3):612-621. doi:10.1002/ijc.28368

46. Fukumoto J, Otake T, Tajima O, et al. Adiponectin and colorectal adenomas: Self Defense Forces Health Study. *Cancer Sci*. 2008;99(4):781-786. doi:10.1111/j.1349-7006.2008.00745.x

47. Li Y, Nicholson RJ, Summers SA. Ceramide signaling in the gut. *Mol Cell Endocrinol*. 2022;544:111554. doi:10.1016/j.mce.2022.111554

48. Li Y, Chaurasia B, Rahman MM, et al. Ceramides Increase Fatty Acid Utilization in Intestinal Progenitors to Enhance Stemness and Increase Tumor Risk. *Gastroenterology*. 2023;165(5):1136-1150. doi:10.1053/j.gastro.2023.07.017

49. Hang D, Zeleznik OA, Lu J, et al. Plasma metabolomic profiles for colorectal cancer precursors in women. *Eur J Epidemiol*. 2022;37(4):413-422. doi:10.1007/s10654-021-00834-5

50. Xia QS, Lu FE, Wu F, et al. New role for ceramide in hypoxia and insulin resistance. *World J Gastroenterol*. 2020;26(18):2177-2186. doi:10.3748/wjg.v26.i18.2177

51. Field BC, Gordillo R, Scherer PE. The Role of Ceramides in Diabetes and Cardiovascular Disease Regulation of Ceramides by Adipokines. *Frontiers in Endocrinology*. 2020;11. Accessed March 16, 2023. https://www.frontiersin.org/articles/10.3389/fendo.2020.569250

52. Wang F, Chandler PD, Zeleznik OA, et al. Plasma Metabolite Profiles of Red Meat, Poultry, and Fish Consumption, and Their Associations with Colorectal Cancer Risk. *Nutrients*. 2022;14(5):978. doi:10.3390/nu14050978

53. Ocvirk S, O'Keefe SJ. Influence of Bile Acids on Colorectal Cancer Risk: Potential Mechanisms Mediated by Diet - Gut Microbiota Interactions. *Curr Nutr Rep*. 2017;6(4):315-322. doi:10.1007/s13668-017-0219-5

54. Kühn T, Stepien M, López-Nogueroles M, et al. Prediagnostic Plasma Bile Acid Levels and Colon Cancer Risk: A Prospective Study. *J Natl Cancer Inst*. 2019;112(5):516-524. doi:10.1093/jnci/djz166

55. Cross AJ, Moore SC, Boca S, et al. A prospective study of serum metabolites and colorectal cancer risk. *Cancer*. 2014;120(19):3049-3057. doi:10.1002/cncr.28799

56. Seidel A, Brunner S, Seidel P, Fritz GI, Herbarth O. Modified nucleosides: an accurate tumour marker for clinical diagnosis of cancer, early detection and therapy control. *Br J Cancer*. 2006;94(11):1726-1733. doi:10.1038/sj.bjc.6603164

57. Zhu J, Djukovic D, Deng L, et al. Targeted serum metabolite profiling and sequential metabolite ratio analysis for colorectal cancer progression monitoring. *Anal Bioanal Chem*. 2015;407(26):7857-7863. doi:10.1007/s00216-015-8984-8

58. Su D, Chen J, Du S, et al. Metabolomic Markers of Ultra-Processed Food and Incident CKD. *Clin J Am Soc Nephrol*. 2023;18(3):327-336. doi:10.2215/CJN.0000000000000062

59. Wang F, Baden MY, Guasch-Ferré M, et al. Plasma metabolite profiles related to plant-based diets and the risk of type 2 diabetes. *Diabetologia*. 2022;65(7):1119-1132. doi:10.1007/s00125-022-05692-8

60. Olinski R, Starczak M, Gackowski D. Enigmatic 5-hydroxymethyluracil: Oxidatively modified base, epigenetic mark or both? *Mutat Res Rev Mutat Res*. 2016;767:59-66. doi:10.1016/j.mrrev.2016.02.001

61. Starczak M, Zarakowska E, Modrzejewska M, et al. In vivo evidence of ascorbate involvement in the generation of epigenetic DNA modifications in leukocytes from patients with colorectal carcinoma, benign adenoma and inflammatory bowel disease. *J Transl Med*. 2018;16:204. doi:10.1186/s12967-018-1581-9

Chapter 3 References

1. Chiuve SE, Rexrode KM, Spiegelman D, Logroscino G, Manson JE, Rimm EB. Primary prevention of stroke by healthy lifestyle. *Circulation*. 2008;118(9):947-954. doi:10.1161/CIRCULATIONAHA.108.781062

2. Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary prevention of coronary heart disease in women through diet and lifestyle. *N Engl J Med*. 2000;343(1):16-22. doi:10.1056/NEJM200007063430103

3. Song M, Giovannucci E. Preventable Incidence and Mortality of Carcinoma Associated With Lifestyle Factors Among White Adults in the United States. *JAMA Oncology*. 2016;2(9):1154-1161. doi:10.1001/jamaoncol.2016.0843

4. Ainsworth BE, Haskell WL, Leon AS, et al. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc*. 1993;25(1):71-80. doi:10.1249/00005768-199301000-00011

5. Chasan-Taber S, Rimm EB, Stampfer MJ, et al. Reproducibility and Validity of a Self-Administered Physical Activity Questionnaire for Male Health Professionals. *Epidemiology*. 1996;7(1):81.

6. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and Validity of a Self-Administered Physical Activity Questionnaire. *Int J Epidemiol*. 1994;23(5):991-999. doi:10.1093/ije/23.5.991

7. Lee DH, Rezende LFM, Ferrari G, et al. Physical activity and all-cause and cause-specific mortality: assessing the impact of reverse causation and measurement error in two large prospective cohorts. *Eur J Epidemiol*. 2021;36(3):275-285. doi:10.1007/s10654-020-00707-3

8. McCullough ML, Feskanich D, Stampfer MJ, et al. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *Am J Clin Nutr*. 2002;76(6):1261-1271. doi:10.1093/ajcn/76.6.1261

9. Chiuve SE, Fung TT, Rimm EB, et al. Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease. *The Journal of Nutrition*. 2012;142(6):1009-1018. doi:10.3945/jn.111.157222

10. Nutrition C for FS and A. Final Determination Regarding Partially Hydrogenated Oils (Removing Trans Fat). *FDA*. Published online February 20, 2020. Accessed April 3, 2023. https://www.fda.gov/food/food-additives-petitions/final-determination-regarding-partially-hydrogenated-oils-removing-trans-fat

11. Yue Y, Yuan C, Wang DD, et al. Reproducibility and validity of diet quality scores derived from food-frequency questionnaires. *The American Journal of Clinical Nutrition*. 2022;115(3):843-853. doi:10.1093/ajcn/nqab368

12. American Diabetes Association. Standards of Medical Care in Diabetes—2010. *Diabetes Care*. 2010;33(Supplement_1):S11-S61. doi:10.2337/dc10-S011

13. National Diabetes Data Group. Classification and Diagnosis of Diabetes Mellitus and Other Categories of Glucose Intolerance. *Diabetes*. 1979;28(12):1039-1057. doi:10.2337/diab.28.12.1039

14. Walker AE, Robins M, Weinfeld FD. The National Survey of Stroke. Clinical findings. *Stroke*. 1981;12(2 Pt 2 Suppl 1):I13-44.

15. Mendis S, Thygesen K, Kuulasmaa K, et al. World Health Organization definition of myocardial infarction: 2008-09 revision. *Int J Epidemiol*. 2011;40(1):139-146. doi:10.1093/ije/dyq165

16. Bullard KM. Prevalence of Diagnosed Diabetes in Adults by Diabetes Type — United States, 2016. *MMWR Morb Mortal Wkly Rep*. 2018;67. doi:10.15585/mmwr.mm6712a2

17. Surveillance - United States Diabetes Surveillance System. Accessed January 7, 2024. https://gis.cdc.gov/grasp/diabetes/diabetesatlas-surveillance.html

18. Rosamond WD, Chambless LE, Heiss G, et al. Twenty-Two–Year Trends in Incidence of Myocardial Infarction, Coronary Heart Disease Mortality, and Case Fatality in 4 US

Communities, 1987–2008. *Circulation*. 2012;125(15):1848-1857. doi:10.1161/CIRCULATIONAHA.111.047480

19. Madsen TE, Khoury JC, Leppert M, et al. Temporal Trends in Stroke Incidence Over Time by Sex and Age in the GCNKSS. *Stroke*. 2020;51(4):1070-1076. doi:10.1161/STROKEAHA.120.028910

20. SEER*Explorer Application. Accessed January 7, 2024. https://seer.cancer.gov/statistics-network/explorer/application.html?site=1&data_type=1&graph_type=2&compareBy=sex&chk_sex_3=3&rate_type=1&race=1&age_range=166&hdn_stage=101&advopt_precision=1&advopt_show_ci=on&hdn_view=0&advopt_show_apc=on&advopt_display=1#resultsRegion0

21. Ferguson J, Alvarez A, Mulligan M, Judge C, O'Donnell M. Bias assessment and correction for Levin's population attributable fraction in the presence of confounding. *Eur J Epidemiol*. Published online January 3, 2024. doi:10.1007/s10654-023-01063-8

22. Miettinen OS. COMPONENTS OF THE CRUDE RISK RATIO. *American Journal of Epidemiology*. 1972;96(2):168-172. doi:10.1093/oxfordjournals.aje.a121443

23. Ng R, Sutradhar R, Yao Z, Wodchis WP, Rosella LC. Smoking, drinking, diet and physical activity—modifiable lifestyle risk factors and their associations with age to first chronic disease. *Int J Epidemiol*. 2020;49(1):113-130. doi:10.1093/ije/dyz078

24. Steinbrecher A, Morimoto Y, Heak S, et al. The preventable proportion of type 2 diabetes by ethnicity: The Multiethnic Cohort. *Ann Epidemiol*. 2011;21(7):526-535. doi:10.1016/j.annepidem.2011.03.009

25. Chomistek AK, Chiuve SE, Eliassen AH, Mukamal KJ, Willett WC, Rimm EB. Healthy Lifestyle in the Primordial Prevention of Cardiovascular Disease Among Young Women. *J Am Coll Cardiol*. 2015;65(1):43-51. doi:10.1016/j.jacc.2014.10.024

26. Zhang FF, Cudhea F, Shan Z, et al. Preventable Cancer Burden Associated With Poor Diet in the United States. *JNCI Cancer Spectr*. 2019;3(2):pkz034. doi:10.1093/jncics/pkz034

APPENDIX ONE

Appendix Table 1.1. Association between genetically predicted or measured BMI and risk of colorectal polyp subtypes in three cohort studies (NHS, NHS2, HPFS).

	Men and Women (NHS, NHS2, HPFS)		Men (HPFS)		Women (NHS, NHS2)	
	Genetically predicted BMI, per 2 kg/m^2 increase [a,b]	Measured BMI, per 2 kg/m^2 increase [c]	Genetically predicted BMI, per 2 kg/m^2 increase	Measured BMI, per 2 kg/m^2 increase	Genetically predicted BMI, per 2 kg/m^2 increase	Measured BMI, per 2 kg/m^2 increase
Conventional Adenoma						
n	2,952	2,952	1,274	1, 274	1,678	1, 678
Mean (GRS or BMI)	87.8	26.3	87.5	26.2	87.9	26.4
OR (95% CI)	0.98 (0.88-1.10)	1.03 (1.01-1.04)	0.93 (0.79-1.10)	1.02 (0.98-1.06)	1.03 (0.89-1.19)	1.03 (1.01-1.05)
Non-Advanced						
n	1,628	1,628	694	694	934	934
Mean (GRS or BMI)	87.7	26.3	87.4	26.1	88.0	26.5
OR (95% CI)	0.99 (0.85-1.14)	1.03 (1.01-1.05)	0.88 (0.70-1.09)	1.01 (0.97-1.06)	1.07 (0.89-1.30)	1.04 (1.01-1.06)
Advanced						
n	1,324	1,324	580	580	744	744
Mean (GRS or BMI)	87.8	26.3	87.7	26.2	87.8	26.3
OR (95% CI)	0.98 (0.83-1.16)	1.02 (0.99-1.04)	1.01 (0.79-1.28)	1.03 (0.98-1.08)	0.98 (0.78-1.22)	1.02 (0.99-1.05)
Serrated Polyp						
n	1,589	1,589	432	432	1,157	1,157
Mean (GRS or BMI)	87.9	26.7	87.8	26.5	87.9	26.7
OR (95% CI)	1.04 (0.90-1.20)	1.06 (1.04-1.08)	1.11 (0.83-1.47)	1.08 (1.02-1.14)	0.99 (0.84-1.18)	1.05 (1.03-1.07)
Small (<10 mm)						
n	1,370	1,370	362	362	1,008	1,008
Mean (GRS or BMI)	87.9	26.7	87.8	26.7	87.9	26.8
OR (95% CI)	1.03 (0.88-1.21)	1.06 (1.04-1.08)	1.11 (0.81-1.52)	1.10 (1.04-1.16)	0.97 (0.81-1.17)	1.05 (1.03-1.08)
Large (>10 mm)						
n	147	147	41	41	106	106
Mean (GRS or BMI)	87.5	26.5	88.0	25.8	87.4	26.7
OR (95% CI)	0.87 (0.57-1.32)	1.03 (0.96-1.11)	1.31 (0.63-2.71)	0.95 (0.79-1.15)	0.80 (0.48-1.33)	1.03 (0.95-1.11)

Appendix Table 1.1 (Continued) Association between genetically predicted or measured BMI and risk of colorectal polyp subtypes in three cohort studies (NHS, NHS2, HPFS).

Synchronous Conventional Adenoma & Serrated Polyp						
n	790	790	336	336	454	454
Mean (GRS or BMI)	88.0	27.1	87.3	26.7	88.5	27.5
OR (95% CI)	1.09 (0.89-1.34)	1.10 (1.07-1.13)	0.88 (0.65-1.20)	1.09 (1.03-1.16)	1.37 (1.05-1.79)	1.11 (1.07-1.14)

BMI: body mass index; GRS: genetic risk score

[a] An 18-unit change in GRS is equivalent to a 2 kg/m^2 change in measured BMI, per regression analysis

[b] Genetically predicted BMI multivariable logistic regression model adjusted for age, study cohort (NHS, NHS2, HPFS), time period of endoscopy (in 2-year intervals), reason for endoscopy (screening or symptoms), number of previous endoscopies, time in years since most recent endoscopy, and top three principal components for population structure. All covariates were treated as continuous variables unless otherwise noted.

[c] Measured BMI multivariable logistic regression model adjusted for age, study cohort (NHS, NHS2, HPFS), time period of endoscopy (in 2-year intervals), reason for endoscopy (screening or symptoms), number of previous endoscopies, time in years since most recent endoscopy, family history of colorectal cancer (yes or no), height, alcohol intake, regular aspirin use (yes or no), and physical activity. All covariates were treated as continuous variables unless otherwise noted.

Appendix Table 1.2. Association between genetically determined or measured BMI and risk of colorectal polyp subtypes, without excluding synchronous polyps, in three cohort studies (NHS, NHS2, HPFS)

	Men and Women (NHS, NHS2, HPFS)		Men (HPFS)		Women (NHS, NHS2)	
	Genetically predicted BMI, per 2 kg/m^2 increase [a,b]	**Measured BMI, per 2 kg/m^2 increase** [c]	**Genetically predicted BMI, per 2 kg/m^2 increase**	**Measured BMI, per 2 kg/m^2 increase**	**Genetically predicted BMI, per 2 kg/m^2 increase**	**Measured BMI, per 2 kg/m^2 increase**
Conventional Adenoma						
n	3,742	3,742	1,610	1,610	2,132	2,132
Mean (GRS or BMI)	87.8	26.5	87.5	26.3	88.1	26.6
OR (95% CI)	1.01 (0.91-1.11)	1.04 (1.03-1.06)	0.92 (0.79-1.07)	1.03 (1.00-1.07)	1.09 (0.96-1.24)	1.05 (1.03-1.06)
Non-Advanced						
n	2,083	2,083	873	873	1,210	1,210
Mean (GRS or BMI)	87.8	26.5	87.4	26.2	88.2	26.8
OR (95% CI)	1.03 (0.91-1.18)	1.05 (1.03-1.07)	0.90 (0.74-1.10)	1.03 (0.99-1.08)	1.14 (0.96-1.35)	1.05 (1.03-1.08)
Advanced						
n	1,659	1,659	737	737	922	922
Mean (GRS or BMI)	87.7	26.4	87.5	26.3	87.9	26.5
OR (95% CI)	0.97 (0.84-1.12)	1.03 (1.01-1.05)	0.95 (0.77-1.17)	1.04 (0.99-1.08)	1.04 (0.85-1.26)	1.04 (1.01-1.06)
Serrated Polyp						
n	2,379	2,379	768	768	1,611	1,611
Mean (GRS or BMI)	87.9	26.8	87.6	26.6	88.1	26.9
OR (95% CI)	1.06 (0.94-1.19)	1.07 (1.05-1.09)	1.00 (0.81-1.24)	1.08 (1.04-1.13)	1.09 (0.94-1.26)	1.07 (1.05-1.09)
Small (<10 mm)						
n	2,007	2,007	623	623	1,384	1,384
Mean (GRS or BMI)	87.9	26.9	87.6	26.7	88.0	26.9
OR (95% CI)	1.05 (0.92-1.20)	1.07 (1.05-1.09)	1.02 (0.80-1.30)	1.10 (1.05-1.15)	1.05 (0.90-1.23)	1.07 (1.04-1.09)

Appendix Table 1.2 (Continued) Association between genetically determined or measured BMI and risk of colorectal polyp subtypes, without excluding synchronous polyps, in three cohort studies (NHS, NHS2, HPFS)

Large (>10 mm)						
n	227	227	82	82	145	145
Mean (GRS or BMI)	87.5	26.7	87.7	26.4	87.4	26.9
OR (95% CI)	0.87 (0.61-1.23)	1.05 (0.99-1.11)	1.09 (0.63-1.89)	1.04 (0.90-1.20)	0.84 (0.54-1.32)	1.05 (0.99-1.12)

BMI: body mass index; GRS: genetic risk score

[a] An 18-unit change in GRS is equivalent to a 2 kg/m^2 change in measured BMI, per regression analysis

[b] Genetically predicted BMI multivariable logistic regression model adjusted for age, study cohort (NHS, NHS2, HPFS), time period of endoscopy (in 2-year intervals), reason for endoscopy (screening or symptoms), number of previous endoscopies, time in years since most recent endoscopy, and top three principal components for population structure. All covariates were treated as continuous variables unless otherwise noted.

[c] Measured BMI multivariable logistic regression model adjusted for age, study cohort (NHS, NHS2, HPFS), time period of endoscopy (in 2-year intervals), reason for endoscopy (screening or symptoms), number of previous endoscopies, time in years since most recent endoscopy, family history of colorectal cancer (yes or no), height, alcohol intake, regular aspirin use (yes or no), and physical activity. All covariates were treated as continuous variables unless otherwise noted.

APPENDIX TWO

Appendix Methods 2.1

Metabolite processing

High resolution, accurate mass profiling metabolomic data were acquired using LC-MS systems comprised of Nexera X2 U-HPLC systems (Shimadzu Corp.; Marlborough, MA) coupled to Q Exactive and Exactive Plus orbitrap mass spectrometers (Thermo Fisher Scientific; Waltham, MA). Hydrophilic interaction liquid chromatography with positive ion mode MS detection (HILIC-pos) was used to separate polar metabolites, while C18 chromatography with negative ion mode detection (C18-neg) and C8 chromatography with positive ion mode detection (C8-pos) were used to profile metabolites of intermediate polarity and lipids, respectively. Raw data were processed using TraceFinder 3.3 software (Thermo Fisher Scientific; Waltham, MA) and Progenesis QI (Nonlinear Dynamics; Newcastle upon Tyne, UK). Metabolite identities were confirmed using authentic reference standards or reference samples; unnamed metabolites and metabolites with no variance were excluded.

Three pilot studies were conducted in 2013 to assess intraassay reproducibility, reproducibility over delays in processing, and within-person reproducibility over time among the archived samples.[1] Approximately 90% of metabolites had CVs <20%, and approximately 75% of metabolites had Spearman correlation coefficients or ICCs ≥ 0.75 when comparing immediately processed samples vs. samples processed after a 24-hour delay. A subset of metabolites; consisting primarily of purines, pyrimidines, and carbohydrates; demonstrated poor stability in 24 hour delayed processing and were excluded.[1] Because the NHS samples used sodium heparin and HPFS samples used EDTA anticoagulants, there were some differences in the excluded failed metabolites in each cohort. There were a total of 362 metabolites assayed among individuals in the CRC study after excluding unnamed metabolites and metabolites with no variance, of which, 66 failed the stability pilot in either the NHS (42 metabolites) or HPFS (33 metabolites). We performed imputation for metabolites with <10% missingness using half of the minimum value measured for that metabolite, under the assumption that missingness was

attributed to a metabolite concentration below the detectable limit; metabolites with >10% missing or that demonstrated poor staibility in pilot studies were excluded.[2] Quality control samples were randomly distributed among the participants' samples and were also profiled, and matched cases and controls were assessed in the same batch.

Inflammation and metabolic dysregulation biomarker processing

We used a highly sensitive immunoturbidimetric assay (Denka Seiken Co.) to measure CRP and enzyme-linked immunosorbent assays (ELISA) to measure IL-6, TNFRSF1B, and GDF15 (R & D Systems);[3,4] C-peptide (Diagnostic Systems Laboratories/Beckman Coulter, Webster, TX); and adiponectin (ALPCO Diagnostics, Salem, New Hampshire). Quality control samples were randomly interspersed among the case-control samples, and personnel blinded to quality control and case-control status conducted all assays. Outliers were removed using the studentized deviate many-outlier procedure with a maximum of 10 outliers.[5]

Assessment of confounding variables

Data on alcohol consumption were categorized as 0 g/day, 0.1-5 g/day, 5-10 g/day, 10-15 g/day, or >15 g/day to account for potential non-linear associations of alcohol consumption with CRC. Postmenopausal hormone use was categorized as never, past, or current; and regular aspirin use was defined as ≥2 tablets/week. Confounders were derived as the cumulative average of all values collected via biennial questionnaires collected prior to blood collection; on average, the most recent questionnaire used to estimate covariates was returned 15 months (NHS) or 12 months (HPFS) before the date of blood collection. After imputation of missing data by last observation carried forward, there was < 5% missingness for all confounders; confounder data that remained missing after last observation carried forward were imputed with the median cohort value.

Comparison of ENR/RRR to alternative methods for signature development

We derived the signatures using four alternative models to compare to the combined ENR and RRR method used in the main analysis: 1) Combined principal components analysis (PCA) and ENR, 2) sparse RRR, 3) RRR alone, and 4) Combined RRR and stepwise linear regression. For the PCA/ENR models, we replaced the RRR step in the primary model with PCA (applied PCA to the 4 markers of inflammation or metabolic dysregulation to derive an inflammatorion or metabolic dysregulation score that was then used as the response variable in ENR). For the sparse RRR model, we used the "rrpack" package in R, with 5-fold cross validation to select the corresponding parameters that minimize mean squared error, and adjusted for the same confounders (age at blood collection, date of blood collection, fasting status, and sex) as the primary ENR model. For the RRR alone model, we used only RRR and retained all 277 metabolites in the final signature with no variable selection. In the RRR/stepwise model, we replaced the ENR step in the primary models with stepwise regression with $p = 0.05$ for forward and backward selection. All four models were derived using the same 10-fold cross validation approach used in the main analysis to produce 10 unique versions of the signatures, one for each hold-out set.

CRC molecular subtypes

We evaluated the association of the signatures with different subtypes of CRC defined based on the combined presence or absence of microsatellite instability (MSI), CpG island methylator phenotype (CIMP), and *BRAF* and *KRAS* mutations. The subtypes were defined as follows: conventional pathway as non-MSI-high, CIMP-low or negative, *BRAF*-wild-type, *KRAS*-wild-type; serrated pathway as any MSI status, CIMP-high, *BRAF*-mutated, *KRAS*-wild-type; alternative pathway as non-MSI-high, CIMP-low or negative, *BRAF*-wild-type, *KRAS*-mutated; and other as any other combination of markers.[9] *KRAS*-wild-type vs. *KRAS*-mutated CRC was

also compared. Inverse probability weights were estimated using logistic regression models with the availability of molecular data (available vs. missing) as the outcome and covariates that could influence the success of tissue collection as predictor variables: disease stage (stage I, stage II, stage III, or stage IV), tumor differentiation (well differentiated, moderately differentiated, or poorly differentiated), tumor subsite (proximal colon, distal colon, or rectum), age at diagnosis (continuous), age at diagnosis-square (continuous), and diagnostic year (continuous).[10]

Appendix 2 References

1. Townsend MK, Clish CB, Kraft P, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin Chem*. 2013;59(11):1657-1667. doi:10.1373/clinchem.2012.199133

2. Wei R, Wang J, Su M, et al. Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Sci Rep*. 2018;8:663. doi:10.1038/s41598-017-19120-0

3. Heidemann C, Sun Q, van Dam RM, et al. Total and High-Molecular Weight Adiponectin and Resistin in Relation to the Risk for Type 2 Diabetes in Women: A Prospective Study. *Ann Intern Med*. 2008;149(5):307-316.

4. Song M, Mehta RS, Wu K, et al. Plasma Inflammatory Markers and Risk of Advanced Colorectal Adenoma in Women. *Cancer Prev Res (Phila)*. 2016;9(1):27-34. doi:10.1158/1940-6207.CAPR-15-0307

5. Percentage Points for a Generalized ESD Many-Outlier Procedure: Technometrics: Vol 25, No 2. Accessed June 13, 2023. https://www-tandfonline-com.ezp-prod1.hul.harvard.edu/doi/abs/10.1080/00401706.1983.10487848

6. Izenman AJ. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*. 1975;5(2):248-264. doi:10.1016/0047-259X(75)90042-1

7. Hilafu H, Safo SE, Haine L. Sparse reduced-rank regression for integrating omics data. *BMC Bioinformatics*. 2020;21(1):283. doi:10.1186/s12859-020-03606-2

8. Zou H, Hastie T. Regression shrinkage and selection via the elastic net, with applications to microarrays. Published online January 5, 2004.

9. Wang L, He X, Ugai T, et al. Risk Factors and Incidence of Colorectal Cancer According to Major Molecular Subtypes. *JNCI Cancer Spectrum*. 2021;5(1). doi:10.1093/jncics/pkaa089

10. Liu L, Nevo D, Nishihara R, et al. Utility of inverse probability weighting in molecular pathological epidemiology. *Eur J Epidemiol*. 2018;33(4):381-392. doi:10.1007/s10654-017-0346-8

Appendix Table 2.1. List of all metabolites with non-zero coefficient in all 10 iterations of 10-fold cross validation elastic net regression for the signature of metabolic dysregulation, signature of inflammation, or both

Superclass	Subclass	Metabolite name	HMDB ID	Direction of coefficient across iterations, inflammation	Direction of coefficient across iterations, metabolic dysregulation
Alkaloids and derivatives	Alkaloids and derivatives	Trigonelline	HMDB0000875	+	
Benzenoids	Benzenoids	Hippurate	HMDB0000714	+	
Lipids and lipid-like molecules	Acyl carnitines	C12 carnitine	HMDB0002250		-
Lipids and lipid-like molecules	Acyl carnitines	C14 carnitine	HMDB0005066	+	
Lipids and lipid-like molecules	Acyl carnitines	C2 carnitine	HMDB0000201	-	-
Lipids and lipid-like molecules	Acyl carnitines	C26 carnitine	HMDB0006347	+	+
Lipids and lipid-like molecules	Acyl carnitines	C3 carnitine	HMDB0000824	+	+
Lipids and lipid-like molecules	Acyl carnitines	C4 carnitine	HMDB0002013	+	
Lipids and lipid-like molecules	Acyl carnitines	C5 carnitine	HMDB0000688		+
Lipids and lipid-like molecules	Acyl carnitines	C5:1 carnitine	HMDB0002366	-	
Lipids and lipid-like molecules	Acyl carnitines	C6 carnitine	HMDB0000705	+	+
Lipids and lipid-like molecules	Carnitines	Carnitine	HMDB0000062	-	
Lipids and lipid-like molecules	Ceramides	C16:0 Ceramide (d18:1)	HMDB0004949	+	+
Lipids and lipid-like molecules	Ceramides	C22:0 Ceramide (d18:1)	HMDB0004952	-	
Lipids and lipid-like molecules	Ceramides	C24:0 Ceramide (d18:1)	HMDB0004956	-	
Lipids and lipid-like molecules	Ceramides	C24:1 Ceramide (d18:1)	HMDB0004953*	-	+
Lipids and lipid-like molecules	Cholesteryl esters	C16:1 CE	HMDB0000658*		-
Lipids and lipid-like molecules	Cholesteryl esters	C18:2 CE	HMDB0000610*	-	
Lipids and lipid-like molecules	Cholesteryl esters	C20:3 CE	HMDB0006736*	+	
Lipids and lipid-like molecules	Cholesteryl esters	C22:5 CE	HMDB0010375*		-
Lipids and lipid-like molecules	Diacylglycerols	C34:1 DAG	HMDB0007102*	+	
Lipids and lipid-like molecules	Diacylglycerols	C36:1 DAG	HMDB0007216*	+	
Lipids and lipid-like molecules	Fatty acids	Butyrobetaine	HMDB0006831	+	-
Lipids and lipid-like molecules	Fatty acids	Myristoleic acid	HMDB0002000	+	+
Lipids and lipid-like molecules	Glycinated bile acids and derivatives	Glycocholate	HMDB0000138	+	+
Lipids and lipid-like molecules	Glycinated bile acids and derivatives	Glycodeoxycholate	HMDB0000631*	+	
Lipids and lipid-like molecules	Lineolic acids and derivatives	C36:4 DAG	HMDB0007248*		+
Lipids and lipid-like molecules	Lysophosphatidylcholines	C22:5 LPC	HMDB0010403*		-
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C18:3 LPE	HMDB0011478*		-
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C20:1 LPE	HMDB0011512*		-
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C22:0 LPE	HMDB0011520		-
Lipids and lipid-like molecules	Phosphatidylcholines	C34:1 PC	HMDB0007972*		-
Lipids and lipid-like molecules	Phosphatidylcholines	C34:1 PC plasmalogen	HMDB0011208*		-
Lipids and lipid-like molecules	Phosphatidylcholines	C34:2 PC plasmalogen	HMDB0011210*		-
Lipids and lipid-like molecules	Phosphatidylcholines	C36:3 PC plasmalogen	HMDB0011244*		-
Lipids and lipid-like molecules	Phosphatidylcholines	C36:4 PC-A	HMDB0007983*		-
Lipids and lipid-like molecules	Phosphatidylcholines	C40:10 PC	HMDB0008511*		-
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:2 PE	HMDB0008942*		-
Lipids and lipid-like molecules	Phosphocholines	Phosphocholine	HMDB0001565	-	-
Lipids and lipid-like molecules	Phosphosphingolipids	C18:0 SM	HMDB0001348		+
Lipids and lipid-like molecules	Steroids and derivatives	21-deoxycortisol	HMDB0004030	-	-
Lipids and lipid-like molecules	Steroids and derivatives	Campesterol	HMDB0002869	-	
Lipids and lipid-like molecules	Steroids and derivatives	Cholesterol	HMDB0000067	-	+
Lipids and lipid-like molecules	Steroids and derivatives	Cortisol	HMDB0000063	+	-

Appendix Table 2.1 (Continued) List of all metabolites with non-zero coefficient in all 10 iterations of 10-fold cross validation elastic net regression for the signature of metabolic dysregulation, signature of inflammation, or both

Lipids and lipid-like molecules	Steroids and derivatives	Cortisone	HMDB0002802	-	-
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C48:0 TAG	HMDB0005356*	-	
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C48:2 TAG	HMDB0005376*	+	
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C50:1 TAG	HMDB0005360*		+
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C51:2 TAG	HMDB0005362*	-	
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C52:1 TAG	HMDB0005367*		+
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C52:2 TAG	HMDB0005369*	+	+
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C54:2 TAG	HMDB0005403*		+
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C56:1 TAG	HMDB0005396*	+	
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C50:3 TAG	HMDB0005433*		+
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C52:3 TAG	HMDB0005384*	+	
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C52:4 TAG	HMDB0005363*		+
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C52:6 TAG	HMDB0005436*	-	
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:3 TAG	HMDB0005405*	+	
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:4 TAG	HMDB0005370*	-	
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:7 TAG	HMDB0005447*		+
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:4 TAG	HMDB0005398*	-	-
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:5 TAG	HMDB0005406*	+	
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:6 TAG	HMDB0005456*	+	
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C58:6 TAG	HMDB0005458*		-
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C58:7 TAG	HMDB0005471*	-	
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C58:8 TAG	HMDB0005413*	-	-
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C58:9 TAG	HMDB0005463*	-	
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C60:12 TAG	HMDB0005478*	-	
Lipids and lipid-like molecules	Ubiquinones	Coenzyme Q10	HMDB0001072		+
Organic acids and derivatives	Alpha amino acid derivatives	Creatine	HMDB0000064	-	+
Organic acids and derivatives	Alpha amino acid derivatives	Creatinine	HMDB0000562	+	
Organic acids and derivatives	Alpha amino acid derivatives	Guanidinoacetic acid	HMDB0000128	-	
Organic acids and derivatives	Alpha amino acid derivatives	Thyroxine	HMDB0000248		-
Organic acids and derivatives	Alpha amino acids (BCAA)	Isoleucine	HMDB0000172	+	
Organic acids and derivatives	Alpha amino acids (BCAA)	Leucine	HMDB0000687	-	
Organic acids and derivatives	Alpha amino acids (BCAA)	Valine	HMDB0000883	-	+
Organic acids and derivatives	Alpha amino acids (non-BCAA)	2-aminooctanoic acid	HMDB0000991*	+	-
Organic acids and derivatives	Alpha amino acids (non-BCAA)	3-methylhistidine	HMDB0000479		-
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Alanine	HMDB0000161	-	+
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Aminoisobutyric acid	HMDB0001906	-	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Asparagine	HMDB0000168		-
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Betaine	HMDB0000043	+	-
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Dimethylglycine	HMDB0000092	+	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Glutamine	HMDB0000641	-	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Glycine	HMDB0000123	+	-
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Histidine	HMDB0000177	-	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Homoarginine	HMDB0000670	+	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Homocitrulline	HMDB0000679	-	-
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Hydroxyproline	HMDB0000725	+	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Methionine	HMDB0000696	+	-
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Methionine sulfoxide	HMDB0002005	-	+
Organic acids and derivatives	Alpha amino acids (non-BCAA)	N6,N6,N6-trimethyllysine	HMDB0001325	+	+

Appendix Table 2.1 (Continued) List of all metabolites with non-zero coefficient in all 10 iterations of 10-fold cross validation elastic net regression for the signature of metabolic dysregulation, signature of inflammation, or both

Organic acids and derivatives	Alpha amino acids (non-BCAA)	Phenylalanine	HMDB0000159	+	+
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Pipecolic acid	HMDB0000716	+	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Proline	HMDB0000162	+	+
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Proline betaine	HMDB0004827	+	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	SDMA	HMDB0003334		-
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Serine	HMDB0000187	-	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Threonine	HMDB0000167	+	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Tryptophan	HMDB0000929	-	
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Tyrosine	HMDB0000158		+
Organic acids and derivatives	Carboximidic acids	N1-acetylspermidine	HMDB0001276	+	
Organic acids and derivatives	Carboximidic acids	Pantothenate	HMDB0000210	-	-
Organic acids and derivatives	N-acyl-alpha amino acids	N-Acetyl-L-alanine	HMDB0000766		+
Organic acids and derivatives	N-acyl-alpha amino acids	N-acetylornithine	HMDB0003357	-	-
Organic acids and derivatives	N-acyl-alpha-hexosamines	Acetyl-galactosamine	HMDB0000212	+	
Organic acids and derivatives	Ureas	N-carbamoyl-beta-alanine	HMDB0000026	+	
Organic acids and derivatives	Ureas	Uric acid	HMDB0000289		+
Organic nitrogen compounds	2-arylethylamines	1-methylhistamine	HMDB0000898	+	-
Organic nitrogen compounds	Amine oxides	Trimethylamine-N-oxide	HMDB0000925	+	
Organic nitrogen compounds	N-acetylarylamines	5-acetylamino-6-amino-3-methyluracil	HMDB0004400	+	
Organoheterocyclic compounds	Bilirubins	Bilirubin	HMDB0000054	-	
Organoheterocyclic compounds	Bilirubins	Biliverdin	HMDB0001008	-	-
Organoheterocyclic compounds	Imidazoles	Allantoin	HMDB0000462	+	
Organoheterocyclic compounds	Imidazoles	Urocanic acid	HMDB0000301	-	
Organoheterocyclic compounds	Nicotinamides	1-methylnicotinamide	HMDB0000699	-	
Organoheterocyclic compounds	Nicotinamides	N1-methyl-2-pyridone-5-carboxamide	HMDB0004193	+	
Organoheterocyclic compounds	Purines and derivatives	1-methylguanosine	HMDB0001563	+	+
Organoheterocyclic compounds	Purines and derivatives	2-methylguanosine	HMDB0005862	+	+
Organoheterocyclic compounds	Purines and derivatives	7-methylguanine	HMDB0000897	+	
Organoheterocyclic compounds	Purines and derivatives	N2,N2-dimethylguanosine	HMDB0004824	+	+
Organoheterocyclic compounds	Pyridines and derivatives	Pyridoxamine	HMDB0001431	+	
Organoheterocyclic compounds	Pyrimidines and derivatives	5-hydroxymethyl-4-methyluracil	HMDB0000544		+
Organoheterocyclic compounds	Pyrimidines and derivatives	Cytosine	HMDB0000630	-	-
Organoheterocyclic compounds	Pyrimidines and derivatives	N4-acetylcytidine	HMDB0005923	-	+
Organoheterocyclic compounds	Pyrimidines and derivatives	Pseudouridine	HMDB0000767	+	
Organoheterocyclic compounds	Pyrimidines and derivatives	Ribothymidine	HMDB0000884	-	-
Organoheterocyclic compounds	Pyrimidines and derivatives	Thiamine	HMDB0000235	-	
Organoheterocyclic compounds	Xanthines	3-methylxanthine	HMDB0001886	-	
Organoheterocyclic compounds	Xanthines	7-methylxanthine	HMDB0001991	+	
Xenobiotics	Xenobiotics	Acetaminophen	HMDB0001859	+	
Xenobiotics	Xenobiotics	Gabapentin	HMDB0005015	-	-
Xenobiotics	Xenobiotics	Metformin	HMDB0001921	-	

* representative ID

Appendix Table 2.2. List of all available metabolites in women (NHS) and men (HPFS) by subclass with parameter estimates for each signature derived from elastic net regression.

Superclass	Subclass	Metabolite name	HMDB ID	Coefficient, metabolomic signature of inflammation	Coefficient, metabolomic signature of metabolic dysregulation
Alkaloids and derivatives	Alkaloids and derivatives	Piperine	HMDB0029377	-0.055	0.029
Alkaloids and derivatives	Alkaloids and derivatives	Trigonelline	HMDB0000875	0.034	0.000
Benzenoids	Benzenoids	2-methyl-4,5-benzoxazole	HMDB0032390	-0.045	0.000
Benzenoids	Benzenoids	4-hydroxyhippurate	HMDB0013678	0.018	0.000
Benzenoids	Benzenoids	Hippurate	HMDB0000714	0.017	0.000
Benzenoids	Benzenoids	Methyl N-methylanthranilate	HMDB0034169	0.048	0.026
Benzenoids	Benzenoids	Trimethylbenzene	HMDB0013733*	-0.023	0.000
Lipids and lipid-like molecules	Acyl carnitines	C10 carnitine	HMDB0000651	0.000	0.000
Lipids and lipid-like molecules	Acyl carnitines	C10:2 carnitine	HMDB0013325	0.013	0.000
Lipids and lipid-like molecules	Acyl carnitines	C12 carnitine	HMDB0002250	0.000	-0.066
Lipids and lipid-like molecules	Acyl carnitines	C12:1 carnitine	HMDB0013326	0.009	0.000
Lipids and lipid-like molecules	Acyl carnitines	C14 carnitine	HMDB0005066	0.061	0.000
Lipids and lipid-like molecules	Acyl carnitines	C14:1 carnitine	HMDB0002014	0.000	0.000
Lipids and lipid-like molecules	Acyl carnitines	C14:2 carnitine	HMDB0013331	0.000	0.000
Lipids and lipid-like molecules	Acyl carnitines	C2 carnitine	HMDB0000201	-0.099	-0.017
Lipids and lipid-like molecules	Acyl carnitines	C26 carnitine	HMDB0006347	0.000	0.000
Lipids and lipid-like molecules	Acyl carnitines	C3 carnitine	HMDB0000824	0.029	0.000
Lipids and lipid-like molecules	Acyl carnitines	C4 carnitine	HMDB0002013	0.000	0.000
Lipids and lipid-like molecules	Acyl carnitines	C4-OH carnitine	HMDB0013127	0.062	0.000
Lipids and lipid-like molecules	Acyl carnitines	C5 carnitine	HMDB0000688	0.000	0.000
Lipids and lipid-like molecules	Acyl carnitines	C5-DC carnitine	HMDB0013130	0.000	0.000
Lipids and lipid-like molecules	Acyl carnitines	C5:1 carnitine	HMDB0002366	0.000	0.000
Lipids and lipid-like molecules	Acyl carnitines	C6 carnitine	HMDB0000705	0.016	0.012
Lipids and lipid-like molecules	Acyl carnitines	C7 carnitine	HMDB0013238	-0.001	0.000
Lipids and lipid-like molecules	Acyl carnitines	C8 carnitine	HMDB0000791	0.000	0.000
Lipids and lipid-like molecules	Acyl carnitines	C9 carnitine	HMDB0013288	0.005	0.000
Lipids and lipid-like molecules	Carnitines	Carnitine	HMDB0000062	-0.123	0.000
Lipids and lipid-like molecules	Ceramides	C16:0 Ceramide (d18:1)	HMDB0004949	0.073	0.024
Lipids and lipid-like molecules	Ceramides	C22:0 Ceramide (d18:1)	HMDB0004952	-0.007	0.000
Lipids and lipid-like molecules	Ceramides	C24:0 Ceramide (d18:1)	HMDB0004956	-0.036	0.000
Lipids and lipid-like molecules	Ceramides	C24:1 Ceramide (d18:1)	HMDB0004953*	0.000	0.015
Lipids and lipid-like molecules	Cholesteryl esters	C16:0 CE	HMDB0000885	0.000	0.000
Lipids and lipid-like molecules	Cholesteryl esters	C16:1 CE	HMDB0000658*	0.000	-0.030
Lipids and lipid-like molecules	Cholesteryl esters	C18:0 CE	HMDB0010368	0.000	0.000
Lipids and lipid-like molecules	Cholesteryl esters	C18:2 CE	HMDB0000610*	-0.068	-0.018
Lipids and lipid-like molecules	Cholesteryl esters	C20:3 CE	HMDB0006736*	0.021	0.000
Lipids and lipid-like molecules	Cholesteryl esters	C20:4 CE	HMDB0006726	-0.006	0.000
Lipids and lipid-like molecules	Cholesteryl esters	C20:5 CE	HMDB0006731	0.000	0.000
Lipids and lipid-like molecules	Cholesteryl esters	C22:5 CE	HMDB0010375*	-0.073	-0.004
Lipids and lipid-like molecules	Cholesteryl esters	C22:6 CE	HMDB0006733	-0.012	0.000
Lipids and lipid-like molecules	Diacylglycerols	C32:0 DAG	HMDB0007098*	0.000	0.000
Lipids and lipid-like molecules	Diacylglycerols	C32:1 DAG	HMDB0007099*	0.000	0.000
Lipids and lipid-like molecules	Diacylglycerols	C34:0 DAG	HMDB0007100*	-0.129	0.000

Appendix Table 2.2 (Continued) List of all available metabolites in women (NHS) and men (HPFS) by subclass with parameter estimates for each signature derived from elastic net regression.

Lipids and lipid-like molecules	Diacylglycerols	C34:1 DAG	HMDB0007102*	0.173	0.000
Lipids and lipid-like molecules	Diacylglycerols	C36:1 DAG	HMDB0007216*	0.002	0.000
Lipids and lipid-like molecules	Diacylglycerols	C36:2 DAG	HMDB0007218*	0.000	0.000
Lipids and lipid-like molecules	Diacylglycerols	C38:5 DAG	HMDB0007199*	0.000	0.000
Lipids and lipid-like molecules	Fatty acids	Butyrobetaine	HMDB0006831	0.053	-0.013
Lipids and lipid-like molecules	Fatty acids	Myristoleic acid	HMDB0002000	0.036	0.000
Lipids and lipid-like molecules	Glycinated bile acids and derivatives	Glycocholate	HMDB0000138	0.073	0.013
Lipids and lipid-like molecules	Glycinated bile acids and derivatives	Glycodeoxycholate	HMDB0000631*	0.009	0.000
Lipids and lipid-like molecules	Lineolic acids and derivatives	C34:2 DAG	HMDB0007103*	-0.011	0.000
Lipids and lipid-like molecules	Lineolic acids and derivatives	C34:3 DAG	HMDB0007132*	-0.067	0.007
Lipids and lipid-like molecules	Lineolic acids and derivatives	C36:3 DAG	HMDB0007219*	0.000	0.000
Lipids and lipid-like molecules	Lineolic acids and derivatives	C36:4 DAG	HMDB0007248*	0.000	0.022
Lipids and lipid-like molecules	Lysophosphatidylcholines	C16:1 LPC plasmalogen	HMDB0010407*	-0.025	0.071
Lipids and lipid-like molecules	Lysophosphatidylcholines	C18:2 LPC	HMDB0010386*	-0.112	0.000
Lipids and lipid-like molecules	Lysophosphatidylcholines	C20:5 LPC	HMDB0010397	0.000	0.000
Lipids and lipid-like molecules	Lysophosphatidylcholines	C22:5 LPC	HMDB0010403*	0.000	-0.125
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C18:1 LPE	HMDB0011506*	0.033	0.000
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C18:2 LPE	HMDB0011507*	0.000	0.000
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C18:3 LPE	HMDB0011478*	-0.157	0.000
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C20:1 LPE	HMDB0011512*	0.046	0.000
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C20:4 LPE	HMDB0011517	0.000	0.000
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C22:0 LPE	HMDB0011520	-0.002	-0.060
Lipids and lipid-like molecules	Lysophosphatidylethanolamines	C22:6 LPE	HMDB0011526	0.025	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C30:0 PC	HMDB0007869*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C30:1 PC	HMDB0007870*	0.083	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C32:0 PC	HMDB0007871*	0.176	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C32:1 PC	HMDB0007873*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C32:2 PC	HMDB0007874*	-0.089	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C34:1 PC	HMDB0007972*	-0.196	-0.154
Lipids and lipid-like molecules	Phosphatidylcholines	C34:1 PC plasmalogen	HMDB0011208*	0.073	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C34:2 PC plasmalogen	HMDB0011210*	-0.096	-0.131
Lipids and lipid-like molecules	Phosphatidylcholines	C34:3 PC plasmalogen	HMDB0011211*	-0.063	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C34:4 PC	HMDB0007883*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C34:5 PC plasmalogen	HMDB0011214*	0.004	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C36:1 PC	HMDB0008038*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C36:2 PC	HMDB0008039*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C36:2 PC plasmalogen	HMDB0011243*	0.117	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C36:3 PC	HMDB0008105*	0.037	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C36:3 PC plasmalogen	HMDB0011244*	0.000	-0.096
Lipids and lipid-like molecules	Phosphatidylcholines	C36:4 PC-A	HMDB0007983*	0.000	-0.036
Lipids and lipid-like molecules	Phosphatidylcholines	C36:4 PC-B	HMDB0008138*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C36:5 PC plasmalogen	HMDB0011220*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C38:2 PC	HMDB0008270*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C38:3 PC	HMDB0008047*	0.094	0.126
Lipids and lipid-like molecules	Phosphatidylcholines	C38:4 PC	HMDB0008048*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C38:4 PC plasmalogen	HMDB0011252*	0.000	0.000

Appendix Table 2.2 (Continued) List of all available metabolites in women (NHS) and men (HPFS) by subclass with parameter estimates for each signature derived from elastic net regression.

Lipids and lipid-like molecules	Phosphatidylcholines	C38:6 PC	HMDB0007991*	-0.097	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C38:7 PC plasmalogen	HMDB0011229*	-0.059	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C40:10 PC	HMDB0008511*	-0.106	-0.019
Lipids and lipid-like molecules	Phosphatidylcholines	C40:6 PC	HMDB0008057*	0.157	0.000
Lipids and lipid-like molecules	Phosphatidylcholines	C40:9 PC	HMDB0008731*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C34:0 PE	HMDB0008925*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C34:2 PE	HMDB0008928*	0.090	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C34:2 PE plasmalogen	HMDB0008952*	0.084	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C34:3 PE plasmalogen	HMDB0011343*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C36:1 PE plasmalogen	HMDB0009016*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C36:2 PE	HMDB0008994*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C36:2 PE plasmalogen	HMDB0009082*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C36:3 PE	HMDB0009060*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C36:3 PE plasmalogen	HMDB0011441*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C36:4 PE	HMDB0008937*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C36:4 PE plasmalogen	HMDB0011442*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C36:5 PE plasmalogen	HMDB0011410*	0.000	0.033
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:2 PE	HMDB0008942*	-0.164	-0.157
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:3 PE plasmalogen	HMDB0011384*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:4 PE	HMDB0009003*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:5 PE	HMDB0009069*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:5 PE plasmalogen	HMDB0011253*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:5 PE plasmalogen	HMDB0011386*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:6 PE	HMDB0009102*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:6 PE plasmalogen	HMDB0011387*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C38:7 PE plasmalogen	HMDB0011420*	-0.040	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C40:6 PE	HMDB0009012*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylethanolamines	C40:7 PE plasmalogen	HMDB0011394*	0.000	0.000
Lipids and lipid-like molecules	Phosphatidylinositols	C38:4 PI	HMDB0009815*	-0.061	0.000
Lipids and lipid-like molecules	Phosphatidylserines	C34:0 PS	HMDB0012356*	0.035	0.000
Lipids and lipid-like molecules	Phosphocholines	Phosphocholine	HMDB0001565	-0.085	-0.121
Lipids and lipid-like molecules	Phosphosphingolipids	C16:0 SM	HMDB0010169	0.000	0.000
Lipids and lipid-like molecules	Phosphosphingolipids	C18:0 SM	HMDB0001348	0.000	0.239
Lipids and lipid-like molecules	Phosphosphingolipids	C20:0 SM	HMDB0012102	-0.017	0.000
Lipids and lipid-like molecules	Phosphosphingolipids	C24:1 SM	HMDB0012107*	0.124	0.000
Lipids and lipid-like molecules	Steroids and derivatives	21-deoxycortisol	HMDB0004030	-0.039	-0.039
Lipids and lipid-like molecules	Steroids and derivatives	Campesterol	HMDB0002869	-0.159	0.000
Lipids and lipid-like molecules	Steroids and derivatives	Cholesterol	HMDB0000067	-0.041	0.000
Lipids and lipid-like molecules	Steroids and derivatives	Cortisol	HMDB0000063	0.110	-0.054
Lipids and lipid-like molecules	Steroids and derivatives	Cortisone	HMDB0002802	-0.056	-0.061
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C43:0 TAG	HMDB0042062*	0.040	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C43:1 TAG	HMDB0042098*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C44:0 TAG	HMDB0042063*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C45:1 TAG	HMDB0042099*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C45:2 TAG	HMDB0043170*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C46:0 TAG	HMDB0010411*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C46:1 TAG	HMDB0010412*	0.019	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C46:2 TAG	HMDB0010419*	0.000	0.000

Appendix Table 2.2 (Continued) List of all available metabolites in women (NHS) and men (HPFS) by subclass with parameter estimates for each signature derived from elastic net regression.

Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C47:1 TAG	HMDB0042100*	0.021	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C47:2 TAG	HMDB0042076*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C48:0 TAG	HMDB0005356*	-0.027	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C48:1 TAG	HMDB0005359*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C48:2 TAG	HMDB0005376*	0.076	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C49:2 TAG	HMDB0011706*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C50:0 TAG	HMDB0005357*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C50:1 TAG	HMDB0005360*	0.000	0.076
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C50:2 TAG	HMDB0005377*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C51:0 TAG	HMDB0031106*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C51:1 TAG	HMDB0042104*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C51:2 TAG	HMDB0005362*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C52:0 TAG	HMDB0005365*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C52:1 TAG	HMDB0005367*	0.000	0.033
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C52:2 TAG	HMDB0005369*	0.033	0.001
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C53:2 TAG	HMDB0042196*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C54:1 TAG	HMDB0005395*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C54:2 TAG	HMDB0005403*	0.009	0.081
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C55:2 TAG	HMDB0042226*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C56:1 TAG	HMDB0005396*	0.016	0.000
Lipids and lipid-like molecules	Triacylglycerols < 3 double bonds	C56:2 TAG	HMDB0005404*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C48:3 TAG	HMDB0005432*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C49:3 TAG	HMDB0042103*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C50:3 TAG	HMDB0005433*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C50:4 TAG	HMDB0005435*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C50:5 TAG	HMDB0010471*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C50:6 TAG	HMDB0010497*	-0.108	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C51:3 TAG	HMDB0011701*	-0.043	0.032
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C52:3 TAG	HMDB0005384*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C52:4 TAG	HMDB0005363*	0.000	0.001
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C52:5 TAG	HMDB0005380*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C52:6 TAG	HMDB0005436*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C52:7 TAG	HMDB0010517*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C53:3 TAG	HMDB0043058*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:3 TAG	HMDB0005405*	0.036	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:4 TAG	HMDB0005370*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:5 TAG	HMDB0005385*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:6 TAG	HMDB0005391*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:7 TAG	HMDB0005447*	0.025	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:8 TAG	HMDB0010518*	0.056	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C54:9 TAG	HMDB0010498*	0.000	0.008
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C55:3 TAG	HMDB0042466*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:3 TAG	HMDB0005410*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:4 TAG	HMDB0005398*	-0.040	-0.021
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:5 TAG	HMDB0005406*	0.104	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:6 TAG	HMDB0005456*	0.058	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:7 TAG	HMDB0005462*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:8 TAG	HMDB0005392*	0.000	0.000

Appendix Table 2.2 (Continued) List of all available metabolites in women (NHS) and men (HPFS) by subclass with parameter estimates for each signature derived from elastic net regression.

Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C56:9 TAG	HMDB0005448*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C58:11 TAG	HMDB0010531*	0.000	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C58:6 TAG	HMDB0005458*	0.000	-0.012
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C58:7 TAG	HMDB0005471*	-0.067	-0.024
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C58:8 TAG	HMDB0005413*	-0.015	-0.058
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C58:9 TAG	HMDB0005463*	-0.022	0.000
Lipids and lipid-like molecules	Triacylglycerols > 3 double bonds	C60:12 TAG	HMDB0005478*	0.000	0.000
Lipids and lipid-like molecules	Ubiquinones	Coenzyme Q10	HMDB0001072	0.000	0.052
Organic acids and derivatives	Alpha amino acid derivatives	Creatine	HMDB0000064	-0.010	0.134
Organic acids and derivatives	Alpha amino acid derivatives	Creatinine	HMDB0000562	0.048	0.000
Organic acids and derivatives	Alpha amino acid derivatives	Guanidinoacetic acid	HMDB0000128	-0.112	0.000
Organic acids and derivatives	Alpha amino acid derivatives	Thyroxine	HMDB0000248	0.000	-0.071
Organic acids and derivatives	Alpha amino acids (BCAA)	Isoleucine	HMDB0000172	0.099	0.000
Organic acids and derivatives	Alpha amino acids (BCAA)	Leucine	HMDB0000687	-0.109	0.000
Organic acids and derivatives	Alpha amino acids (BCAA)	Valine	HMDB0000883	0.000	0.232
Organic acids and derivatives	Alpha amino acids (non-BCAA)	2-aminooctanoic acid	HMDB0000991*	0.023	-0.009
Organic acids and derivatives	Alpha amino acids (non-BCAA)	3-methylhistidine	HMDB0000479	0.000	-0.015
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Alanine	HMDB0000161	-0.103	0.095
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Aminoisobutyric acid	HMDB0001906	-0.049	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Asparagine	HMDB0000168	-0.003	-0.123
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Betaine	HMDB0000043	0.099	-0.051
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Citrulline	HMDB0000904	0.000	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Dimethylglycine	HMDB0000092	0.000	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Glutamine	HMDB0000641	-0.053	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Glycine	HMDB0000123	0.022	-0.116
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Histidine	HMDB0000177	-0.112	-0.006
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Homoarginine	HMDB0000670	0.044	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Homocitrulline	HMDB0000679	0.000	-0.011
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Hydroxyproline	HMDB0000725	0.011	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Lysine	HMDB0000182	-0.006	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Methionine	HMDB0000696	0.020	-0.035
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Methionine sulfoxide	HMDB0002005	-0.016	0.015
Organic acids and derivatives	Alpha amino acids (non-BCAA)	N6,N6-dimethyllysine	HMDB0013287	-0.016	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	N6,N6,N6-trimethyllysine	HMDB0001325	0.061	0.021
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Phenylalanine	HMDB0000159	0.005	0.036
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Pipecolic acid	HMDB0000716	0.026	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Proline	HMDB0000162	0.000	0.078
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Proline betaine	HMDB0004827	0.000	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	SDMA	HMDB0003334	0.000	-0.047
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Serine	HMDB0000187	-0.026	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Threonine	HMDB0000167	0.035	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Tryptophan	HMDB0000929	-0.075	0.000
Organic acids and derivatives	Alpha amino acids (non-BCAA)	Tyrosine	HMDB0000158	0.000	0.084
Organic acids and derivatives	Carboximidic acids	N1-acetylspermidine	HMDB0001276	0.015	0.000
Organic acids and derivatives	Carboximidic acids	Palmitoylethanolamide	HMDB0002100	0.000	0.000
Organic acids and derivatives	Carboximidic acids	Pantothenate	HMDB0000210	-0.001	-0.013
Organic acids and derivatives	Gamma amino acids and derivatives	4-acetamidobutanoate	HMDB0003681	-0.003	0.000

Appendix Table 2.2 (Continued) List of all available metabolites in women (NHS) and men (HPFS) by subclass with parameter estimates for each signature derived from elastic net regression.

Organic acids and derivatives	N-acyl-alpha amino acids	Cinnamoylglycine	HMDB0011621	-0.072	-0.019
Organic acids and derivatives	N-acyl-alpha amino acids	N-Acetyl-L-alanine	HMDB0000766	0.020	0.000
Organic acids and derivatives	N-acyl-alpha amino acids	N-acetylhistidine	HMDB0032055	0.016	-0.067
Organic acids and derivatives	N-acyl-alpha amino acids	N-acetylornithine	HMDB0003357	-0.007	-0.045
Organic acids and derivatives	N-acyl-alpha amino acids	N-acetyltryptophan	HMDB0013713	0.002	0.000
Organic acids and derivatives	N-acyl-alpha amino acids	N-alpha-acetylarginine	HMDB0004620	0.000	0.000
Organic acids and derivatives	N-acyl-alpha amino acids	N-lauroylglycine	HMDB0013272	-0.001	0.004
Organic acids and derivatives	N-acyl-alpha amino acids	N6-acetyllysine	HMDB0000206	0.011	0.000
Organic acids and derivatives	N-acyl-alpha amino acids	Oleoyl glycine	HMDB0013631	-0.041	-0.009
Organic acids and derivatives	N-acyl-alpha amino acids	Phenylacetylglutamine	HMDB0006344	-0.003	0.000
Organic acids and derivatives	N-acyl-alpha-hexosamines	Acetyl-galactosamine	HMDB0000212	0.159	0.000
Organic acids and derivatives	Ureas	N-carbamoyl-beta-alanine	HMDB0000026	0.053	0.000
Organic acids and derivatives	Ureas	Uric acid	HMDB0000289	0.000	0.090
Organic nitrogen compounds	2-arylethylamines	1-methylhistamine	HMDB0000898	0.029	0.000
Organic nitrogen compounds	Amine oxides	Trimethylamine-N-oxide	HMDB0000925	0.000	0.000
Organic nitrogen compounds	N-acetylarylamines	5-acetylamino-6-amino-3-methyluracil	HMDB0004400	0.057	0.000
Organic oxygen compounds	Alkyl-phenylketones	4-hydroxy-3-methylacetophenone	HMDB0059824	0.065	0.043
Organoheterocyclic compounds	Bilirubins	Bilirubin	HMDB0000054	-0.023	0.000
Organoheterocyclic compounds	Bilirubins	Biliverdin	HMDB0001008	-0.076	-0.008
Organoheterocyclic compounds	Imidazoles	Allantoin	HMDB0000462	0.043	0.000
Organoheterocyclic compounds	Imidazoles	Methylimidazole acetic acid	HMDB0002820	0.000	0.000
Organoheterocyclic compounds	Imidazoles	Urocanic acid	HMDB0000301	-0.033	-0.001
Organoheterocyclic compounds	Nicotinamides	1-methylnicotinamide	HMDB0000699	-0.037	0.000
Organoheterocyclic compounds	Nicotinamides	N1-methyl-2-pyridone-5-carboxamide	HMDB0004193	0.023	0.000
Organoheterocyclic compounds	Oxirane carboxylic acid or derivatives	Cerulenin	HMDB0015168	0.052	0.000
Organoheterocyclic compounds	Purines and derivatives	1-methylguanine	HMDB0003282	-0.004	0.000
Organoheterocyclic compounds	Purines and derivatives	1-methylguanosine	HMDB0001563	0.078	0.025
Organoheterocyclic compounds	Purines and derivatives	2-methylguanosine	HMDB0005862	0.039	0.015
Organoheterocyclic compounds	Purines and derivatives	7-methylguanine	HMDB0000897	0.077	0.000
Organoheterocyclic compounds	Purines and derivatives	N2,N2-dimethylguanosine	HMDB0004824	0.027	0.156
Organoheterocyclic compounds	Pyridines and derivatives	Pyridoxamine	HMDB0001431	0.000	0.000
Organoheterocyclic compounds	Pyrimidines and derivatives	5-hydroxymethyl-4-methyluracil	HMDB0000544	0.000	0.050
Organoheterocyclic compounds	Pyrimidines and derivatives	Cytosine	HMDB0000630	-0.060	-0.073
Organoheterocyclic compounds	Pyrimidines and derivatives	N4-acetylcytidine	HMDB0005923	-0.017	0.062
Organoheterocyclic compounds	Pyrimidines and derivatives	Pseudouridine	HMDB0000767	0.380	0.000
Organoheterocyclic compounds	Pyrimidines and derivatives	Ribothymidine	HMDB0000884	-0.200	-0.007
Organoheterocyclic compounds	Pyrimidines and derivatives	Thiamine	HMDB0000235	-0.032	0.000
Organoheterocyclic compounds	Xanthines	1,7-dimethyluric acid	HMDB0011103	-0.042	0.000
Organoheterocyclic compounds	Xanthines	3-methylxanthine	HMDB0001886	-0.050	0.000
Organoheterocyclic compounds	Xanthines	7-methylxanthine	HMDB0001991	0.029	0.000
Organoheterocyclic compounds	Xenobiotics	Caffeine	HMDB0001847	-0.022	0.000
Xenobiotics	Xenobiotics	Acetaminophen	HMDB0001859	0.000	0.000
Xenobiotics	Xenobiotics	Gabapentin	HMDB0005015	-0.068	-0.036
Xenobiotics	Xenobiotics	Metformin	HMDB0001921	-0.052	0.000
Xenobiotics	Xenobiotics	Sulfamethoxazole	HMDB0015150	-0.008	-0.018

* representative ID

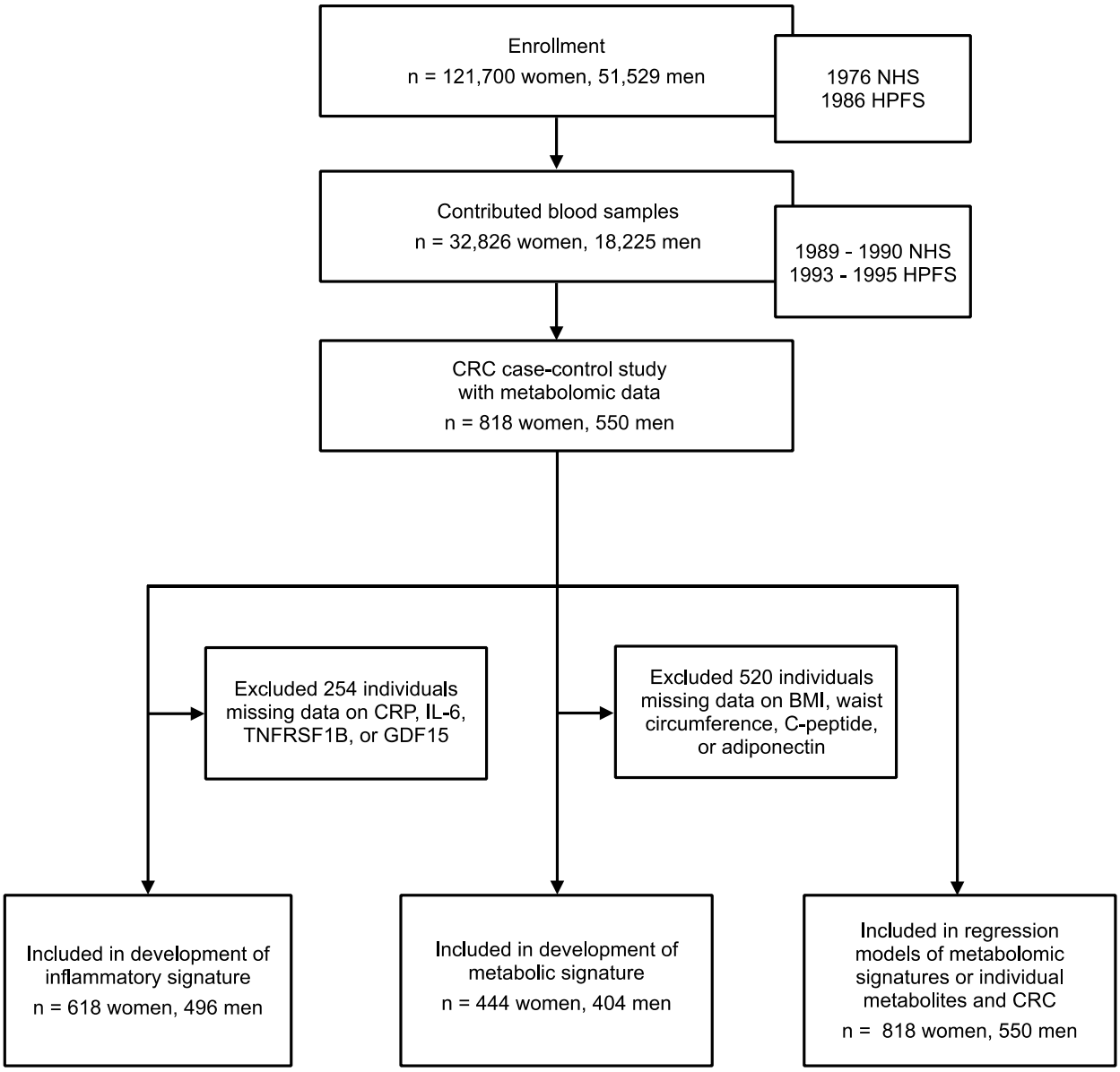
Appendix Table 2.3. Multivariable-adjusted[a] odds of CRC associated with a 1-SD increase in metabolomic signatures of inflammation and metabolic dysregulation, in men and women, by cancer subsite, common molecular subtype, and KRAS mutation status

	Signature of inflammation		Signature of metabolic dysregulation	
	Women (NHS)	Men (HPFS)	Women (NHS)	Men (HPFS)
Tumor site				
Proximal Colon				
Cases	201	104	201	104
OR (95%CI)	1.04 (0.87, 1.24)	1.23 (0.92, 1.65)	1.01 (0.85, 1.19)	1.14 (0.86, 1.51)
Distal Colon				
Cases	109	75	109	75
OR (95%CI)	1.07 (0.86, 1.34)	1.34 (0.98, 1.85)	1.03 (0.82, 1.28)	1.21 (0.89, 1.66)
Rectum				
Cases	85	68	85	68
OR (95%CI)	1.06 (0.82, 1.36)	1.17 (0.83, 1.66)	0.98 (0.76, 1.26)	1.21 (0.86, 1.70)
$P_{heterogeneity}$[a]	0.89	0.75	0.90	0.89
Molecular subtype				
Conventional pathway				
Cases	45	68	45	68
OR (95%CI)	0.98 (0.70, 1.39)	1.32 (0.94, 1.84)	1.17 (0.84, 1.63)	1.02 (0.73, 1.42)
Serrated pathway				
Cases	23	9	23	9
OR (95%CI)	1.02 (0.66, 1.60)	0.94 (0.36, 2.48)	0.79 (0.47, 1.33)	1.20 (0.49, 2.91)
Alternate pathway				
Cases	58	48	58	48
OR (95%CI)	1.04 (0.78, 1.40)	0.98 (0.64, 1.51)	1.00 (0.74, 1.35)	1.38 (0.93, 2.07)
Other pathway				
Cases	27	15	27	15
OR (95%CI)	0.94 (0.54, 1.30)	2.00 (0.96, 4.16)	0.99 (0.66, 1.48)	2.60 (1.25, 5.43)
$P_{heterogeneity}$[a]	0.64	0.45	0.66	0.11
KRAS mutation status				
KRAS mutation				
Cases	72	61	72	61
OR (95%CI)	0.92 (0.70, 1.22)	1.07 (0.73, 1.56)	0.92 (0.70, 1.22)	1.43 (1.00, 2.04)
KRAS wild type				
Cases	94	102	94	102
OR (95%CI)	1.04 (0.82, 1.32)	1.09 (0.81, 1.46)	1.18 (0.94, 1.49)	1.06 (0.80, 1.41)
$P_{heterogeneity}$[a]	0.65	0.94	0.15	0.17

All models show the odds ratio for a 1-SD increase in metabolomic signature and are adjusted for age at blood collection, date of blood collection, fasting status (yes or no), height, physical activity (METS-hours/week), alcohol intake (0, 0.1-5 g/day, 5-10 g/day, 10-15 g/day, 15+ g/day), pack-years of smoking, history of CRC in parent or sibling (yes or no), history of prior endoscopy (yes or no), regular aspirin use (>2 tablets/week, yes or no), alternative healthy eating index (AHEI) score. Models in women additionally adjusted for menopausal status (pre- or postmenopausal) and postmenopausal hormone use (never, past, or current).

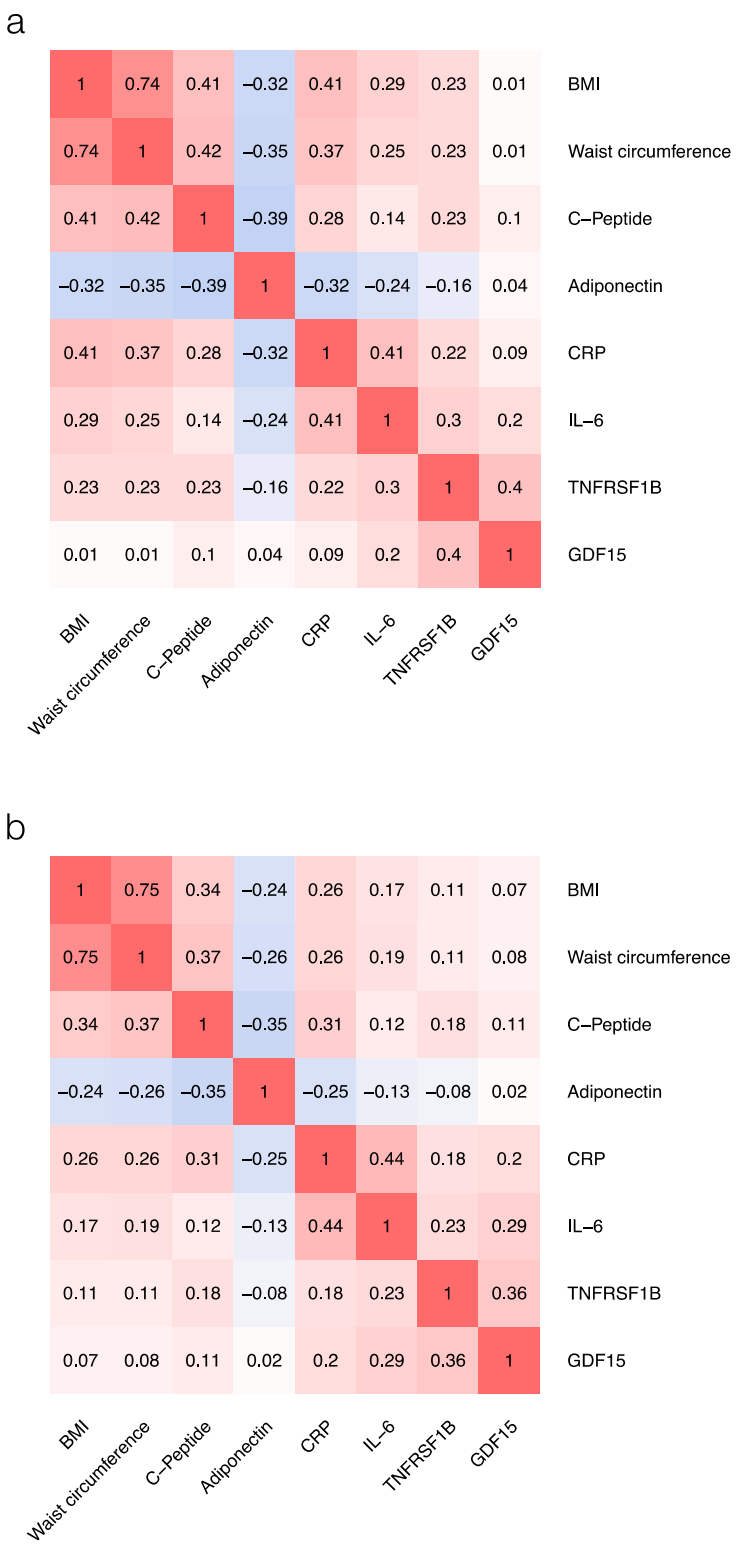
[a] Heterogeneity was estimated using the contrast test method for a global difference between subtypes.

Abbreviations: CI, confidence interval; CRC, colorectal cancer; Health Professionals Follow-up Study; METS, metabolic equivalent of task score; NHSI, Nurses' Health Study I; OR, odds ratio



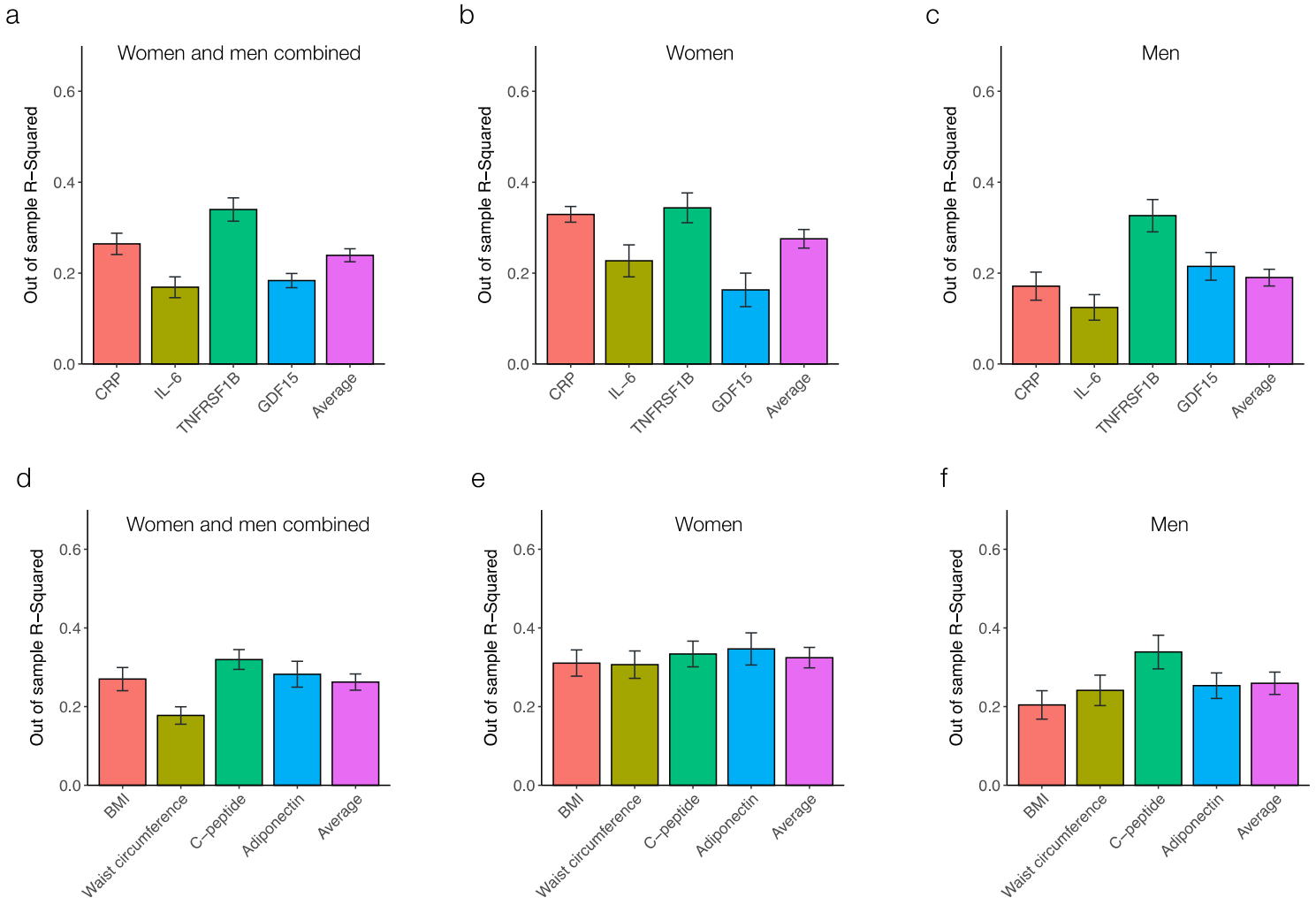
Appendix Figure 2.1. Participant flowchart

Abbreviations: BMI, body mass index; CRC, colorectal cancer; CRP, C-reactive protein; GDF15, growth differentiation factor 15; HPFS, Health Professionals Follow-up Study; IL-6, interleukin-6; NHS, Nurses' Health Study; TNFRSF1B, tumor necrosis factor receptor super family 1B



Appendix Figure 2.2. Correlation among the response variables used to derive the signatures of inflammation (CRP, IL-6, TNFRSF1B, GDF15) and metabolic dysregulation (BMI, waist circumference, C-peptide, adiponectin) in **a)** Nurses' Health Study (n = 478) and **b)** Health Professionals Follow-up Study (n = 613). Spearman partial correlation coefficient after adjustment for age and fasting status at blood draw is shown for each pairing.

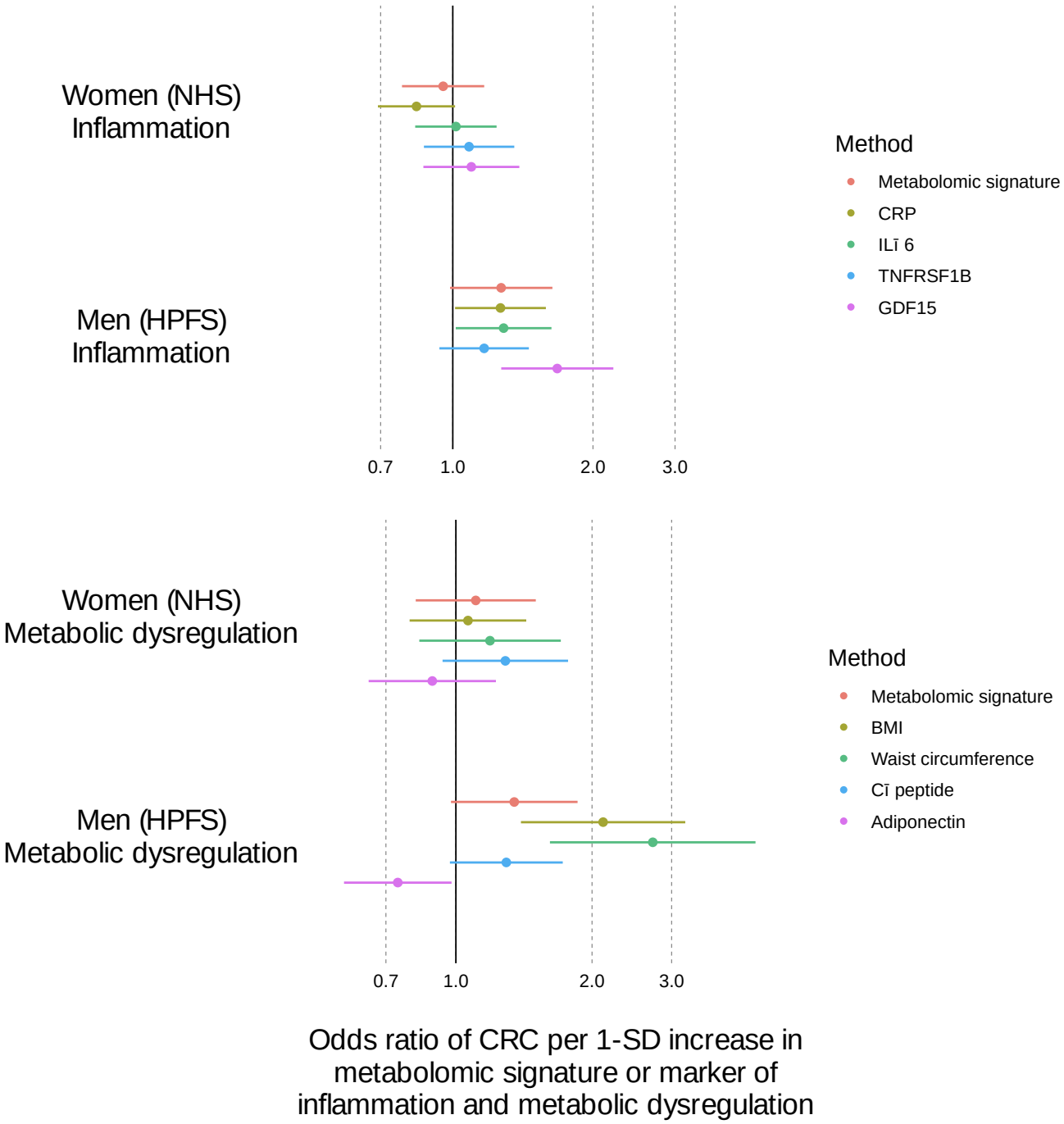
Abbreviations: BMI, body mass index; CRP, C-reactive protein; GDF15, growth differentiation factor 15; IL-6, interleukin-6; TNFRSF1B, tumor necrosis factor receptor super family 1B



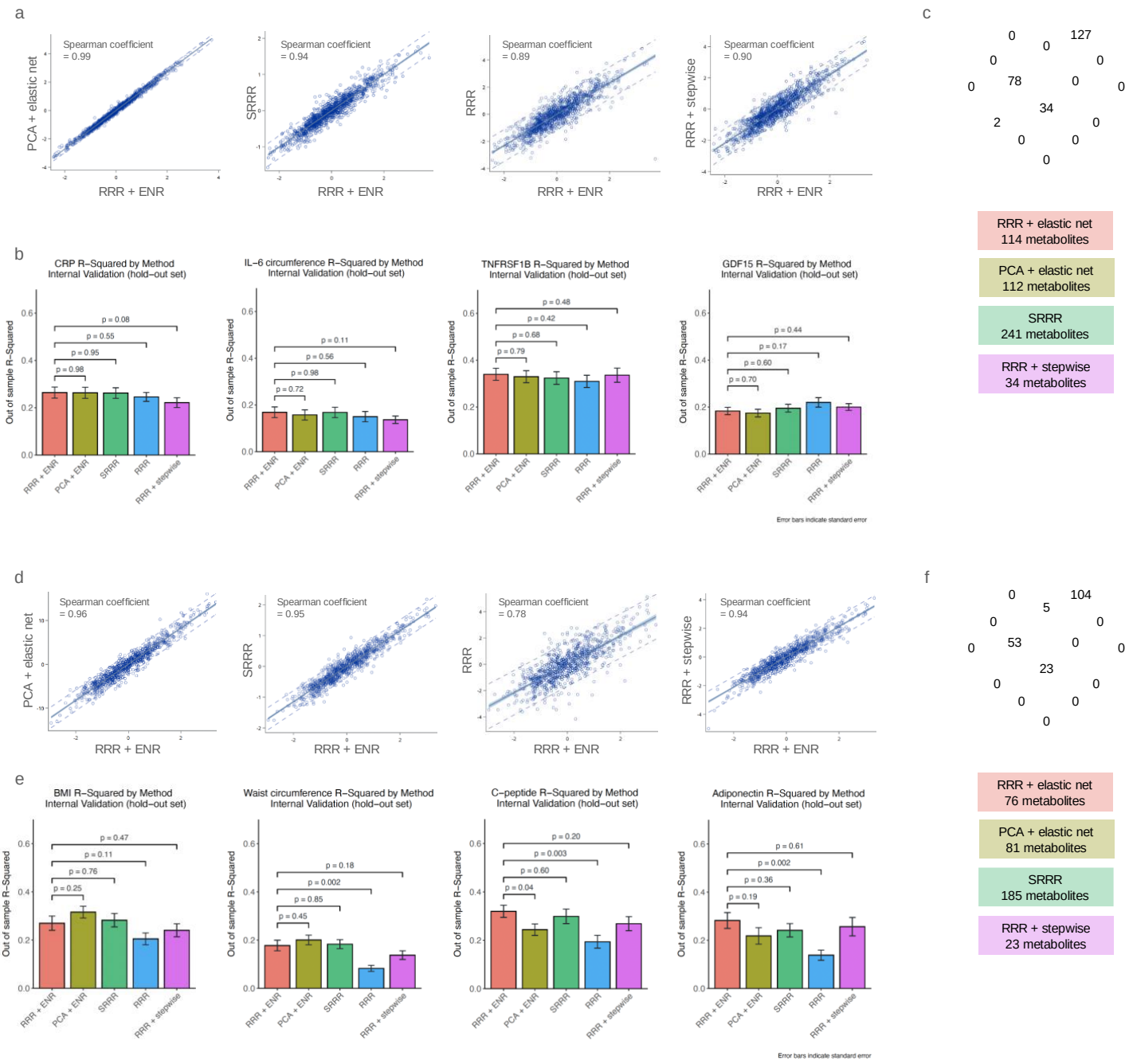
Appendix Figure 2.3. Proportion of variance in markers of inflammation and metabolic dysregulation explained by the metabolomic signatures

Proportion of variance explained (R-squared) by the metabolomic signature of inflammation in each of the markers of inflammation (a-c) and proportion of variance explained by the metabolomic signature of metabolic dysregulation in each of the markers of metabolic dysregulation (d-f) among women and men combined (a,d), women (b,e), and men (c,f). Error bars indicate standard error. R-squared is the average adjusted R-squared from 10 out of sample iterations.

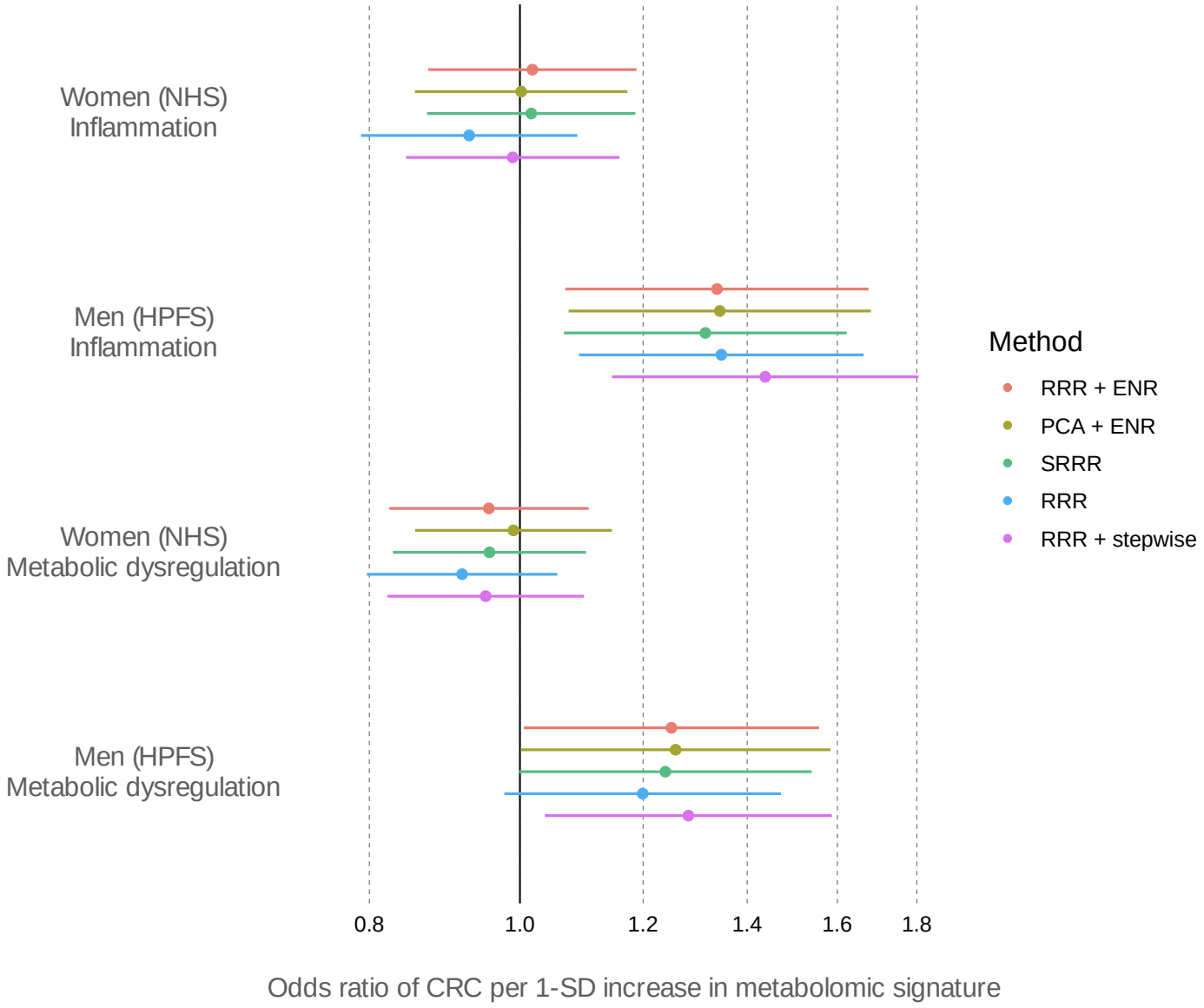
Abbreviations: BMI, body mass index; CRP, C-reactive protein; GDF15, growth differentiation factor 15; IL-6, interleukin-6; NHS, Nurses' Health Study; TNFRSF1B, tumor necrosis factor receptor super family 1B



Appendix Figure 2.4. The association of the metabolomic signatures and the individual markers used to derive the signatures with CRC, among women and men with complete data on all individual markers and metabolomic data (NHS, inflammation n = 618; HPFS, inflammation n = 496; NHS, metabolic n = 444; HPFS, metabolic n = 404). The multivariable adjusted OR per 1-SD increase in metabolomic signature is shown. All models were adjusted for age at blood collection, date of blood collection, fasting status (yes or no), height, physical activity (METS-hours/week), alcohol intake (0, 0.1-5 g/day, 5-10 g/day, 10-15 g/day, 15+ g/day), pack-years of smoking, history of CRC in parent or sibling (yes or no), history of prior endoscopy (yes or no), regular aspirin use (>2 tablets/week, yes or no), alternative healthy eating index (AHEI) score. Models in women additionally adjusted for menopausal status (pre- or postmenopausal) and postmenopausal hormone use (never, past, or current).



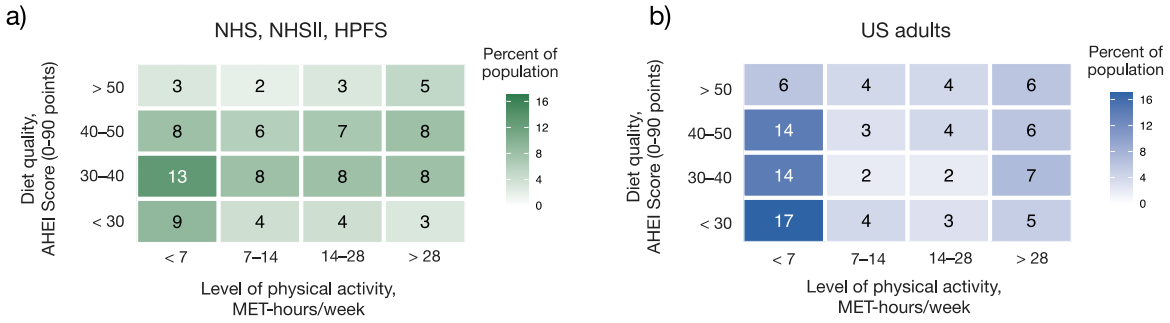
Appendix Figure 2.5. Comparing methods to derive the metabolomic signatures. Spearman correlation coefficient was estimated for metabolomic signatures derived using various methods and metabolomic signatures derived using RRR/ENR (main analysis), a) signature of inflammation and d) signature of metabolic dysregulation. Estimated R-squared for the association between the signatures derived using each method and individual markers of b) inflammation and e) metabolic dysregulation in the ten hold-out sets, with average R-squared compared using t-test. The common metabolites retained in each signature are shown c) for inflammation and f) metabolic dysregulation.



Appendix Figure 2.6. The association of metabolomic signatures derived using various methods with CRC. The multivariable adjusted OR per 1-SD increase in metabolomic signature is shown. All models were adjusted for age at blood collection, date of blood collection, fasting status (yes or no), height, physical activity (METS-hours/week), alcohol intake (0, 0.1-5 g/day, 5-10 g/day, 10-15 g/day, 15+ g/day), pack-years of smoking, history of CRC in parent or sibling (yes or no), history of prior endoscopy (yes or no), regular aspirin use (>2 tablets/week, yes or no), alternative healthy eating index (AHEI) score. Models in women additionally adjusted for menopausal status (pre- or postmenopausal) and postmenopausal hormone use (never, past, or current).

APPENDIX THREE

Appendix Figure 3.1



Appendix Figure 3.1. Prevalence of diet quality and physical activity levels in a) NHS, NHSII, and HPFS cohorts and b) the estimated prevalence among US adults in 2018.

Abbreviations: AHEI, alternative healthy eating index; HPFS, Health Professionals Follow-up Study; MET, metabolic equivalent of task MET, metabolic equivalent of task; NHS, Nurses' Health Study; US, United States