Document Cover page

Official Title of the study: Dietary Patterns, Metabolomics and Colorectal Cancer Risk and Mortality in the Nurses' Health Study and Health Professionals Follow-Up Study

NCT number: not available

Date of the document: originally created on October 1, 2015

History of changes:

C. APPROACH

C1. Aim 1 – K99 phase: To use 3 different approaches to derive 8 dietary patterns and utilize a standardized methodology to compare the best diet for overall health with the best diet for CRC prevention; and further determine if associations with CRC prevention are mediated by biological mechanisms involving inflammation and insulin

Rationale / Preliminary data

We have previously reported that several *a priori* (AHEI, AMED or OMNI) and *a posteriori* (Western or prudent) dietary patterns are associated with overall health (with all-cause mortality as indicator)[58,59] and with CRC prevention[12,60] in separate studies. The current proposal goes further to determine the extent to which diets beneficial for overall health are consistent with diets beneficial for CRC prevention. This is important for designing effective dietary guidelines for overall health that are optimized for CRC prevention, given that it is not feasible to have an optimal diet for every disease. We also increase the variety of dietary patterns by including two new patterns – the EDII and CDHR derived using a third and novel approach – the hypothesis-driven approach, that combines the strengths of both *a priori* and *a posteriori* approaches.

Dietary patterns research has overcome the limitations of single nutrient research to a large extent but not completely. Methodological issues have hindered a wider translation of findings from dietary patterns research, the major critique being lack of consistency and scientific rigor in the methods. In Aim 1a we will utilize a standardized methodology to evaluate the associations of the 8 dietary patterns with all-cause mortality as a marker for overall health and with CRC risk, adjusting for the same set of potential confounders, and further investigate in Aim 1b several inflammatory markers (CRP, IL6, TNFR, and adiponectin) and hyperinsulinemia (C-peptide) as mediators of the association between dietary patterns and CRC risk. This will provide an unprecedented opportunity to evaluate the consistency of evidence for the benefits of diet quality for overall health and for CRC prevention, supported by strong epidemiological methods and biological plausibility.

Methodological Approach

Study Population: The NHS was established in 1976 as a prospective cohort of 121,701 U.S. registered female nurses who were aged 30 to 55 years at enrollment.[61,62] The HPFS was initiated in 1986 as a prospective cohort of 51,529 U.S. male health professionals who were aged 40 to 75 years.[63] Blood samples were collected from subpopulations of the NHS (n=32,826) in 1989-1990 and HPFS (n=18,225) from 1993 to 1994, using similar protocols for all cohorts.[64,65]

Assessment of dietary data: Dietary data are updated every four years in the NHS (since 1980), NHS-II (since 1991) and in the HPFS (since 1986) with a validated semi quantitative food frequency questionnaire (FFQ) assessing habitual intake of foods and beverages over the preceding year. Participants reported how often, on average, they consumed each food of a standard portion size.[66-68] We will use average dietary data cumulatively across the multiple FFQs, to reduce within-subject variation and best represent long-term diet.[69,70]

Assessment of lifestyle factors: We will calculate cumulative body mass index (BMI), physical activity levels, and pack-years smoked.[71] Correlation between technician-measured and self-reported weight was r=0.96 in 1980 and 0.97 in 1986 and did not differ by level of BMI.[72] Our methods that measure physical activity and diet have been rigorously validated in our cohorts.[73-76] We have also routinely collected information regarding a CRC family history in a first-degree relative, post-menopausal hormone therapy, and aspirin use.[77,78] These long-term lifestyle data at multiple time points provide an unprecedented documentation of individual's lifestyle factors and will allow us to account for within-person variation in the effect of lifestyle on CRC risk.

Assessment of biomarkers: Procedures for the measurement of plasma inflammatory markers (IL6, CRP, TNFαR2, and adiponectin), [64,79,80] and plasma insulinemic markers (C-peptide, TG, and HDL)[80,81] have been described. The intra-assay coefficient of variation from blinded quality control (QC) samples ranged from 2.9% to 12.8% for IL6, 1.0% to 9.1% for CRP, 4.0% to 10.0% for TNFαR2, 8.1% to 11.1% for adiponectin, <12% for C-peptide and <1.8% for TG and HDL across batches. In the nested case-control studies in which these biomarkers were measured, samples from cases and their matched controls were analyzed in the same batch. QC samples were randomly interspersed among the case-control samples, and laboratory personnel were blinded to QC and case-control status for all assays. Our group has previously documented the stability of biomarkers in whole blood for 24 to 48 hours.[82] Participants who provided blood samples were cancer free and similar to their overall cohort with respect to BMI, parity, age at menarche, and past oral contraceptive use.[80]

Ascertainment of CRC incidence[83,97,98] and mortality:[80,81] Information on incident cancer and other disease

outcomes have been collected through participants, next-of-kin, or death certification. For any report of CRC, we seek written permission from study participants or next-of-kin to review their medical records. Study physicians review these records and successfully confirm 97% of reported CRC. The National Death Index is used to ascertain deaths and any diagnosis of CRC that contributed to death or was a secondary diagnosis.
Dietary patterns definitions: We will use 3 different approaches to define 8 dietary patterns: 4 *a priori* patterns (AHEI-2010, DASH, AMED, OMNI); 2 *a posteriori* patterns (Western and prudent); and 2 hypotheses-driven patterns (EDII, CDHR). ***1. AHEI-2010***: The 11 components of the AHEI-2010 include vegetables, fruits, whole grains, sugar-sweetened beverages and fruit juice, nuts and legumes, red/processed meat, trans fat, long chain (ω-3) fats, polyunsaturated fatty acids, sodium and alcohol. We will award points for intake of each dietary component on a scale from 0-10, with 10 indicating adherence, 0 no adherence and intermediate scores categorized proportionately.[13] ***2. DASH*:**[84,85] Several different DASH scores exist,[6] and we will use the one developed (DASH-Fung)[86] and applied by my mentors,[60] which is most commonly found in the literature with US populations. DASH-Fung scores 8 components (7 food groups and 1 nutrient)—each worth 5 points—for a total of 40 points.[86] ***3. AMED*:** we will use the AMED score that our group adapted for use in a US population.[59,87] The score rewards 1 point if intake is above the cohort-specific median for vegetables, legumes, fruit, nuts, whole-grain cereals, fish, and MUFAs:SFAs and 1 point for intake below the cohort-specific median for red and processed meats. In addition, alcohol intake of 5 to 10 g/d for women and 10 to 15 g/d for men received 1 point. The total score ranges between 0 and 9.[87,88] ***4.OMNI***: the OMNI diet is an improved DASH diet with partial substitution of carbohydrates with protein, about half from plant sources, or with unsaturated fat, predominantly monounsaturated fat.[89,90] ***5. Western pattern and 6. Prudent pattern***: we will derive these two *a posteriori* patterns using principal components analysis.[91-93] The factors will be rotated by an orthogonal transformation to achieve a simpler structure with greater interpretability, and the number of factors to retain will be determined using the diagram of eigenvalues.[91] The factor score for each pattern will be calculated by summing intakes of food groups weighted by their factor loadings. ***7. EDII and 8. CDHR***: I developed and validated these two patterns (described below) as part of my postdoctoral fellowship work. The manuscripts have been submitted.

Preliminary study 1 - Development and validation of an empirical index of dietary inflammatory potential (EDII): The goal for developing the **EDII** was to empirically create a score based on food groups to assess the overall inflammatory potential of whole diets. We first calculated average daily intakes (per 1000kcal) of 39 previously defined food groups[70] from the 1986 and 1990 FFQs in the NHS. We then applied reduced rank regression (RRR) to derive a dietary pattern predictive of three inflammatory biomarkers: CRP, TNFαR2 and IL6. RRR identifies linear functions of predictors (e.g., food groups) that simultaneously explain as much variation in the responses of interest (e.g., biomarkers) as possible. The first factor obtained by RRR was retained for subsequent analyses; (we called this the RRR dietary pattern). We then used stepwise linear regression analyses to identify the most important component food groups contributing to the RRR dietary pattern with the biomarker response score (RRR dietary pattern) as the dependent variable, the 39 food groups as independent variables, and a significance level of p=0.05 for entry into, and retention in the model. Intakes of the food groups identified in the stepwise linear regression analyses were weighted by the beta coefficients derived from the final stepwise regression model and then algebraically summed to constitute the EDII score. Higher (more positive) scores indicate more pro-inflammatory diets while lower (more negative) scores indicate anti-inflammatory diets. The EDII is comprised of 18 food groups; 9 are anti-inflammatory and 9 pro-inflammatory. In two independent samples, the NHS-II and HPFS, the EDII significantly

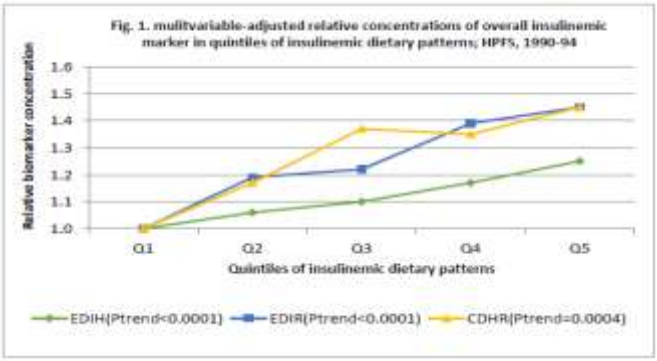
Table 2: Multivariable-adjusted relative concentrations (95% confidence intervals) of inflammatory markers across quintiles of the EDII in the NHS-II and HPFS

Quintile 1	Quintile 3	Quintile 5	Ptrend[1]	Quintile 1	Quintile 3	Quintile 5	Ptrend[1]
Nurses' Health Study-II (n=1,002); 1995-1999				Health Professionals Follow-up Study (n=2,633); 1990-1994			
IL6[2]							
reference	0.99 (0.86, 1.13)	1.13 (0.99, 1.30)	0.09	reference	1.04 (0.96, 1.16)	1.15 (1.05, 1.25)	<0.0001
CRP[2,3]							
reference	1.25 (0.97, 1.60)	1.43 (1.11, 1.84)	0.08	reference	1.23 (1.09, 1.39)	1.39 (1.23, 1.57)	<0.0001
TNFαR2[2]							
reference	1.06 (1.01, 1.10)	1.07 (1.03, 1.12)	0.04	reference	1.05 (1.02, 1.09)	1.09 (1.06, 1.13)	<0.0001
Adiponectin[2]							
reference	1.01 (0.92, 1.10)	0.88 (0.81, 0.97)	0.008	reference	0.95 (0.90, 1.00)	0.87 (0.82, 0.92)	<0.0001
Overall inflammatory marker score[3,4]							
reference	1.45 (0.89, 2.37)	2.69 (1.64, 4.41)	0.002	reference	1.61 (1.26, 2.07)	2.64 (2.05, 3.39)	<0.0001

[1]the p-value for trend is the p-value of the EDII as a continuous variable adjusted for all covariates listed in footnote #3; [2]values are ratio relative concentrations, comparing higher EDII quintiles to the lowest quintile (e.g., Q5/Q1), and back transformed (e^x) since biomarker data were log transformed using natural logarithms prior to analyses; [3]All fully adjusted models were adjusted for age, physical activity, smoking status, batch effects for biomarker measurements, regular aspirin/NSAIDs use and an inflammation-related chronic disease comorbidity score, and in NHS-II, additionally adjusted for menopausal status and postmenopausal hormone use. Chronic diseases/conditions included in the score were hypercholesterolemia, cancer, diabetes, high blood pressure, heart disease, and rheumatoid/other arthritis; [4]the overall inflammatory score was computed by summing z-scores of all four biomarkers for each participant

predicted concentrations of several biomarkers (IL6, CRP, TNFαR2, adiponectin and an overall inflammatory marker score as construct validators) (**Table 2**).

Preliminary study 2 - Development and validation of three indices of dietary insulinemic potential: We developed the three indices based on food groups, to assess the insulinemic potential of whole diets: 1) the empirical dietary index for hyperinsulinemia (EDIH), 2) the empirical dietary index for insulin resistance (EDIR) and 3) the composite dietary index for hyperinsulinemia and insulin resistance (**CDHR**). We used dietary and biomarker data (C-peptide, TG and HDL) in the NHS to develop the three indices, by fitting 2 separate stepwise linear regression models to identify the most important component food groups contributing to hyperinsulinemia (with C-peptide as the dependent variable) and to insulin resistance (ratio of TG/HDL as the dependent variable), the 39 food groups as independent variables, and a significance level of p=0.1 for entry into, and retention in the model. Intakes of the food groups identified in the stepwise linear regression analyses were weighted by the beta coefficients derived from the final stepwise regression model and then algebraically summed to constitute the scores. To create the composite score, we algebraically summed up the different components of both the EDIH and the EDIR to constitute the CDHR. Higher scores on all 3 indices indicate more highly insulinemic diets (hyperinsulinemia or insulin resistance) while lower scores indicate low insulinemic or insulin sensitive diets. The EDHR comprises 24 food groups; 12 are anti-insulinemic and 12 pro-insulinemic. In the HPFS, the EDII significantly predicted concentrations of the overall insulinemic marker score (combination of C-peptide and TG/HDL) as the construct validator (**Figure 1**) and of C-peptide and TG/HDL in separate models.



Statistical Analysis: The primary analysis for **Aim 1a** will prospectively assess associations between each of the 8 dietary patterns and all-cause mortality, and with incident CRC, using Cox proportional hazards regression, with separate models for men and women and in pooled analyses. For CRC, we will investigate associations by anatomic subsite (proximal colon, distal colon, rectum). After >30 years of follow-up, we have confirmed 31,286 all-cause deaths in the NHS and 12,555 in the HPFS.[94] Also, we have identified 1,764 CRC cases (823 proximal, 521 distal, and 379 rectal cancers) in NHS, and in the HPFS, we documented 1,640 CRC cases (640 proximal, 574 distal, and 426 rectal cancers).[95-97] Person-time (in months) for each participant will be calculated from the date of return of the baseline questionnaire in 1980 for women and 1986 for men, until the date of death from any cause or until CRC diagnosis, or June 1, 2012 whichever came first. Cumulatively updated average dietary pattern scores will be calculated using information from all available FFQs up to the 2-year interval prior to the follow-up cycle in which death or CRC diagnosis was confirmed. We will categorize the dietary patterns into quintiles and use the lowest quintile as reference, stratifying the Cox models on calendar year, age in months, to estimate hazard ratios (HR) and 95% confidence intervals (CI) to determine risk of dying from any cause and risk of developing CRC in higher quintiles of dietary patterns. Test for trend will be performed by modeling the dietary pattern scores as continuous variable and deriving the p-value for trend from the Wald test. In multivariable analyses, we will adjust for: BMI, physical activity, age, energy intake, race/ethnicity, educational status, smoking status, family history of CRC, NSAIDs use and additionally for menopausal status and postmenopausal hormone use in women. We will assess effect modification by BMI and physical activity, and incorporate 2-6 year time lags to reduce the potential for reverse causation. Based on the magnitude of risk explained, we will select the best diet in each of the three approaches (*a priori*, *a posteriori* and hypothesis-driven) and conduct a comparative analysis to derive the overall best diet for a healthful lifestyle that maximizes CRC prevention.

Aim 1b will be investigated using dietary and biomarker data from the two cohorts in a cross-sectional study. The 8 dietary patterns will be the main exposures while CRC diagnosis will be the outcome of interest. We will calculate dietary patterns from the 2 FFQs closest to the period of blood draw. That is, the 1986 and 1990

FFQs for the NHS, and the 1990 and 1994 FFQs for HPFS, averaging data across the 2 FFQs. We will then conduct mediation analyses, with the main objective of identifying the direct and indirect effects of each dietary pattern on CRC development, i.e., the effect of the dietary pattern that acts through the proposed mediators (CRP, TNFαR2, IL6, adiponectin, C-peptide) (indirect effect) and the effects that are unexplained by these same mediators (direct effects). The traditional approach to mediation analysis is based on adjusting for the mediator in standard regression models to estimate the direct effect, however, this approach may lead to flawed conclusions.[98] We will therefore use the regression-based approach described by VanderWeele and Vansteelandt[99] to estimate the direct and indirect effects of each potential mediator one at a time and cumulative effect for multiple mediators. This approach consists of fitting one model for the mediator, and another for the outcome. The approach has the advantage of assessing several mediators in the same model which would efficiently assess the potential interrelatedness of these mediators. We will use the SAS macro developed by the same authors[99] to implement the mediation analysis. All statistical tests will be two sided, and we will consider 95% confidence intervals not including one as statistically significant.

Expected outcomes for Aim 1 and impact on the field: To our knowledge, an evaluation of the consistency of evidence for the benefits of diet quality for overall health and for CRC prevention, based on a standardized methodology and plausible biological mechanisms has not previously been conducted. With the completion of this study, we anticipate an increased ability to widely translate findings from dietary patterns research and to design guidelines for healthful lifestyles that are optimized for CRC prevention.

C2. Aim 2 – R00 phase: To identify the set of metabolites predicted by both inflammatory diets and by circulating inflammatory markers and then evaluate the association of this unique inflammatory metabolome with CRC risk.

Aim 3 – R00 phase: To identify the set of metabolites predicted by both insulinemic diets and by circulating insulin markers and then evaluate the association of the unique insulinemic metabolome with CRC risk.

Rationale

Metabolomics, being the sum of the total metabolic processes in a cell, and situated downstream to all the other "omics,"[49,100] is uniquely suited to assess metabolic responses to dietary stimuli. Few studies have integrated metabolomics data in nutritional epidemiology and have determined the metabolic phenotype of behavioral/psychological dietary preferences.[49,50] The observed metabolic "fingerprint" or metabolome provides evidence for a link between specific dietary patterns and metabolic phenotypes. Additionally, no prior study has integrated dietary patterns with biomarker and metabolomics data to determine the metabolic phenotype of dietary behavior corroborated by relevant biomarkers. Integrating dietary, biomarker and metabolomics data provides an opportunity to derive unique metabolomes that may help to clearly differentiate between CRC cases and non-cases within the context of CRC early detection or diagnosis.

Methodological Approach

Assessment of metabolite data: plasma metabolites were measured as peak areas by a targeted LC-MS metabolomics platform directed by Dr. Clary Clish at the Broad Institute of the Massachusetts Institute of Technology and Harvard University (Cambridge, MA). Metabolite profiling methods were developed using reference standards of metabolites to determine chromatographic retention times, MS multiple reaction monitoring transitions, declustering potentials and collision energies.[101] In prior pilot work,[102] our group quantified 257 metabolites from archived plasma to evaluate metabolite interassay reproducibility, reproducibility with delayed processing, and within-person reproducibility over time. Interassay reproducibility was assessed with CVs from 60 duplicate plasma samples donated by participants in the NHS and HPFS, and 20 QC pool plasma replicates. CVs were <20% for 92% of metabolites and generally were similar by plasma anticoagulant type (heparin or EDTA) and fasting time. Metabolite reproducibility over a 24- to 48-h processing delay (n=48 samples) and within-person reproducibility over 1–2 years (n=80 samples) were assessed using

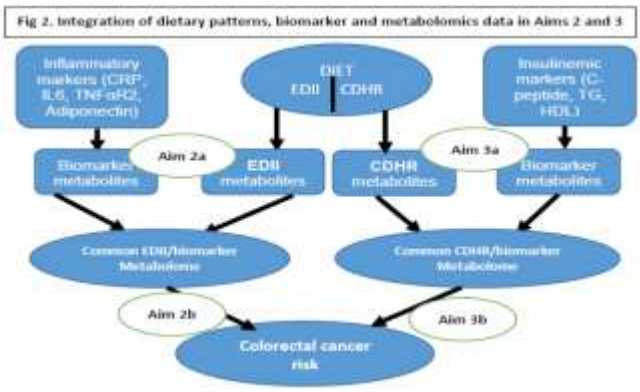
Spearman and intraclass correlation coefficients (ICCs). Approximately 75% of metabolites were reproducible over delays in processing of blood samples (Spearman or ICC ≥0.75, comparing immediate and 24-h delayed processing). Carbohydrates and purine/pyrimidine derivatives were most adversely affected by the processing delay; and 90% of metabolites were reproducible over 1–2 years within individuals (Spearman or ICC ≥0.4).[102] 174 metabolites passed these QC criteria and will be included in the analysis for Aims 2 and 3. We have used this platform in a nested case-control study of pancreatic cancer, with the novel finding that the 3 branched chain amino acids (isoleucine, leucine and valine) were associated with a higher risk of pancreatic cancer (RR, top vs. bottom quintile of sum of 3 amino acids=2.13; 95%CI=1.43-3.15), independent of other risk factors.[103]

Derivation of metabolomic fingerprint scores and statistical analysis*: (metabolomic fingerprint is used interchangeably with metabolomic signature, profile, or pattern).* **Table 3** lists the sample sizes for participants with biomarker (CRP, TNFαR2, IL6, adiponectin, and C-peptide) data from several matched case-control studies, nested within the NHS and HPFS. Participants were matched on cohort, sex, year of birth, smoking status (never, former, current), and fasting status (<8 hours, ≥8 hours). In addition to the two dietary pattern scores (EDII and CDHR), we will create two different overall biomarker scores for each participant as follows:

Overall inflammatory marker score (infMS) = z-score (logIL6) + z-score (logCRP) + z-score (log TNFαR2) – z-score (logAdiponectin);

Overall insulinemic marker score (insMS) = z-score (logC-peptide) + z-score (logTG/logHDL).

Prior to analysis, concentrations of all metabolites will be log transformed and values normalized to have mean: 0, standard deviation: ±1. The two dietary pattern scores (EDII and CDHR) and the two overall biomarker scores (infMS and insMS) will be trichotomized into tertiles. All data will then be merged and participants randomized to either the training dataset or the validation dataset in a 2 to 1 manner. We will use the training dataset to develop the metabolomic fingerprint scores, and then assess their validity in the validation dataset. We will construct four separate multivariable-adjusted Full Canonical Discriminant Analysis models, to create the metabolomic fingerprint scores, with all 174 metabolites as dependent variables and the highest and lowest tertiles of the dietary pattern scores and two overall biomarker scores as independent variables (one in each of the four models). To maximize the degree of discrimination, we will exclude participants in the middle tertiles. The canonical coefficients from the principal components' eigenvector encapsulate the greatest degree of discrimination. As such, all 174 metabolites contribute to the computation of the Full Canonical Variable. The metabolomic fingerprint score will include the 20 metabolites with the largest canonical discriminant coefficients (10 largest negative and 10 largest positive), identified by the full canonical discriminant analysis. Ten was determined as the cutoff for inclusion in the metabolomic fingerprint score as we deemed that the identification of 20 (potentially related) metabolites would be sufficient for biological interpretation of the results. The multivariable-adjusted models will include case/control status, cohort, quintiles of caloric intake, smoking status, and age at blood collection. We will not adjust for BMI or physical activity because we showed in **preliminary studies 1 and 2** that these variables played a more mediating or modification role than a confounding role in the association between inflammatory and insulinemic dietary patterns and respective biomarkers (*Tabung et al, submitted*). We will therefore assess effect modification by BMI and physical activity.



Equation 1. Computation of the Metabolite Fingerprint Score

$$\text{Metabolomic fingerprint score} = \sum_{k=1}^{10} hi - \sum_{k=1}^{10} li$$

h=peak area of metabolites with highest canonical discriminant coefficient
l=peak area of metabolites with lowest canonical discriminant coefficient

To evaluate the performance of the discrimnatory variables (i.e., the 4 metabolomic fingerprint scores), we will compute them in the validation dataset, and then conduct four separate un-adjusted and multivariable-adjusted ANCOVAs for each discriminatory variable, with tertiles of each of EDII, CDHR, infMS and insMS as the

independent variable in each validation model. Participants in the middle tertiles will be included. We will then analyze least square means ± SEM of the metabolomic fingerprint scores among tertiles of the independent variables, adjusting for multiple comparisons using the Bonferroni approach (0.05/3=0.017).

Matched nested case-control study to evaluate CRC risk: Metabolomic assays are currently being conducted for 1,056 CRC cases and 1,056 matched controls for the 174 identified metabolites as part of funded grants (NHS: P01 CA 087969, CA 186107; HPFS: CA 167552) (see assessment of metabolite data). We will compute the 4 validated metabolomic fingerprint scores and the 2 common metabolome profiles developed and categorize into quintiles. The date of blood donation will be used as baseline for follow-up (1989-90 for NHS and 1993-94 for HPFS). We will employ multivariable conditional logistic regression models to determine the risk of CRC in quintiles of the 2 common metabolome profiles (**Fig. 2**), using the lowest quintile as reference. In secondary analyses, we will utilize each of the 4 metabolomic fingerprint scores in the models to evaluate CRC risk and compare findings with the main analyses (2 common metabolome profiles).

Table 3. Study design, sample sizes (NHS + HPFS) and minimum detectable associations

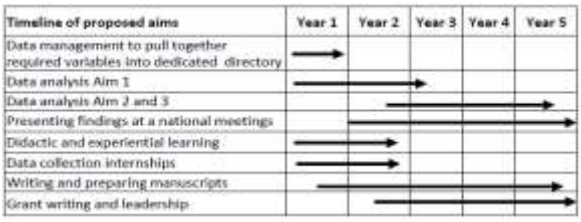
Hypothesis	Outcome	Case numbers	Min. detect. association
1a	All-cause deaths	43,841	1.04
	CRC	3,404	1.12
	Colon	2,599	1.22
	Colon proximal	1,463	1.21
	Colon distal	1,095	1.18
	Rectal	805	1.33
1b	CRP	909/1,365	1.20 – 1.22
	IL6	882/1,345	1.25 – 1.17
	TNFaR2	882/1,345	1.02 – 1.10
	Adiponectin	1179/1,642	0.94 – 0.98
	C-peptide	666/881	1.22 – 1.27
2a	EDII - Metabolome	6,490	0.06
	Invfla - metabolome	7,868	0.04
2b	CRC	1056/1056	1.08
3a	CDHR - Metabolome	6,490	0.05
	Insul - metabolome	1,275	0.04
3b	CRC	1056/1056	1.08

For 1a, numbers are cases only; for 1b, 2b and 3b numbers are CRC cases/controls; associations are RRs for 1a, 1b, 2b and 3b; and metabolomic fingerprint difference scores (validation dataset) for 2a and 3a; for 1b, the 1st RR is for the dietary pattern- biomarker association and the 2nd is for the biomarker-CRC association;

Expected outcomes for aims 2 and 3 and impact on the field: We anticipate identifying key metabolic pathways in CRC based on unique metabolomes that are the joint effect of dietary preferences and circulating biomarkers. In addition, we expect to find metabolites of uncertain, and possible pathogenic significance related to CRC risk. Therefore we will uncover potentially actionable metabolomes and specific metabolites for prevention and early detection. Findings from these aims will identify the biological mechanisms by which dietary patterns relate to CRC risk, and along with our findings from Aim 1, we will develop a more comprehensive understanding of how to improve the efficiency of dietary recommendations optimized for CRC prevention.

Sample size and statistical power considerations (**all Aims**): **Table 3** illustrates case counts up to 2012 for all-cause death and CRC in NHS and HPFS, including the magnitude of association that can be achieved with 80% power given the sample sizes for the various hypotheses.[104-106] We will thus have ample power to detect significant associations. We also anticipate large precision (narrow 95%CIs) in these analyses.

Limitations and alternative strategies (**all Aims**): *Bias due to incomplete follow-up is possible*. Thus we devote great effort to follow-up, with 93% cumulative follow-up. Differential CRC screening intensity could create bias, but we do not expect endoscopy rates to vary appreciably by exposure status. *Misclassification of exposures* is always present to some degree. We have documented extensively the validity of our dietary questionnaire, which is further enhanced by repeated administrations. The data from our validation/calibration studies and studies of within-person variation of biomarkers allow us to adjust RRs and CIs for measurement error using methods we have developed.[107-110] We also conducted extensive pilot studies to assess metabolites based on our *a priori* QC criteria, and excluded metabolites that did not pass these criteria.[102] *Confounding* due to unmeasured or imperfectly measured covariates can be present. Thus, we strive to assess any plausible confounding variables and will conduct extensive multivariable-adjusted analyses. *Representativeness of a single blood sample* to reflect long-term levels is a potential concern. However, our data above indicate that the ICCs over 2-3 years for most proposed metabolites are 0.48-0.72.



Summary and future research directions: Despite advances in screening, CRC remains the third most commonly diagnosed cancer and the second-leading cause of cancer death in the US. In this unique cohort with data on dietary and lifestyle factors and biological specimens, we will use an innovative, integrated

approach to evaluate hypotheses related to inflammation-mediated, insulin-mediated and metabolic signaling pathways. The findings will elucidate the etiology of CRC and identify potential strategies for prevention, and early detection. In addition, the cross-cutting nature of our hypotheses across aims will allow for a productive synergy and further our understanding of CRC etiology. Dr. Tabung will acquire an important set of skills necessary to achieve his long term goal of establishing an independent interdisciplinary research program to elucidate the role of diet in cancer prevention and control, including the identification of plausible underlying biological mechanisms. We expect to identify metabolomes or specific metabolites generated by diet, inflammation, insulin or their complex interactions, which are up- or downregulated in specific biological pathways and that may be important to further investigate in the context of CRC prevention and control. These data will form the basis of my future R01 in Year 4.

LIST OF REFERENCES CITED

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer Statistics, 2014. *CA: A Cancer Journal for Clinicians* 2014;**64**(1):9-29.

2. Harvey AE, Lashinger LM, Hursting SD. The growing challenge of obesity and cancer: an inflammatory issue. *Ann N Y Acad Sci* 2011;**1229**:45-52.

3. Lumeng CN, Saltiel AR. Inflammatory links between obesity and metabolic disease. *J Clin Invest* 2011;**121**(6):2111-7.

4. Chan AT, Giovannucci EL. Primary prevention of colorectal cancer. *Gastroenterology* 2010;**138**(6):2029-2043 e10.

5. Miller PE, Lazarus P, Lesko SM, Muscat JE, Harper G, Cross AJ, Sinha R, Ryczak K, Escobar G, Mauger DT, Hartman TJ. Diet Index-Based and Empirically Derived Dietary Patterns are Associated with Colorectal Cancer Risk. *The Journal of Nutrition* 2010;**140**(7):1267-1273.

6. Miller PE, Cross AJ, Subar AF, Krebs-Smith SM, Park Y, Powell-Wiley T, Hollenbeck A, Reedy J. Comparison of 4 Established DASH Diet Indexes: Examining Associations of Index Scores and Colorectal Cancer. *The American Journal of Clinical Nutrition* 2013;**98**(3):794-803.

7. Flood A, Rastogi T, Wirfält E, Mitrou PN, Reedy J, Subar AF, Kipnis V, Mouw T, Hollenbeck AR, Leitzmann M, Schatzkin A. Dietary patterns as identified by factor analysis and colorectal cancer among middle-aged Americans. *The American Journal of Clinical Nutrition* 2008;**88**(1):176-184.

8. Reedy J, Mitrou PN, Krebs-Smith SM, Wirfält E, Flood A, Kipnis V, Leitzmann M, Mouw T, Hollenbeck A, Schatzkin A, Subar AF. Index-based Dietary Patterns and Risk of Colorectal Cancer: The NIH-AARP Diet and Health Study. *American Journal of Epidemiology* 2008;**168**(1):38-48.

9. Yusof A, Isa ZM, Shah SA. Dietary Patterns and Risk of Colorectal Cancer: A Systematic Review of Cohort Studies (2000-2011). *Asian Pacific Journal of Cancer Prevention* 2012;**13**(9):4713-7.

10. Randi G, Edefonti V, Ferraroni M, La Vecchia C, Decarli A. Dietary Patterns and the Risk of Colorectal Cancer and Adenomas. *Nutrition Reviews* 2010;**68**(7):389-408.

11. Fung T, Hu F.B., Fuchs C., Giovannucci E., Hunter D.J., Stampfer M.J., Colditz G.A., Willett W.C. Major dietary patterns and the risk of colorectal cancer in women. 2003;**163**(3):309-314.

12. Fung T, Brown LS. Dietary Patterns and the Risk of Colorectal Cancer. *Current Nutrition Reports* 2013;**2**(148-55).

13. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, Stampfer MJ, Willett WC. Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease. *The Journal of Nutrition* 2012;**142**(6):1009-1018.

14. World Cancer Research Fund and the American Institute for Cancer Research. Food, Nutrition and Prevention of Cancer: A Global Perspective *Washington, DC: American Institute for Cancer Research*, 2007.

15. Varraso R, Garcia-Aymerich J, Monier F, Le Moual N, De Batlle J, Miranda G, Pison C, Romieu I, Kauffmann F, Maccario J. Assessment of Dietary Patterns in Nutritional Epidemiology: Principal

Component Analysis Compared with Confirmatory Factor Analysis. *The American Journal of Clinical Nutrition* 2012;**96**(5):1079-1092.

16. Moeller SM, Reedy J, Millen AE, Dixon LB, Newby PK, Tucker KL, Krebs-Smith SM, Guenther PM. Dietary Patterns: Challenges and Opportunities in Dietary Patterns Research: An Experimental Biology Workshop, April 1, 2006. *Journal of the American Dietetic Association* 2007;**107**(7):1233-1239.

17. Schulze MB, Hoffmann K, Kroke A, Boeing H. An approach to construct simplified measures of dietary patterns from exploratory factor analysis. *British Journal of Nutrition* 2003;**89**(03):409-418.

18. Tabung FK, Steck SE, Su LJ, Mohler JL, Fontham ET, Bensen JT, Hebert JR, Zhang H, Arab L. Intake of grains and dietary fiber and prostate cancer aggressiveness by race. *Prostate Cancer* 2012;**2012**(Article ID 323296):10 pages.

19. Tabung FK, Steck SE, Burch JB, Chen CF, Zhang H, Hurley TG, Cavicchia P, Alexander M, Shivappa N, Creek KE, Lloyd SC, Hebert JR. A healthy lifestyle index is associated with reduced risk of colorectal adenomatous polyps among non-users of non-steroidal anti-Inflammatory drugs. *The Journal of Primary Prevention* 2015;**36**(1):21-31.

20. Tabung FK, Steck SE, Ma Y, Liese AD, Zhang J, Caan B, Hou L, Johnson KC, Mossavar-Rahmani Y, Shivappa N, Wactawski-Wende J, Ockene JK, Hebert JR. The association between dietary inflammatory index and risk of colorectal cancer among postmenopausal women: results from the Women's Health Initiative. *Cancer Causes & Control* 2015;**26**(3):399-408.

21. Tabung FK, Steck SE, Zhang J, Ma Y, Liese AD, Agalliu I, Hou L, Hurley TG, Hingle M, Jiao L, Martin LW, Millen EA, Park HL, Rosal CM, Shikany JM, Shivappa N, Ockene JK, Hebert JR. Construct validation of the dietary inflammatory index among postmenopausal women. *Annals of Epidemiology* 2015;**25**(6):398-405.

22. Tabung FK, Daguise V, Lydiard D, Steck SE. An integrated approach to addressing chronic disease risk factors in financially disadvantaged women in South Carolina *American Journal of Health Promotion* 2015(In Press (doi: 10.4278/ajhp.141010-QUAN-512)).

23. Steck SE, Shivappa N, Tabung FK, Harmon BE, Wirth MD, Hurley TG, Hebert JR. The dietary inflammatory index: a new tool for assessing diet quality based on inflammatory potential. *The Digest* 2014;**49**(3):1-9.

24. Shivappa N, Steck SE, Hurley TG, Hussey JR, Ma Y, Ockene IS, Tabung FK, Hebert JR. A population-based dietary inflammatory index predicts levels of C-reactive protein in the Seasonal Variation of Blood Cholesterol Study (SEASONS). *Public Health Nutrition* 2013;**17**(8):1825-33.

25. Hébert JR, Hurley TG, Steck SE, Miller DR, Tabung FK, Peterson KE, Kushi LH, Frongillo EA. Considering the value of dietary assessment data in informing nutrition-related health policy. *Advances in Nutrition: An International Review Journal* 2014;**5**(4):447-455.

26. Hébert JR, Hurley TG, Steck SE, Miller DR, Tabung FK, Kushi LH, Frongillo EA. Reply to Archer E. and Blair SN: Implausible data, false memories, and the status quo in dietary assessment. *Advances in Nutrition: An International Review Journal* 2015;**6**(2):230-233.

27. Cerhan JR BS, Vijai J, Ghesquières H, McKay J, Wang SS, Wang Z, Yeager M, Conde L, de Bakker PI, Nieters A, Cox D, Burdett L, Monnereau A, Flowers CR, De Roos AJ, Brooks-Wilson AR, Lan Q, Severi G, Melbye M, Gu J, Jackson RD, Kane E, Teras LR, Purdue MP, Vajdic CM, Spinelli JJ, Giles

GG, Albanes D, Kelly RS, Zucca M, Bertrand KA, Zeleniuch-Jacquotte A, Lawrence C, Hutchinson A, Zhi D, Habermann TM, Link BK, Novak AJ, Dogan A, Asmann YW, Liebow M, Thompson CA, Ansell SM, Witzig TE, Weiner GJ, Veron AS, Zelenika D, Tilly H, Haioun C, Molina TJ, Hjalgrim H, Glimelius B, Adami HO, Bracci PM, Riby J, Smith MT, Holly EA, Cozen W, Hartge P, Morton LM, Severson RK, Tinker LF, North KE, Becker N, Benavente Y, Boffetta P, Brennan P, Foretova L, Maynadie M, Staines A, Lightfoot T, Crouch S, Smith A, Roman E, Diver WR, Offit K, Zelenetz A, Klein RJ, Villano DJ, Zheng T, Zhang Y, Holford TR, Kricker A, Turner J, Southey MC, Clavel J, Virtamo J, Weinstein S, Riboli E, Vineis P, Kaaks R, Trichopoulos D, Vermeulen RC, Boeing H, Tjonneland A, Angelucci E, Di Lollo S, Rais M, Birmann BM, Laden F, Giovannucci E, Kraft P, Huang J, Ma B, Ye Y, Chiu BC, Sampson J, Liang L, Park JH, Chung CC, Weisenburger DD, Chatterjee N, Fraumeni JF Jr, Slager SL, Wu X, de Sanjose S, Smedby KE, Salles G, Skibola CF, Rothman N, Chanock SJ. Genome-wide association study identifies multiple susceptibility loci for diffuse large B-cell lymphoma. *Nature genetics* 2014;**46**(11):1233-1238.

28. Fung TT, Hu FB, Schulze M, Pollak M, Wu T, Fuchs CS, Giovannucci E. A dietary pattern that is associated with C-peptide and risk of colorectal cancer in women. *Cancer Causes & Control* 2012;**23**(6):959-965.

29. Vijai J WZ, Berndt SI, Skibola CF, Slager SL, de Sanjose S, Melbye M, Glimelius B, Bracci PM, Conde L, Birmann BM, Wang SS, Brooks-Wilson AR, Lan Q, de Bakker PI, Vermeulen RC, Portlock C, Ansell SM, Link BK, Riby J, North KE, Gu J, Hjalgrim H, Cozen W, Becker N, Teras LR, Spinelli JJ, Turner J, Zhang Y, Purdue MP, Giles GG, Kelly RS, Zeleniuch-Jacquotte A, Ennas MG, Monnereau A, Bertrand KA, Albanes D, Lightfoot T, Yeager M, Chung CC, Burdett L, Hutchinson A, Lawrence C, Montalvan R, Liang L, Huang J, Ma B, Villano DJ, Maria A, Corines M, Thomas T, Novak AJ, Dogan A, Liebow M, Thompson CA, Witzig TE, Habermann TM, Weiner GJ, Smith MT, Holly EA, Jackson RD, Tinker LF, Ye Y, Adami HO, Smedby KE, De Roos AJ, Hartge P, Morton LM, Severson RK, Benavente Y, Boffetta P, Brennan P, Foretova L, Maynadie M, McKay J, Staines A, Diver WR, Vajdic CM, Armstrong BK, Kricker A, Zheng T, Holford TR, Severi G, Vineis P, Ferri GM, Ricco R, Miligi L, Clavel J, Giovannucci E, Kraft P, Virtamo J, Smith A, Kane E, Roman E, Chiu BC, Fraumeni JF, Wu X, Cerhan JR, Offit K, Chanock SJ, Rothman N, Nieters A. A genome-wide association study of marginal zone lymphoma shows association to the HLA region. *Nature Communications* 2015;**6**:5751.

30. Giovannucci E. Insulin and colon cancer. *Cancer Causes and Control* 1995;**6**(2):164-179.

31. Giovannucci E. Modifiable risk factors for colon cancer. *Gastroenterol Clin North Am.* 2002;**31**(4):925-43.

32. Hu FB. Dietary Pattern Analysis: A New Direction in Nutritional Epidemiology. *Current Opinion in Lipidology* 2002;**13**(1):3-9.

33. Bruce W.R WT, Giacca A. Mechanisms Linking Diet and Colorectal Cancer: The Possible Role of Insulin Resistance. *Nutrition and Cancer* 2000;**37**(1):19-26.

34. Liese AD, Krebs-Smith S.M, Subar A.F, George S.M, Harmon B.E, Neuhouser M.L, Boushey C.J, Schap T.E, Reedy J. The Dietary Patterns Methods Project: Synthesis of Findings across Cohorts and Relevance to Dietary Guidance. *The Journal of Nutrition* 2015;**145**(3):393-402.

35. Harmon BE, Boushey CJ, Shvetsov YB, Ettienne R, Reedy J, Wilkens LR, Le Marchand L, Henderson BE, Kolonel LN. Associations of key diet-quality indexes with mortality in the Multiethnic Cohort: the Dietary Patterns Methods Project. *The American Journal of Clinical Nutrition* 2015.

36. Reedy J, Krebs-Smith SM, Miller PE, Liese AD, Kahle LL, Park Y, Subar AF. Higher Diet Quality is Associated with Decreased Risk of All-Cause, Cardiovascular Disease, and Cancer Mortality among Older Adults. *The Journal of Nutrition* 2014.

37. George S, Ballard-Barbash R, Manson JE, Reedy J, Shikany JM, Subar AF, Tinker LF, Vitolins M, Neuhouser ML. Comparing Indices of Diet Quality With Chronic Disease Mortality Risk in Postmenopausal Women in the Women's Health Initiative Observational Study: Evidence to Inform National Dietary Guidance. *American Journal of Epidemiology* 2014;**180**(6):616-625.

38. Russell J, Flood V, Rochtchina E, Gopinath B, Allman-Farinelli M, Bauman A, Mitchell P. Adherence to dietary guidelines and 15-year risk of all-cause mortality. *British Journal of Nutrition* 2013;**109**(03):547-555.

39. Miller PE, Lesko SM, Muscat JE, Lazarus P, Hartman TJ. Dietary Patterns and Colorectal Adenoma and Cancer Risk: A Review of the Epidemiological Evidence. *Nutrition & Cancer* 2010;**62**(4):413-424.

40. Chan AT, Ogino Shuji, Giovannucci Edward L., Fuchs, Charles S. Inflammatory Markers Are Associated With Risk of Colorectal Cancer and Chemopreventive Response to Anti-Inflammatory Drugs. *Gastroenterology* 2011;**140**(3):799-808.e2.

41. 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.

42. Aleksandrova K, Jenab M, Boeing H, Jansen E, Bueno-de-Mesquita HB, Rinaldi S, Riboli E, Overvad K, Dahm CC, Olsen A, Tjønneland A, Boutron-Ruault MC, Clavel-Chapelon F, Morois S, Palli D, Krogh V, Tumino R, Vineis P, Panico S, Kaaks R, Rohrmann S, Trichopoulou A, Lagiou P, Trichopoulos D, van Duijnhoven FJ, Leufkens AM, Peeters PH, Rodríguez L, Bonet C, Sánchez MJ, Dorronsoro M, Navarro C, Barricarte A, Palmqvist R, Hallmans G, Khaw KT, Wareham N, Allen NE, Spencer E, Romaguera D, Norat T, Pischon TPischon, Tobias. Circulating C-Reactive Protein Concentrations and Risks of Colon and Rectal Cancer: A Nested Case-Control Study Within the European Prospective Investigation into Cancer and Nutrition. *American Journal of Epidemiology* 2010;**172**(4):407-418.

43. Song M, Zhang X, Wu K, Ogino S, Fuchs CS, Giovannucci EL, Chan AT. Plasma Adiponectin and Soluble Leptin Receptor and Risk of Colorectal Cancer: A Prospective Study. *Cancer Prevention Research* 2013;**6**(9):875-885.

44. Macarthur M, Hold GL, El-Omar EM. Inflammation and Cancer II. Role of Chronic Inflammation and Cytokine Gene Polymorphisms in the Pathogenesis of Gastrointestinal Malignancy. *American Journal of Physiology - Gastrointestinal and Liver Physiology* 2004;**286**(4):G515-G520.

45. Yoon Y, Keum N, Zhang X, Cho E, Giovannucci EL. Hyperinsulinemia, insulin resistance and colorectal adenomas: A meta-analysis. *Metabolism - Clinical and Experimental* 2015;**In Press**.

46. Giovannucci E. Metabolic syndrome, hyperinsulinemia, and colon cancer: a review. *American Journal of Clinical Nutrition* 2007;**86**(3):s836-42.

47. Aleksandrova K, Jenab M, Bueno-de-Mesquita HB, Fedirko V, Kaaks R, Lukanova A, van Duijnhoven FJ, Jansen E, Rinaldi S, Romieu I, Ferrari P, Murphy N, Gunter MJ, Riboli E, Westhpal S, Overvad K, Tjønneland A, Halkjær J, Boutron-Ruault MC, Dossus L, Racine A, Trichopoulou A, Bamia C, Orfanos P, Agnoli C, Palli D, Panico S, Tumino R, Vineis P, Peeters PH, Duell EJ, Molina-Montes E, Quirós JR, Dorronsoro M, Chirlaque MD, Barricarte A, Ljuslinder I, Palmqvist R, Travis RC, Khaw KT, Wareham N, Pischon T, Boeing H. Biomarker patterns of inflammatory and metabolic pathways are associated with

risk of colorectal cancer: results from the European Prospective Investigation into Cancer and Nutrition (EPIC). *European Journal of Epidemiology* 2014;**29**(4):261-275.

48. Nambiar P, Gupta RR, Misra V. An “Omics” based survey of human colon cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 2010;**693**(1–2):3-18.

49. Rezzi S, Ramadan Z, Fay LB, Kochhar S. Nutritional Metabonomics: Applications and Perspectives. *Journal of Proteome Research* 2007;**6**(2):513-525.

50. Rezzi S, Ramadan Z, Martin FP, Fay LB, van Bladeren P, Lindon JC, Nicholson JK, Kochhar S. Human Metabolic Phenotypes Link Directly to Specific Dietary Preferences in Healthy Individuals. *Journal of Proteome Research* 2007;**6**(11):4469-4477.

51. Hu F. Metabolic Profiling of Diabetes: From Black-Box Epidemiology to Systems Epidemiology. *Clinical Chemistry* 2011;**57**(9):1224-1226.

52. Cornelis M, Hu FB. Systems Epidemiology: A New Direction in Nutrition and Metabolic Disease Research. *Current nutrition reports* 2013;**2**(4):10.1007/s13668-013-0052-4.

53. Medina M. Systems biology for molecular life sciences and its impact in biomedicine. *Cellular and Molecular Life Sciences* 2013;**70**(6):1035-1053.

54. Costantini S, Colonna G, Castello G. A Holistic Approach to Study the Effects of Natural Antioxidants on Inflammation and Liver Cancer. In: Zappia V, Panico S, Russo GL, Budillon A, Della Ragione F, eds. *Advances in Nutrition and Cancer*. Cancer Treatment and Research. Vol. 159 Springer Berlin Heidelberg, 2014;311-323.

55. Satija A, Yu E, Willett WC, Hu FB. Understanding Nutritional Epidemiology and Its Role in Policy. *Advances in Nutrition: An International Review Journal* 2015;**6**(1):5-18.

56. Willett W, Stampfer M, Chu NF, Spiegelman D, Holmes M, Rimm E. Assessment of Questionnaire Validity for Measuring Total Fat Intake using Plasma Lipid Levels as Criteria. *American Journal of Epidemiology* 2001;**154**(12):1107-1112.

57. Brunner E, Stallone D, Juneja M, Bingham S, Marmot M. Dietary assessment in Whitehall II: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. *British Journal of Nutrition* 2001;**86**(03):405-414.

58. Fung TT, van Dam RM, Hankinson SE, Stampfer M, Willett WC, Hu FB. Low-Carbohydrate Diets and All-Cause and Cause-Specific MortalityTwo Cohort Studies. *Annals of Internal Medicine* 2010;**153**(5):289-298.

59. Lopez-Garcia E, Rodriguez-Artalejo F, Li TY, Fung TT, Li S, Willett WC, Rimm EB, Hu FB. The Mediterranean-style dietary pattern and mortality among men and women with cardiovascular disease. *The American Journal of Clinical Nutrition* 2014;**99**(1):172-180.

60. Fung T, Hu FB, Wu K, Chiuve SE, Fuchs CS, Giovannucci E. The Mediterranean and Dietary Approaches to Stop Hypertension (DASH) Diets and Colorectal Cancer. *The American Journal of Clinical Nutrition* 2010;**92**(6):1429-1435.

61. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer* 2005;**5**(5):388-96.

62. Colditz G, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer* 2005;**5**(5):388-396.

63. 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.

64. Pai J, Pischon T, Ma J, Manson JE, Hankinson SE, Joshipura K, Curhan GC, Rifai N, Cannuscio CC, Stampfer MJ, Rimm EB. Inflammatory Markers and the Risk of Coronary Heart Disease in Men and Women. *New England Journal of Medicine* 2004;**351**(25):2599-2610.

65. Hankinson SE, Willett WC, Manson JE, Hunter DJ, Colditz GA, Stampfer MJ, Longcope C, Speizer FE. Alcohol, height, and adiposity in relation to estrogen and prolactin levels in postmenopausal women. *Journal of the National Cancer Institute* 1995;**87**(17):1297-302.

66. Willett W, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *American Journal of Epidemiology* 1985;**122**(1):51-65.

67. Feskanich D, Rimm EB, Giovannucci EL, Colditz GA, Stampfer MJ, Litin LB, Willett WC. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *Journal of the American Dietetic Association* 1993;**93**(7):790-796.

68. Rimm E, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and Validity of an Expanded Self-Administered Semiquantitative Food Frequency Questionnaire among Male Health Professionals. *American Journal of Epidemiology* 1992;**135**(10):1114-1126.

69. Hu FB, Stampfer MJ, Rimm E, Ascherio A, Rosner BA, Spiegelman D, Willett WC. Dietary Fat and Coronary Heart Disease: A Comparison of Approaches for Adjusting for Total Energy Intake and Modeling Repeated Dietary Measurements. *American Journal of Epidemiology* 1999;**149**(6):531-540.

70. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L, Willett WC. Reproducibility and Validity of Dietary Patterns Assessed with a Food-Frequency Questionnaire. *The American Journal of Clinical Nutrition* 1999;**69**(2):243-249.

71. Nishihara R, Morikawa T, Kuchiba A, Lochhead P, Yamauchi M, Liao X, Imamura Y, Nosho K, Shima K, Kawachi I, Qian ZR, Fuchs CS, Chan AT, Giovannucci E, Ogino S. A Prospective Study of Duration of Smoking Cessation and Colorectal Cancer Risk by Epigenetics-related Tumor Classification. *Am J Epidemiol* 2013;**178**(1):84-100.

72. Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB, Willett WC. Validity of self-reported waist and hip circumferences in men and women. *Epidemiology* 1990;**1**(6):466-73.

73. Wolf AM, Hunter DJ, Colditz GA, Manson JE, Stampfer MJ, Corsano KA, Rosner B, Kriska A, Willett WC. Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol* 1994;**23**(5):991-9.

74. Chasan-Taber S, Rimm EB, Stampfer MJ, Spiegelman D, Colditz GA, Giovannucci E, Ascherio A, Willett WC. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. *Epidemiology* 1996;**7**(1):81-6.

75. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;**122**(1):51-65.

76. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 1992;**135**(10):1114-26; discussion 1127-36.

77. Wei EK, Colditz GA, Giovannucci EL, Fuchs CS, Rosner BA. Cumulative risk of colon cancer up to age 70 years by risk factor status using data from the Nurses' Health Study. *Am J Epidemiol* 2009;**170**(7):863-72.

78. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, Berry DA. Meta-analyses of colorectal cancer risk factors. *Cancer Causes Control* 2013;**24**(6):1207-22.

79. Hankinson S, Willett WC, Manson JE, Colditz GA, Hunter DJ, Spiegelman D, Barbieri RL, Speizer FE. Plasma Sex Steroid Hormone Levels and Risk of Breast Cancer in Postmenopausal Women. *Journal of the National Cancer Institute* 1998;**90**(17):1292-1299.

80. Eliassen A, Missmer SA, Tworoger SS, Hankinson SE. Endogenous Steroid Hormone Concentrations and Risk of Breast Cancer: Does the Association Vary by a Woman's Predicted Breast Cancer Risk? *Journal of Clinical Oncology* 2006;**24**(12):1823-1830.

81. Tworoger S, Mantzoros C, Hankinson SE. Relationship of Plasma Adiponectin With Sex Hormone and Insulin-like Growth Factor Levels. *Obesity* 2007;**15**(9):2217-2224.

82. Hankinson S, London SJ, Chute CG, Barbieri RL, Jones L, Kaplan LA, Sacks FM, Stampfer MJ. Effect of transport conditions on the stability of biochemical markers in blood. *Clinical Chemistry* 1989;**35**(12):2313-6.

83. Giovannucci E, Colditz GA, Stampfer MJ, Hunter D, Rosner BA, Willett WC, Speizer FE. A Prospective Study of Cigarette Smoking and Risk of Colorectal Adenoma and Colorectal Cancer in U.S. Women. *Journal of the National Cancer Institute* 1994;**86**(3):192-199.

84. Karamanos B, Thanopoulou A, Anastasiou E, Assaad-Khalil S, Albache N, Bachaoui M, Slama CB, El Ghomari H, Jotic A, Lalic N, Lapolla A, Saab C, Marre M, Vassallo J, Savona-Ventura C. Relation of the Mediterranean Diet with the Incidence of Gestational Diabetes. *European Journal of Clinical Nutrition* 2013.

85. Sundaram S JA, Makowski L. *Obesity, metabolism and the microenvironment: Links to cancer*. Vol. 12, 2013.

86. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a dash-style diet and risk of coronary heart disease and stroke in women. *Archives of Internal Medicine* 2008;**168**(7):713-720.

87. Fung T, Hu FB, McCullough ML, Newby PK, Willett WC, Holmes MD. Diet Quality is Associated with the Risk of Estrogen Receptor–Negative Breast Cancer in Postmenopausal Women. *The Journal of Nutrition* 2006;**136**(2):466-472.

88. Fung T, Rexrode KM, Mantzoros CS, Manson JE, Willett WC, Hu FB. Mediterranean Diet and Incidence of and Mortality from Coronary Heart Disease and Stroke in Women. *Circulation* 2009;**119**(8):1093-1100.

89. Appel LJ SF, Carey VJ, Obarzanek E, Swain JF, Miller ER 3rd, Conlin PR, Erlinger TP, Rosner BA, Laranjo NM, Charleston J, McCarron P, Bishop LM; OmniHeart Collaborative Research Group. Effects of protein, monounsaturated fat, and carbohydrate intake on blood pressure and serum lipids: Results of the omniheart randomized trial. *JAMA* 2005;**294**(19):2455-2464.

90. Gadgil MD AL, Yeung E, Anderson CA, Sacks FM, Miller ER 3rd. The Effects of Carbohydrate, Unsaturated Fat, and Protein Intake on Measures of Insulin Sensitivity: Results from the OmniHeart Trial. *Diabetes Care* 2013;**36**(5):1132-1137.

91. Varraso R, Fung TT, Hu FB, Willett W, Camargo CA. Prospective study of dietary patterns and chronic obstructive pulmonary disease among US men. *Thorax* 2007;**62**(9):786-791.

92. Hu F, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D, Willett WC. Prospective study of major dietary patterns and risk of coronary heart disease in men. *The American Journal of Clinical Nutrition* 2000;**72**(4):912-921.

93. 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):935-944.

94. Charlton BM, Rich-Edwards JW2, Colditz GA3, Missmer SA4, Rosner BA5, Hankinson SE6, Speizer FE7, Michels KB8. Oral contraceptive use and mortality after 36 years of follow-up in the Nurses' Health Study: prospective cohort study. *BMJ : British Medical Journal* 2014;**349**:g6356.

95. Charlton BM, Wu K, Zhang X, Giovannucci EL, Fuchs CS, Missmer SA, Rosner B, Hankinson SE, Willett WC, Michels KB. Oral Contraceptive Use and Colorectal Cancer in the Nurses' Health Study I and II. *Cancer Epidemiology Biomarkers & Prevention* 2015;**24**(8):1214-1221.

96. Cao Y, Meyerhardt JA, Chan AT, Wu K, Fuchs CS, Giovannucci EL. Television watching and colorectal cancer survival in men. *Cancer Causes & Control* 2015:1-10.

97. Song M, Hu FB, Spiegelman D, Chan AT, Wu K, Ogino S, Fuchs CS, Willett WC, Giovannucci EL. Adulthood Weight Change and Risk of Colorectal Cancer in the Nurses' Health Study and Health Professionals Follow-up Study. *Cancer Prevention Research* 2015;**8**(7):620-627.

98. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *International Journal of Epidemiology* 2013;**42**(5):1511-1519.

99. VanderWeele T, Vansteelandt S. Mediation Analysis with Multiple Mediators. *Epidemiol Method* 2014;**2**(5):91-115.

100. Rochfort S. Metabolomics Reviewed: A New "Omics" Platform Technology for Systems Biology and Implications for Natural Products Research. *Journal of Natural Products* 2005;**68**(12):1813-1820.

101. Cheng S, Larson MG, McCabe EL, Murabito JM, Rhee EP, Ho JE, Jacques PF, Ghorbani A, Magnusson M, Souza AL, Deik AA, Pierce KA, Bullock K, O'Donnell CJ, Melander O, Clish CB, Vasan RS, Gerszten RE, Wang TJ. Distinct metabolomic signatures are associated with longevity in humans. *Nat Commun* 2015;**6**.

102. Townsend M, Clish CB, Kraft P, Wu C, Souza AL, Deik AA, Tworoger SS, Wolpin BM. Reproducibility of Metabolomic Profiles among Men and Women in 2 Large Cohort Studies. *Clinical Chemistry* 2013;**59**(11):1657-1667.

103. Mayers J, Wu C, Clish CB, Kraft P, Torrence ME, Fiske BP, Yuan C, Bao Y, Townsend MK, Tworoger SS, Davidson SM, Papagiannakopoulos T, Yang A, Dayton TL, Ogino S, Stampfer MJ, Giovannucci EL, Qian ZR, Rubinson DA, Ma J, Sesso HD, Gaziano JM, Cochrane BB, Liu S, Wactawski-Wende J, Manson JE, Pollak MN, Kimmelman AC, Souza A, Pierce K, Wang TJ, Gerszten RE, Fuchs CS, Vander Heiden MG, Wolpin BM. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat Med* 2014;**20**(10):1193-1198.

104. Rosner B. *Fundamentals of biostatistics*. 3rd ed. Boston: Duxbury Press, 1990.

105. Dupont WD. Power calculations for matched case-control studies. *Biometrics* 1988;**44**:1157-1168.

106. Chapman D, Nam J. Asymptotic power of chi square tests for linear trends in proportions. *Biometrics* 1968;**24**:315-327.

107. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *Am J Epidemiol* 1992;**136**(11):1400-13.

108. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* 1990;**132**:734-45.

109. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics Med* 1989;**8**:1051-1069.

110. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *Am J Epidemiol* 1992;**136**:1400-1413.