Prospective association of energy balance scores based on metabolic biomarkers with colorectal cancer risk.

Mark A. Guinter,[1] Susan M. Gapstur,[1] Marjorie L. McCullough,[1] W. Dana Flanders,[2,3] Ying Wang,[1] Erika Rees-Punia,[1] Kassandra I. Alcaraz,[1] Michael N. Pollak,[4] Peter T. Campbell,[1]

[1]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA; [2]Department of Epidemiology, Rollins School of Public Health, Emory University; [3]Winship Cancer Institute, Emory University, Atlanta, GA; [4]Center for Translational Research in Cancer, McGill University, Montreal, QC.

Running title: Energy balance scores and colorectal cancer

Correspondence:
Mark Guinter
250 Williams St #6D.439
Atlanta, GA 30303
404.329.4455
mark.guinter@cancer.org

All authors declare no conflict of interest

Abbreviations:

BMI	Body mass index
CI	Confidence interval
CIMP	CpG island methylator phenotype
CPS-II	Cancer Prevention Study-II
CV	Coefficient of variation
HbA_{1c}	HemoglobinA_{1c}
HR	Hazard ratio
hsCRP	High-sensitivity CRP
ICC	Intraclass correlation
IPW	Inverse probability weighting
MET	Metabolic equivalent of task
MHT	Menopausal hormone therapy
MSI	Microsatellite instability
MSS	Microsatellite stable
PCR	Polymerase chain reaction
PMR	Percent of methylated reference
RRR	Reduced rank regression
WT	Wildtype

ABSTRACT:

Background: Energy balance-related factors such as body mass index (BMI), diet, and physical activity may influence colorectal cancer (CRC) etiology through inter-connected metabolic pathways, but their combined influence is less clear.

Methods: We used reduced rank regression to derive three energy balance scores that associate lifestyle factors with combinations of pre-diagnostic, circulating levels of high-sensitivity C-reactive protein (hsCRP), C-peptide, and hemoglobin A_{1c} (HbA_{1c}) among 2,498 participants in the Cancer Prevention Study-II Nutrition Cohort. Among 114,989 participants, we verified 2,228 CRC cases. We assessed associations of each score with CRC incidence and by tumor molecular phenotypes using Cox proportional hazards regression.

Results: The derived scores comprised BMI, physical activity, screen time, and 14 food groups, and explained 5.1 to 10.5% of the variation in biomarkers. The hazard ratio (HR) and 95% confidence interval (CI) for quartile 4 vs. 1 of the HbA_{1c}+C-peptide-based score and CRC was 1.30 (1.15, 1.47), the hsCRP-based score was 1.35 (1.19, 1.53), and the hsCRP, C-peptide, and HbA_{1c}-based score was 1.35 (1.19, 1.52). The latter score was associated with non-CIMP tumors (HR_{Q4vsQ1}: 1.59; 95% CI: 1.17, 2.16) but not CIMP-positive tumors ($p_{heterogeneity}$=0.04).

Conclusions: These results further support hypotheses that systemic biomarkers of metabolic health –inflammation and abnormal glucose homeostasis– mediate part of the relationship between several energy balance-related modifiable factors and CRC risk.

Impact: Results support cancer prevention guidelines for maintaining a healthful body weight, consuming a healthful diet and being physically active. More research is needed on these clusters of exposures with molecular phenotypes of tumors.

Introduction

An estimated 55% of colorectal cancers diagnosed in the US in 2014 were attributed to modifiable lifestyle factors, underscoring the importance of lifestyle in the development of colorectal cancer (1). Most of the evidence linking lifestyle and colorectal cancer comes from investigations of individual lifestyle factors without consideration of their downstream metabolic effects. Moreover, lifestyle risk factors are highly correlated and have synergistic effects on health (2).

Energy balance-related lifestyle risk factors associated with colorectal cancer include excess body fat, poor diet, physical inactivity, and sedentary behavior, which are particularly susceptible to clustering (3) and share common pathways in colorectal carcinogenesis, including chronic systemic inflammation and insulin resistance (4). While risk estimates are available for the association between individual energy balance-related factors and colorectal cancer (5) their combined influence is unclear.

Associations of multiple highly correlated lifestyle factors with disease outcomes are examined using indices or scores. These scores are rarely based on mechanisms or clinical biomarkers of disease risk. Reduced rank regression (RRR) provides a robust method to derive weighted lifestyle scores with an *a priori* definition of hypothesized mechanisms via use of biomarkers reflective of certain pathways or processes. RRR is a method for reducing large amounts of correlated explanatory variables to a smaller set of latent variables that maximize the explained variation in a single or set of responses variables, often an intermediate marker for disease risk (6). RRR has been used to identify associations between dietary patterns with disease risk (7,8), but to date, no published study has developed a score of lifestyle factors using RRR.

In the present analysis, we used RRR to derive and validate three energy balance scores in a subset of Cancer Prevention Study II (CPS-II) Nutrition Cohort participants with pre-diagnostic measures of high sensitivity C-reactive protein (hsCRP), C-peptide, and hemoglobin A_{1c} (HbA_{1c}), which are, respectively, established clinical markers for general inflammation, hyperinsulinemia, and hyperglycemia; the latter two together represent glucose homeostasis. We examined associations of the derived energy balance scores with incident colorectal cancer risk and of tumor molecular phenotypes (where available), among all eligible men and women in the CPS-II Nutrition Cohort.

Materials and Methods

Participants were from the CPS-II Nutrition Cohort, a prospective study of cancer incidence and mortality (9). Briefly, at enrollment (1992/1993), 184,185 men and women completed a 10-page self-administered questionnaire on medical history and lifestyle factors. Follow-up questionnaires were sent biennially starting in 1997 to update exposure information and to ascertain new cancer diagnoses. From 1998 to 2001, a subset of 39,371 participants provided non-fasting blood samples. Blood samples were shipped chilled overnight to a central repository for long-term storage. The CPS-II Nutrition Cohort is approved by the Institutional Review Board of Emory University.

In the present analysis, all participants who returned the 1999 survey were eligible (n=151,342). Exclusions included: prevalent cancer except for non-melanoma skin cancer (n=28,472), loss to follow up (n=4,497), unverified diagnosis (n=174), invalid end-of-study time (n=7), invalid dietary data (n=2,184), and missing lifestyle data from all applicable survey cycles (1992, 1997, 1999; n=1,019). The final analytic sample comprised 114,989 men and women.

Exposure Assessment

Diet, physical activity, sedentary behaviors, and body mass index (BMI) were used to develop the scores because they represent a comprehensive characterization of modifiable exposures pertaining to energy balance. Intakes of 33 food groups were assessed in 1999 using a modified Willett food frequency questionnaire (FFQ, refs. 10,11) or in 1992 using a modified Block FFQ (10,11) if the 1999 FFQ was incomplete. Other self-reported information in 1999 (or in 1997 and 1992 if 1999 was missing) was used to characterize BMI and moderate-to-vigorous intensity physical activity (MVPA) MET-hours/week. Screen time hours/week, a valid proxy for sedentary time (12), was assessed in 1999 (or in 1992 if missing in 1999; not assessed in 1997). To mitigate the potential for misclassification bias between information on exposures carried forward (due to missingness) from past surveys and baseline in 1999, we weighted individuals in Cox models based on the proportion of the lifestyle factors that were measured at the 1999 survey. Complete information on dietary was factors considered to be one of four energy balance-related factors for the purpose of calculating the weight. Thus, individuals with complete data in 1999 were given a full weight of 1 and those with fewer available data in 1999 received a lower weight (e.g., 3/4 lifestyle factors available received a weight of 0.75), thereby having a smaller influence on HR estimates. Complete data in 1999 was available for 81.7% of participants. Missingness of the exposures was not associated with incident colorectal cancer.

Biomarker Measurement

Using non-fasting blood samples, circulating concentrations of hsCRP, C-peptide, and HbA_{1c} were measured from serum in a case-cohort study consisting of a random subcohort of 3,000 participants and 2,962 diabetes-related cancers (including colorectal cancers). All three

biomarkers have been shown to be reliable and clinically useful when measured from a non-fasting state (13-15). The biomarkers represent states of metabolic dysfunction that develop over a period of multiple years.

Lab personnel were blinded to case-control status and all plates included blinded QC samples. Human CRP Immunoassay (R&D Systems, Inc., Minneapolis, MN), a quantitative sandwich enzyme immunoassay technique, was used to measure hsCRP. The coefficient of variation (CV) for hsCRP was 7.4%, with an intraclass correlation coefficient (ICC) of 99.8% for CPS-II samples. The C-peptide ELISA (Ansh Labs, Webster, TX), an enzymatically amplified one-step sandwich-type immunoassay, was used to measure C-peptide. The CV for C-peptide was 7.7%, with an ICC of 97.5% for CPS-II samples. The HbA_{1c} assay is an enzymatic measurement in which lysed whole blood samples are subjected to extensive protease digestion. The CV for HbA_{1c} was 8.9%, with an ICC of 74.7% for CPS-II samples.

Derivation and Validation of Energy Balance Scores

The 3,000 subcohort participants from the CPS-II nested case-cohort were used to derive energy balance scores. Seventeen participants were excluded for lack of biomarker data, and an additional 299 participants were excluded because of incomplete lifestyle information in 1999. We additionally excluded 186 participants who self-reported a diabetes diagnosis, as their reported lifestyle information may reflect post-diagnosis lifestyle changes. Ten participants were excluded from models with hsCRP because levels were indicative of acute inflammation (>40,000 µg/L). Of the remaining 2,498 participants, 80% (n=1,999) were used to derive energy balance scores and the remaining 20% (n=499) were used for validation. Values for all three biomarkers were log-transformed. RRR was used to create a factor score, which is a single

continuous value that represents a linear combination of the energy balance-related factors, as determined by eigenvectors of the biomarker covariance matrix. In all models, only the first factor score was retained as it represents the weighted combination of energy balance factors that explains the greatest amount of variation in the biomarkers. First, RRR was performed in a single model with men and women using all three biomarkers as dependent variables to identify which individual factors were most strongly associated with the biomarkers for the overall sample. The individual lifestyle factors that explained at least 1% of the variation in the biomarkers in the overall model were retained for future models to derive the scores. This approach aims to limit the set of explanatory variables to those that are most important in predicting biomarker values, which facilitates the interpretation and application of the scores. All remaining RRR models were then performed in men (n=835) and women (n=1,164), separately. For both men and women, three different RRR models were performed corresponding to three different combinations of biomarkers used in the model: (1) poor glucose homeostasis and inflammation (all 3 biomarkers); (2) poor glucose homeostasis (C-peptide and HbA_{1c}); and (3) inflammation separately (hsCRP alone). The RRR model produces a factor score, which represents the overall correlation of all input variables to the biomarker(s) as a single continuous number.

Validity of the scores was examined among the remaining 499 men and women with biomarker data who were not used to derive the scores. Age- and sex-adjusted multivariate linear regression models were used to examine if the energy balance scores were associated with the biomarker or combination of biomarkers used in their derivation.

Outcome Ascertainment and Molecular Phenotypes

Self-reported cancer diagnoses from follow-up surveys were verified through medical records, registry linkage, or death certificates. A total of 2,228 incident cases of colorectal cancer (1,778 colon, and 450 rectum) were verified between 1999 and June 30, 2015. Among the cases, 1,108 were men and 1,120 were women. We collected formalin-fixed, paraffin-embedded tumor tissue specimens to assess molecular characteristics of the tumors in a sample of 627 cases (16). Polymerase chain reaction (PCR)-based assessment was used to categorize microsatellite instability (MSI) status among 409 tumors, and was based on the Bethesda Consensus Panel (17). Tumors were classified as MSI-high if ≥30% of the markers showed instability, and microsatellite stable (MSS)/MSI-low if <30% of the markers showed instability. Classification was based on ≥5 interpretable markers (unless all four markers were unstable, in which case the tumor was classified as MSI-high). CpG island methylator phenotype (CIMP) status was determined for 498 tumors using the eight-panel MethyLight array (18); tumors were classified as CIMP-high if ≥6 of the 8 genes had a percent of methylated reference (PMR) value ≥10, and non-CIMP if tumors had <6 genes with a PMR ≥10. *BRAF* mutation status c.1799T>A (p.V600E) was determined for 426 tumors using a fluorescent allele-specific PCR assay (19). Sanger sequencing was used to classify mutations in *KRAS* codons 12 and 13 for 352 tumors, and was considered mutated if any mutation was found in either codon (20).

Prospective Analysis

Energy balance scores for each CPS-II Nutrition Cohort study participant for whom we do not have biomarker data were computed using model-based parameters (i.e., RRR weights) in a way that mimics how the scores were calculated in the biomarker sub-sample. First, all lifestyle factors were centered and scaled, as is done at the onset of the RRR procedure. Then, each

lifestyle factor was multiplied by the RRR model weights (similar to a regression coefficient) produced in the RRR models when deriving the scores. The weighted scores for all components were summed to calculate a final energy balance score for each participant.

Hazard ratios (HR) and 95% confidence intervals (CI) for the associations of energy balance scores with colorectal cancer risk were computed using Cox proportional hazards regression among 114, 989 CPS-II Nutrition Cohort participants. Time-on-study was used for the time scale and was calculated from the date of completion of the 1999 survey until date of colorectal cancer diagnosis, date of death, date of last returned survey, or June 30, 2015, whichever came first. Each score was categorized into sex-specific quartiles; the reference group was the first quartile score representing the lifestyles correlated with the lowest levels of the energy balance biomarker(s). Trend tests were assessed by assigning the median value of the quartile as a continuous exposure variable. We also computed the HRs for a 1-standard deviation (SD) increase in the scores. All models were adjusted for age, sex, race/ethnicity, smoking status, alcohol intake, non-steroidal anti-inflammatory drug use, multivitamin use, and menopausal hormone therapy (women only). We examined associations stratified by age (<70, ≥ 70 years), sex, and anatomical subsite (colon, rectum). In sensitivity analyses we included adjustment for diabetes which was excluded from the main models as it may lie on the causal pathway between lifestyle and colorectal cancer. No violations of the proportional hazards assumption were observed using the Likelihood Ratio Test.

Cox models using the duplication method were used to examine potential heterogeneity for the association between the combined energy balance score on molecular subtypes of colorectal cancer (21). Molecular phenotype data of tumors is dependent upon the availability of tumor tissue and selection bias may be introduced into the HR estimates if the availability of

tumor tissue is dependent upon the colorectal cancer subtype (i.e., non-random missingness). Inverse probability weighting (IPW) was used to remediate this potential bias (22). Logistic regression was used among verified CPS-II colorectal cancer cases in the present analysis to calculate the probability of selection into the present study given their age at diagnosis, year of diagnosis, anatomical location of the tumor, sex, and stage. The inverse of that calculated probability served as the selection bias weight so colorectal cancer cases with phenotype data represents all colorectal cancer cases ascertained in the full analytic cohort of similar diagnoses.

Results

Mean (SD) levels of hsCRP, C-peptide, and HbA_{1c} in the 2,498 participants with biomarker data were 3,838.3 µg/L (5,013.7), 5.5 nmol/L (3.0), and 5.3% (0.7), respectively. The following factors explained >1% of the variation in the biomarkers and were used to derive the energy balance scores: BMI, MVPA, screen time, and servings/day of fruit juice, dark green vegetables, cruciferous vegetables, red/orange vegetables, whole grains, red meat, cured meat, organ meat, other fish (i.e., non-fried fish), eggs, high-fat dairy, oily fats, solid fats, and sugar-sweetened beverages.

The results from RRR and scoring weights used to calculate each of the sex-specific energy balance scores with different combinations of biomarkers are shown in **Table 1**. The highest weights for each energy balance score was with BMI (range from 0.20 to 0.38). The amount of variation in the biomarkers explained by the energy balance factors varied across the derived scores (**Table 1**). The largest amount of variation in the biomarkers explained by the derived energy balance scores was for the hsCRP score in women (10.5%). The smallest amount of variation in the biomarkers explained by the scores was for the HbA_{1c} + C-peptide score in

men (5.1%). All of the energy balance scores were strongly correlated (r>0.83; **Supplemental Table 1**). Multivariate linear regression results indicated that all energy balance scores were associated with the biomarker(s) used in their derivation among the validation subset of participants (**Supplemental Table 2**).

Descriptive characteristics across quartiles of the combined energy balance score are shown in **Table 2**. Participants in the fourth quartile compared to the first quartile of the score were less likely to be college educated, never smokers or to report multivitamin use, and more likely to report NSAID use, current menopausal hormone use, and a personal history of diabetes. The mean (SD) and median time from baseline to colorectal cancer diagnosis was 6.5 (4.1) and 6.0 years, respectively.

The associations of the three energy balance scores with risk of incident colorectal cancer are shown in **Table 3**. The higher risk observed in the fourth quartiles, compared to the first quartile, ranged from 30% for the HbA_{1c} + C-peptide lifestyle score to 35% for the hsCRP-alone score and the combined score based on all three biomarkers. All HR estimates from continuous models indicated a 10% higher risk of incident colorectal cancer per 1-SD increase.

Results from IPW-weighted models for associations between the combined energy balance score and molecular subtypes are shown in **Table 4**. Statistically significant heterogeneity was observed for the association when stratified by CIMP status of the tumor, where the fourth quartile was associated with a 58% higher risk compared to the first quartile for non-CIMP tumors (HR: 1.59; 95% CI: 1.17, 2.16) but not CIMP-positive tumors (HR: 0.72; 95% CI: 0.39, 1.30). Other statistically significant estimates were observed for the fourth quartiles when examining MSS/MSI-L (HR: 1.55; 95% CI: 1.10, 2.19), BRAF-wildtype (WT) (HR: 1.70; 95% CI: 1.21, 2.38), and for KRAS-mut tumors (HR: 2.00; 95% CI: 1.14, 3.50).

Supplemental Tables 3, 4, 5, and 6 show results for associations among strata of sex, age, and anatomical subsite, respectively. Stronger associations were observed in participants <70 years old, but otherwise little evidence of heterogeneity was observed. No substantive differences were observed after adjusting for self-reported diabetes.

Discussion

We empirically derived three energy balance scores based on circulating levels of hsCRP, C-peptide, and HbA_{1c}. All three scores were associated with higher risk of incident colorectal cancer in a large study population of predominantly non-Hispanic White men and women. These results indicate that men and women whose lifestyles reflect high potential for systemic inflammation and poor glucose homeostasis are at a higher subsequent risk of developing colorectal cancer. The relative role of excess body fat in poor metabolic health and subsequent colorectal cancer risk was evident by consistently high scoring weights for BMI. This study further supports long held hypotheses that systemic biomarkers of metabolic health mediate part of the relationship between several modifiable behaviors and colorectal cancer risk (23,24).

These biomarkers may reflect synergistic interactions in metabolic pathways that link unhealthy energy balance-related lifestyles to colorectal cancer risk. Pre-diagnostic levels of hsCRP levels, which has been used to evaluate a chronic inflammatory state, were positively associated with colorectal cancer risk in a meta-analysis of 18 studies (25). Pro-inflammatory conditions may promote tumor malignant progression, invasion, and metastasis (26). C-peptide is a marker of insulin production from the β-cells in the pancreas, uninfluenced by fasting status and with a longer half-life than insulin, and has been positively associated with colorectal cancer in multiple meta-analyses of prospective studies (27,28). HbA_{1c}, a stable indicator of circulating

glucose over the previous two-to-three months, also has been positively associated with colorectal cancer risk (28). Hyperglycemia may influence colorectal cancer etiology through multiple biologic mechanisms, such as through angiogenesis (29) or through mitogenic effects of insulin-like growth factor, among others (30).

Previous studies examining the relationship between lifestyle scores and colorectal cancer risk did not focus specifically on energy balance-related risk factors, nor were the scores derived based on empirical biomarker data; nonetheless, all reported statistically significant associations in the hypothesized direction (31-42). In the only other study that derived a score based on a biomarker (42), Tabung et al. developed a lifestyle score comprising BMI, physical activity, and 12 food groups based on circulating C-peptide concentrations (43). Similar to the combined energy balance score derived herein, positive weights were observed for BMI, solid fats, and fruit juice; a negative weight was observed for physical activity. In that study, the highest quintile of the score was associated with a 49% higher risk of colorectal cancer (CI: 1.10, 2.01), with no heterogeneity observed by sex (42). Differences observed in the current scores, such as weighting and direction of some food groups, may be explained by our use of multiple biomarkers of downstream effects of energy balance, not solely a marker of hyperinsulinemia. For example, the association of high fat dairy was stronger for hsCRP than with C-peptide and HbA_{1c}, which would tend to weaken the associations with all 3 biomarkers combined into one score. Nevertheless, BMI, physical activity and sugar-sweetened beverages were most strongly associated with the combined score compared to the other two scores, supporting energy balance as an important predictor of metabolic health. In contrast to the Tabung et al. score based on hyperinsulinemic potential, the present combined score was additionally based on inflammation and hyperglycemia, which may provide a more comprehensive characterization of poor energy

balance and colorectal cancer etiology. As other biologic pathways may connect energy balance-related factors to colorectal cancer risk (44,45), the scores in the present analysis may explain only a portion of the total effect.

It is possible the consistently stronger association observed among individuals <70 years old at baseline is explained by the slightly attenuated correlation between BMI and adiposity in older individuals (46), which subsequently limits our ability to estimate adipose-related inflammatory, hyperinsulinemic, hyperglycemic status in the empirical scores. Additionally, the relative contribution to these biomarkers from adipose tissue may be less in older individuals compared to age-related declines of metabolic function that are independent of adiposity (47,48).

This is the first study to examine an aggregate measure of energy balance-related factors in relation to molecular phenotypes of colorectal cancer, although our limited power made it difficult to examine associations for rarer sub-types of colorectal cancer and our ability to test heterogeneity was limited to large differences. Even with limited power, we observed a differential association of the combined energy balance score on CIMP status, with the association limited to non-CIMP tumors. There is evidence that genes associated with epigenetic silencing via methylation, such as SIRT1, have decreased expression in obesity resulting in lower levels of methylation (49). Non-CIMP tumors are also usually MSS (50) which have shown more consistent associations with excess adiposity (51,52). Although we did not observe a statistically significant association, we observed similar patterns of association for MSI status of tumors as we saw for CIMP status. Characterization of lifestyles that may promote the progression of certain mechanistic pathways indicative of molecular subtypes may be useful when monitoring patient risk profiles for personalized prevention, although more research in this area is needed.

Some limitations to our study should be noted. We did not have biomarker data on all participants, thus we can only hypothesize that the derived scores represent associations between energy balance-related factors and biomarkers in the entire cohort, as supported by the validation of the scores in a subset of participants with biomarker data. Further, we did not have sufficient power to perform a traditional mediation analysis. Self-reported exposure data may introduce misclassification bias, and there are known limitations in using BMI to assess adiposity (53). The study population comprised predominately older, non-Hispanic white participants, thus our results may not be generalizable to other age or racial/ethnic groups. Limited data on molecular subtypes did not allow for adequate power to examine associations in rarer molecular sub-types; future studies with larger numbers should consider this research question. The use of IPW accounts for non-random missing subtype data to help mitigate the role of selection bias in our models of molecular tumor phenotypes. Furthermore, energy balance scores and general lifestyle exposures did not previously differ across colorectal cancer cases with and without available tumor tissue (16). The term "energy balance score" was used given the stronger weighting of BMI, physical activity and the combination of fruit juice and sugar-sweetened beverages relative to other components in the scores; however, we recognize that non-energy balance-related pathways may also be involved. Additionally, the biomarkers in the present analysis are likely influenced by other external and inherited factors not included in our analysis, such as smoking with respect to inflammation. There are many strengths in our approach. We used data from a large, prospective study with detailed assessment on lifestyle factors and covariates. Use of RRR allowed for *a priori* identification of mechanistic pathways along with empirically based scoring. Further, data on established clinical markers that provide a comprehensive characterization of energy balance-related metabolic function were used to derive the scores.

In conclusion, the present analysis suggests that the clustering of energy balance-related lifestyle factors indicative of high levels of inflammation and poor glucose homeostasis are associated with higher risk of colorectal cancer. Focusing on energy balance-related factors that lower inflammation and ameliorate abnormal insulin/glucose levels may be effective methods for reducing risk of colorectal cancer, particularly for some molecular subtypes of colorectal cancer, and should be incorporated into public health recommendations. Future analyses should include other risk biomarkers and a more complete examination of associations of energy balance-related lifestyle factors with molecular subtypes of colorectal cancer in more populations with greater racial and ethnic diversity.

Acknowledgements

This work was supported by the American Cancer Society's Cancer Prevention Studies Postdoctoral Fellowship Program (M.G., E.R-P.) and by American Cancer Society funds for the creation, maintenance, and updating of the Cancer Prevention Study-II cohort. We sincerely appreciate all Cancer Prevention Study-II participants and each member of the study and biospecimen management group. We acknowledge the contributions to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries and cancer registries supported by the National Cancer Institute's SEER Program. The authors declare no conflicts of interest.

References

1. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ*, et al.* Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA: a cancer journal for clinicians **2018**;68(1):31-54 doi 10.3322/caac.21440.

2. Liu Y, Croft JB, Wheaton AG, Kanny D, Cunningham TJ, Lu H*, et al.* Clustering of Five Health-Related Behaviors for Chronic Disease Prevention Among Adults, United States, 2013. Preventing chronic disease **2016**;13:E70 doi 10.5888/pcd13.160054.
3. Kremers SPJ, De Bruijn G-J, Schaalma H, Brug J. Clustering of energy balance-related behaviours and their intrapersonal determinants. Psychology & Health **2004**;19(5):595-606 doi 10.1080/08870440412331279630.
4. Gunter MJ, Leitzmann MF. Obesity and colorectal cancer: epidemiology, mechanisms and candidate genes. The Journal of nutritional biochemistry **2006**;17(3):145-56 doi 10.1016/j.jnutbio.2005.06.011.
5. Continuous Update Project: Diet, Nutrition, Physical Activity and Colorectal Cancer. Available at: wcrf.org/colorectal-cancer-2017: World Cancer Research Fund International/American Institute for Cancer Research; 2017.
6. Hoffmann K, Schulze MB, Schienkiewitz A, Nöthlings U, Boeing H. Application of a New Statistical Method to Derive Dietary Patterns in Nutritional Epidemiology. American journal of epidemiology **2004**;159(10):935-44 doi 10.1093/aje/kwh134 %J American Journal of Epidemiology.
7. Guinter MA, McLain AC, Merchant AT, Sandler DP, Steck SE. A dietary pattern based on estrogen metabolism is associated with breast cancer risk in a prospective cohort of postmenopausal women. International journal of cancer **2018**;143(3):580-90 doi 10.1002/ijc.31387.
8. Tabung FK, Smith-Warner SA, Chavarro JE, Wu K, Fuchs CS, Hu FB*, et al.* Development and Validation of an Empirical Dietary Inflammatory Index. The Journal of nutrition **2016**;146(8):1560-70 doi 10.3945/jn.115.228718.
9. Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML*, et al.* The American Cancer Society cancer prevention study II nutrition cohort - Rationale, study design, and baseline characteristics. Cancer **2002**;94(9):2490-501 doi 10.1002/cncr.101970.abs.
10. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. Epidemiology (Cambridge, Mass) **1990**;1(1):58-64.
11. Flagg EW, Coates RJ, Calle EE, Potischman N, Thun MJ. Validation of the American Cancer Society Cancer Prevention Study II Nutrition Survey Cohort Food Frequency Questionnaire. Epidemiology (Cambridge, Mass) **2000**;11(4):462-8.
12. Healy GN, Clark BK, Winkler EA, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults' sedentary time in population-based studies. American journal of preventive medicine **2011**;41(2):216-27 doi 10.1016/j.amepre.2011.05.005.
13. Bonora E, Tuomilehto J. The Pros and Cons of Diagnosing Diabetes With A1C. **2011**;34(Supplement 2):S184-S90 doi 10.2337/dc11-s216 %J Diabetes Care.
14. Hope SV, Knight BA, Shields BM, Hattersley AT, McDonald TJ, Jones AG. Random non-fasting C-peptide: bringing robust assessment of endogenous insulin secretion to the clinic. Diabetic medicine : a journal of the British Diabetic Association **2016**;33(11):1554-8 doi 10.1111/dme.13142.
15. Musunuru K, Kral BG, Blumenthal RS, Fuster V, Campbell CY, Gluckman TJ*, et al.* The use of high-sensitivity assays for C-reactive protein in clinical practice. Nat Clin Pract Cardiovasc Med **2008**;5(10):621-35 doi 10.1038/ncpcardio1322.
16. Campbell PT, Deka A, Briggs P, Cicek M, Farris AB, Gaudet MM*, et al.* Establishment of the cancer prevention study II nutrition cohort colorectal tissue repository. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology **2014**;23(12):2694-702 doi 10.1158/1055-9965.Epi-14-0541.

17. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW*, et al.* A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res **1998**;58(22):5248-57.
18. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D*, et al.* MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res **2000**;28(8):E32.
19. Buchanan DD, Sweet K, Drini M, Jenkins MA, Win AK, English DR*, et al.* Risk factors for colorectal cancer in patients with multiple serrated polyps: a cross-sectional case series from genetics clinics. PLoS One **2010**;5(7):e11636 doi 10.1371/journal.pone.0011636.
20. Alsop K, Mead L, Smith LD, Royce SG, Tesoriero AA, Young JP*, et al.* Low somatic K-ras mutation frequency in colorectal cancer diagnosed under the age of 45 years. Eur J Cancer **2006**;42(10):1357-61 doi 10.1016/j.ejca.2006.02.023.
21. Wang M, Spiegelman D, Kuchiba A, Lochhead P, Kim S, Chan AT*, et al.* Statistical methods for studying disease subtype heterogeneity. Stat Med **2016**;35(5):782-800 doi 10.1002/sim.6793.
22. Liu L, Nevo D, Nishihara R, Cao Y, Song M, Twombly TS*, et al.* Utility of inverse probability weighting in molecular pathological epidemiology. European journal of epidemiology **2018**;33(4):381-92 doi 10.1007/s10654-017-0346-8.
23. Slattery ML, Potter J, Caan B, Edwards S, Coates A, Ma KN*, et al.* Energy balance and colon cancer - Beyond physical activity. Cancer Res **1997**;57(1):75-80.
24. Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. PHYSICAL-ACTIVITY, OBESITY, AND RISK FOR COLON-CANCER AND ADENOMA IN MEN. Ann Intern Med **1995**;122(5):327-34 doi 10.7326/0003-4819-122-5-199503010-00002.
25. Zhou B, Shu B, Yang J, Liu J, Xi T, Xing Y. C-reactive protein, interleukin-6 and the risk of colorectal cancer: a meta-analysis. Cancer causes & control : CCC **2014**;25(10):1397-405 doi 10.1007/s10552-014-0445-8.
26. Janakiram NB, Rao CV. The role of inflammation in colon cancer. Advances in experimental medicine and biology **2014**;816:25-52 doi 10.1007/978-3-0348-0837-8_2.
27. Chen L, Li L, Wang Y, Li P, Luo L, Yang B*, et al.* Circulating C-peptide level is a predictive factor for colorectal neoplasia: evidence from the meta-analysis of prospective studies. Cancer causes & control : CCC **2013**;24(10):1837-47 doi 10.1007/s10552-013-0261-6.
28. Xu J, Ye Y, Wu H, Duerksen-Hughes P, Zhang H, Li P*, et al.* Association between markers of glucose metabolism and risk of colorectal cancer. BMJ open **2016**;6(6):e011430 doi 10.1136/bmjopen-2016-011430.
29. Liu JJ, Druta M, Shibata D, Coppola D, Boler I, Elahi A*, et al.* Metabolic syndrome and colorectal cancer: is hyperinsulinemia/insulin receptor-mediated angiogenesis a critical process? Journal of geriatric oncology **2014**;5(1):40-8 doi 10.1016/j.jgo.2013.11.004.
30. Ma J, Pollak MN, Giovannucci E, Chan JM, Tao Y, Hennekens CH*, et al.* Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3. Journal of the National Cancer Institute **1999**;91(7):620-5.
31. Turati F, Bravi F, Di Maso M, Bosetti C, Polesel J, Serraino D*, et al.* Adherence to the World Cancer Research Fund/American Institute for Cancer Research recommendations and colorectal cancer risk. European journal of cancer (Oxford, England : 1990) **2017**;85:86-94 doi 10.1016/j.ejca.2017.08.015.
32. Thomson CA, McCullough ML, Wertheim BC, Chlebowski RT, Martinez ME, Stefanick ML*, et al.* Nutrition and physical activity cancer prevention guidelines, cancer risk, and mortality in the women's health initiative. Cancer prevention research (Philadelphia, Pa) **2014**;7(1):42-53 doi 10.1158/1940-6207.Capr-13-0258.

33. Romaguera D, Gracia-Lavedan E, Molinuevo A, de Batlle J, Mendez M, Moreno V*, et al.* Adherence to nutrition-based cancer prevention guidelines and breast, prostate and colorectal cancer risk in the MCC-Spain case-control study. International journal of cancer **2017**;141(1):83-93 doi 10.1002/ijc.30722.
34. Romaguera D, Vergnaud AC, Peeters PH, van Gils CH, Chan DS, Ferrari P*, et al.* Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? Results from the EPIC study. The American journal of clinical nutrition **2012**;96(1):150-63 doi 10.3945/ajcn.111.031674.
35. Platz EA, Willett WC, Colditz GA, Rimm EB, Spiegelman D, Giovannucci E. Proportion of colon cancer risk that might be preventable in a cohort of middle-aged US men. Cancer causes & control : CCC **2000**;11(7):579-88.
36. Odegaard AO, Koh WP, Yuan JM. Combined lifestyle factors and risk of incident colorectal cancer in a Chinese population. Cancer prevention research (Philadelphia, Pa) **2013**;6(4):360-7 doi 10.1158/1940-6207.Capr-12-0384.
37. Kirkegaard H, Johnsen NF, Christensen J, Frederiksen K, Overvad K, Tjonneland A. Association of adherence to lifestyle recommendations and risk of colorectal cancer: a prospective Danish cohort study. BMJ (Clinical research ed) **2010**;341:c5504 doi 10.1136/bmj.c5504.
38. Hang J, Cai B, Xue P, Wang L, Hu H, Zhou Y*, et al.* The Joint Effects of Lifestyle Factors and Comorbidities on the Risk of Colorectal Cancer: A Large Chinese Retrospective Case-Control Study. PloS one **2015**;10(12):e0143696 doi 10.1371/journal.pone.0143696.
39. Erdrich J, Zhang X, Giovannucci E, Willett W. Proportion of colon cancer attributable to lifestyle in a cohort of US women. Cancer causes & control : CCC **2015**;26(9):1271-9 doi 10.1007/s10552-015-0619-z.
40. Aleksandrova K, Pischon T, Jenab M, Bueno-de-Mesquita HB, Fedirko V, Norat T*, et al.* Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. BMC medicine **2014**;12:168 doi 10.1186/s12916-014-0168-4.
41. Kabat GC, Matthews CE, Kamensky V, Hollenbeck AR, Rohan TE. Adherence to cancer prevention guidelines and cancer incidence, cancer mortality, and total mortality: a prospective cohort study. The American journal of clinical nutrition **2015**;101(3):558-69 doi 10.3945/ajcn.114.094854.
42. Wang W, Fung TT, Wang M, Smith-Warner SA, Giovannucci EL, Tabung FK. Association of the Insulinemic Potential of Diet and Lifestyle With Risk of Digestive System Cancers in Men and Women. JNCI cancer spectrum **2018**;2(4):pky080 doi 10.1093/jncics/pky080.
43. Tabung FK, Wang W, Fung TT, Hu FB, Smith-Warner SA, Chavarro JE*, et al.* Development and validation of empirical indices to assess the insulinaemic potential of diet and lifestyle. The British journal of nutrition **2016**:1-12 doi 10.1017/s0007114516003755.
44. Aleman JO, Eusebi LH, Ricciardiello L, Patidar K, Sanyal AJ, Holt PR. Mechanisms of obesity-induced gastrointestinal neoplasia. Gastroenterology **2014**;146(2):357-73 doi 10.1053/j.gastro.2013.11.051.
45. Tamakoshi K, Toyoshima H, Wakai K, Kojima M, Suzuki K, Watanabe Y*, et al.* Leptin is associated with an increased female colorectal cancer risk: a nested case-control study in Japan. Oncology **2005**;68(4-6):454-61 doi 10.1159/000086988.
46. Flegal KM, Shepherd JA, Looker AC, Graubard BI, Borrud LG, Ogden CL*, et al.* Comparisons of percentage body fat, body mass index, waist circumference, and waist-stature ratio in adults. The American journal of clinical nutrition **2009**;89(2):500-8 doi 10.3945/ajcn.2008.26847.
47. Sanada F, Taniyama Y, Muratsu J, Otsu R, Shimizu H, Rakugi H*, et al.* Source of Chronic Inflammation in Aging. Front Cardiovasc Med **2018**;5:12- doi 10.3389/fcvm.2018.00012.

48. Reaven G. Age and Glucose Intolerance. Effect of fitness and fatness **2003**;26(2):539-40 doi 10.2337/diacare.26.2.539 %J Diabetes Care.
49. Hursting SD, Berger NA. Energy balance, host-related factors, and cancer progression. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2010**;28(26):4058-65 doi 10.1200/jco.2010.27.9935.
50. Ogino S, Kawasaki T, Kirkner GJ, Suemoto Y, Meyerhardt JA, Fuchs CS. Molecular correlates with MGMT promoter methylation and silencing support CpG island methylator phenotype-low (CIMP-low) in colorectal cancer. Gut **2007**;56(11):1564-71 doi 10.1136/gut.2007.119750.
51. Hughes LA, Williamson EJ, van Engeland M, Jenkins MA, Giles GG, Hopper JL*, et al.* Body size and risk for colorectal cancers showing BRAF mutations or microsatellite instability: a pooled analysis. International journal of epidemiology **2012**;41(4):1060-72 doi 10.1093/ije/dys055.
52. Campbell PT, Jacobs ET, Ulrich CM, Figueiredo JC, Poynter JN, McLaughlin JR*, et al.* Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumor microsatellite instability status. Journal of the National Cancer Institute **2010**;102(6):391-400 doi 10.1093/jnci/djq011.
53. Silva BR, Mialich MS, Hoffman DJ, Jordao AA. BMI, BMIfat, BAI or BAIFels - Which is the best adiposity index for the detection of excess weight? Nutricion hospitalaria **2017**;34(2):389-95 doi 10.20960/nh.366.

Tables

Table 1. Sex-specific reduced rank regression model weights used for the scoring of each energy balance score[a]

	All 3 Biomarkers		C-peptide + HbA1c		hsCRP	
	Male	Female	Male	Female	Male	Female
BMI	0.34	0.38	0.28	0.28	0.20	0.27
Physical activity	-0.10	-0.10	-0.08	-0.06	-0.06	-0.08
Screen time	0.02	0.04	0.01	0.02	0.03	0.04
Fruit juice	0.09	0.07	0.06	0.08	0.06	0.01
Dark green vegetables	-0.02	-0.01	-0.01	-0.04	-0.01	0.03
Cruciferous vegetables	0.03	-0.05	0.04	-0.01	0.01	-0.06
Red/orange vegetables	0.00	-0.04	0.04	-0.03	-0.05	-0.03
Whole grains	-0.07	0.00	-0.02	0.00	-0.07	-0.01
Red meat	0.00	-0.04	-0.02	-0.02	0.02	-0.04
Cured meat	-0.02	0.02	0.01	0.00	-0.03	0.03
Organ meat	0.02	0.04	0.05	0.02	-0.02	0.03
Other fish (i.e., non-fried)	-0.04	0.02	-0.02	0.02	-0.04	0.01
Eggs	0.04	0.01	0.01	0.01	0.05	0.01
High-fat dairy	0.06	0.01	-0.02	0.02	0.12	-0.01
Oil fats	-0.03	-0.04	-0.05	-0.06	0.02	0.00
Solid fats	-0.04	0.05	-0.05	0.05	-0.01	0.02
Sugar-sweetened beverages	0.07	0.04	0.06	0.02	0.04	0.04
Percent of variation in biomarker(s) explained[b]	5.2	7.1	5.1	5.8	8.2	10.5

BMI: body mass index; CRP: hsC-reactive protein; HbA1c: hemoglobinA1c

[a]Models for women also included current menopausal hormone therapy use. Weights represent coefficients for center and scaled input variables.

[b]Percentages represent the amount of variation in the biomarkers explained by the factor scores derived in the reduced rank regression models. Factor scores are a linear combination of the energy balance-related exposures that maximizes the explained variation in the biomarkers.

Table 2. Descriptive statistics stratified by quartiles of the comprehensive energy balance score

	1st	2nd	3rd	4th
n	28746	28748	28748	28747
Continuous[a]				
Age (years)	69.9 (6.28)	69.6 (6.14)	69.3 (6.02)	68.3 (5.88)
BMI (kg/m^2)	21.9 (2.53)	24.4 (2.0)	26.7 (2.02)	31.8 (4.41)
Physical activity (MVPA MET hrs/wk)	24 (20.25)	16.2 (13.73)	13.1 (11.77)	10.3 (10.45)
Screen time (min/wk)	512.3 (506.83)	592.1 (554.73)	646.5 (597.34)	756.7 (661.47)
Caloric intake (kcal/day)	1702.2 (544.95)	1681.9 (542.91)	1700.7 (558.52)	1781.7 (605.31)
Alcohol (drinks/day)	0.6 (1.0)	0.6 (1.0)	0.5 (0.98)	0.4 (0.97)
Categorical[b]				
Sex				
Female	56.2	56.2	56.2	56.2
Male	43.8	43.8	43.8	43.8
Education				
<High school degree	3.8	4.6	6.0	7.5
High school graduate	19.3	24.5	27.4	30.7
Some college	27.7	28.4	29.0	29.3
College graduate	48.6	41.9	37.0	31.8
Unknown/missing	0.6	0.5	0.6	0.7
Race/Ethnicity				
Non-Hispanic White	97.5	97.8	97.6	97.2
Non-Hispanic Black	0.7	0.9	1.3	1.8
Hispanic	0.4	0.4	0.4	0.4
Other/Unknown	1.4	0.9	0.7	0.6
Smoking Status				
Current	5.1	5.2	4.8	4.3
Former	50.6	51.8	52.3	53.8
Never	44.1	42.9	42.8	41.8
Missing	0.1	0.1	0.1	0.1
NSAID use				
No pills/month	40.3	37.4	36.3	35.5
1 to 14 pills/month	13.0	12.8	12.1	11.1
15 to 29 pills/month	7.8	7.7	7.2	6.4
30 to 59 pills/month	25.8	27.1	27.2	25.9
60+ pills/month	7.6	8.8	10.4	13.8
Unknown/missing	5.5	6.2	6.7	7.4
Multivitamin use				
Non-user	31.1	34.5	36.4	40.0
Non-daily user	7.5	8.0	7.8	7.4
Daily user	47.3	44.3	42.7	39.2

Unknown/missing	14.1	13.3	13.1	13.4
Comorbid diabetes				
No	94.4	93.7	91.9	87.5
Yes	5.6	6.3	8.1	12.5
Menopausal hormone therapy				
Current user	70.3	70.9	72.8	75.8
Non-user	28.4	27.7	25.8	22.9
Unknown/missing	1.3	1.4	1.4	1.3

BMI: body mass index; MET: metabolic equivalent of task; MVPA: moderate-to-vigorous intensity physical activity.
[a]Continuous variables expressed as mean (standard deviation)
[b]Categorical variables expressed as column percentage

Table 3. Relationship between empirically derived energy balance scores and colorectal cancer risk[a]

	Energy Balance Score Quartiles				Continuous Score[b]
	1st	2nd	3rd	4th	
All 3 biomarkers score					
No. of CRC cases	495	513	577	643	
HR (95% CI)	1.00 (ref)	1.03 (0.90 ,1.17)	1.14 (1.01 ,1.30)	1.35 (1.19 ,1.52)	1.10 (1.06 ,1.14)
p for trend	<0.0001				
HbA1c + C-peptide score					
No. of CRC cases	502	525	565	636	
HR (95% CI)	1.00 (ref)	1.02 (0.90 ,1.16)	1.10 (0.97 ,1.25)	1.30 (1.15 ,1.47)	1.10 (1.05 ,1.14)
p for trend	<0.0001				
hsCRP score					
No. of CRC cases	479	541	586	622	
HR (95% CI)	1.00 (ref)	1.12 (0.99 ,1.27)	1.21 (1.07 ,1.37)	1.35 (1.19 ,1.53)	1.10 (1.05 ,1.14)
p for trend	<0.0001				

CI: confidence interval; CRC: colorectal cancer; hsCRP: C-reactive protein; HbA1c: hemoglobin A1c; HR: hazard ratio

[a]Cox proportional hazards regression including multivariable adjustment for age, sex, race/ethnicity, NSAID use, multivitamin use, and MHT use.

[b]HRs shown for a 1-standard deviation increase in the respective score.

Table 4. Relationship between the combined energy balance score and molecular subtypes of colorectal cancer[a]

		Energy Balance Score Quartiles				Continuous Score[b]	p for heterogeneity[c]
		1st	2nd	3rd	4th		
MSI							0.17
High	cases	21	23	17	20		
	HR (95% CI)	1.00 (ref)	1.11 (0.58, 2.11)	0.82 (0.41, 1.65)	1.09 (0.58, 2.04)	0.98 (0.80, 1.20)	
Low/stable	cases	69	75	80	104		
	HR (95% CI)	1.00 (ref)	0.98 (0.68, 1.41)	0.97 (0.68, 1.40)	1.55 (1.10, 2.19)	1.15 (1.05, 1.25)	
CIMP							0.04
CIMP	cases	31	19	24	23		
	HR (95% CI)	1.00 (ref)	0.57 (0.30, 1.09)	0.70 (0.39, 1.25)	0.72 (0.39, 1.30)	0.89 (0.71, 1.12)	
Non-CIMP	cases	80	97	101	122		
	HR (95% CI)	1.00 (ref)	1.19 (0.86, 1.63)	1.19 (0.87, 1.63)	1.59 (1.17, 2.16)	1.15 (1.07, 1.23)	
BRAF							0.29
Mutant	cases	25	14	14	22		
	HR (95% CI)	1.00 (ref)	0.63 (0.31, 1.27)	0.63 (0.32, 1.26)	1.00 (0.57, 1.76)	1.01 (0.80, 1.27)	
Wild-type	cases	70	87	84	110		
	HR (95% CI)	1.00 (ref)	1.19 (0.84, 1.68)	1.03 (0.73, 1.47)	1.70 (1.21, 2.38)	1.15 (1.06, 1.24)	
KRAS							0.34
Mutant	cases	25	25	23	46		
	HR (95% CI)	1.00 (ref)	0.86 (0.47, 1.58)	0.70 (0.37, 1.34)	2.00 (1.14, 3.50)	1.19 (1.09, 1.30)	
Wild-type	cases	50	55	60	68		
	HR (95% CI)	1.00 (ref)	0.95 (0.63, 1.43)	1.12 (0.74, 1.67)	1.45 (0.98, 2.15)	1.12 (1.00, 1.24)	

CI: confidence interval; CIMP: CpG island methylator phenotype; CRC: colorectal cancer; HR: hazard ratio; MSI: microsatellite instability

[a]Cox proportional hazards regression including multivariable adjustment for age, sex, race/ethnicity, NSAID use, multivitamin use, and MHT use.

[b]HRs shown for a 1-standard deviation increase in the respective score.

[c]P-value from likelihood ratio test.