Interactions Between Folate Intake and Genetic Predictors of Gene Expression Levels Associated with Colorectal Cancer Risk

Cameron B. Haas (✉ chaas@uw.edu)

Fred Hutchinson Cancer Research Center

Yu-Ru Su

Fred Hutchinson Cancer Research Center

Paneen Petersen

Fred Hutchinson Cancer Research Center

Xiaoliang Wang

Fred Hutchinson Cancer Research Center

Stephanie A Bien

Fred Hutchinson Cancer Research Center

Yi Lin

Fred Hutchinson Cancer Research Center

Demetrius Albanes

National Cancer Institute, National Institutes of Health

Stephanie J Weinstein

National Cancer Institute, National Institutes of Health

Mark A Jenkins

The University of Melbourne

Jane C Figueiredo

Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center

Polly A Newcomb

Fred Hutchinson Cancer Research Center

Graham Casey

University of Virginia

Loic Le Marchand

University of Hawaii Cancer Center

Peter T Campbell

American Cancer Society

Ulrike Peters

Fred Hutchinson Cancer Research Center

Li Hsu

Fred Hutchinson Cancer Research Center

Victor Moreno

L'Hospitalet de Llobregat

John D Potter

Fred Hutchinson Cancer Research Center

Lori C Sakoda

Fred Hutchinson Cancer Research Center

Martha L Slattery

University of Utah

Andrew T Chan

Massachusetts General Hospital, Harvard Medical School

Li Li

University of Virginia

Graham G Giles

Cancer Council Victoria

Roger L Milne

Cancer Council Victoria

Stephen B Gruber

University of Southern California

Gad Rennert

Lady Davis Carmel Medical Center

Michael O Woods

Memorial University of Newfoundland

Steven J Gallinger

Lunenfeld Tanenbaum Research Institute, University of Toronto

Sonja Berndt

National Cancer Institute, National Institutes of Health

Richard B Hayes

New York University School of Medicine

Wen-Yi Huang

National Cancer Institute, National Institutes of Health

Alicja Wolk

Karolinska Institutet

Emily White

University of Washington

Hongmei Nan

Indiana University

Rami Nassir

Umm Al-Qura'a University

Noralane M Lindor

Mayo Clinic Arizona

Juan P Lewinger

University of Southern California

Andrew E Kim

University of Southern California

David Conti

University of Southern California

W. James Gauderman

University of Southern California

Daniel D Buchanan

University of Southern California

Research Article

Keywords: Cancer, genetics, colorectal, interaction, folate, nutrition

Posted Date: December 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1097039/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Observational studies have shown higher folate consumption to be associated with lower risk of colorectal cancer (CRC). Understanding whether and how genetic risk factors interact with folate could further elucidate the underlying mechanism. Aggregating functionally relevant genetic variants in set-based variant testing has higher power to detect gene-environment (GxE) interactions and may provide information on the underlying biological pathway.

Objective

We investigated interactions between folate consumption and predicted gene expression on colorectal cancer risk across the genome.

Methods

We used variant weights from the PrediXcan models of colon tissue-specific gene expression as *a priori* variant information for a set-based GxE approach. We harmonized total folate intake (mcg/day) based on dietary intake and supplemental use across cohort and case-control studies and calculated sex and study specific quantiles. Analyses were performed using a mixed effects score tests for interactions between folate and genetically predicted expression of 4,839 genes with available genetically predicted expression. We meta-analyzed results across 23 studies for a total of 13,498 cases with colorectal tumors and 13,918 controls of European ancestry.

Results

We found suggestive evidence of interaction with folate intake for genes including glutathione S-Transferase Alpha 1 (*GSTA1*; $p=4.3E-4$), Tonsuko Like, DNA Repair Protein (*TONSL*; $p=4.3E-4$), and Aspartylglucosaminidase (*AGA*; $p=4.5E-4$). Glutathione is an antioxidant, preventing cellular damage and is a downstream metabolite of homocysteine and metabolized by *GSTA1*. *TONSL* is part of a complex that functions in the recovery of double strand breaks and *AGA* plays a role in lysosomal breakdown of glycoprotein.

Conclusion

We identified three genes involved in preventing or repairing DNA damage that may interact with folate consumption to alter CRC risk.

Background

Folate is a naturally occurring, water-soluble B vitamin that cannot be produced by the human body and plays a key role in DNA formation and is necessary for cellular division and tissue differentiation. It is found abundantly in green leafy vegetables, legumes, fruits, and its more potent form, folic acid, is found in supplements and fortified foods (Food and Drug Administration 1996). Supplementary folic acid is routinely prescribed during pregnancy as an evidence-based intervention to prevent neural tube defects in utero (Crider et al. 2014) (Choumenkovitch et al. 2002). Dietary deficiency is typically found in persons subsisting on inadequate diets, as well as chronic alcoholics with

diminished absorption (Finglas 2000). Fortification of grains with folic acid began in the early 90's to prevent nutritional deficiencies (Initiative n.d.; Hoey et al. 2007). To date, 71 countries have legislative mandates for including folate in the fortification of milled grains (Initiative n.d.). Results pre- and post-fortification and risk of CRC have been somewhat inconsistent (Hirsch et al. 2009; Lee et al. 2011; Mason et al. 2007; Kennedy et al. 2011; Edward Giovannucci 2002b; Zschäbitz et al. 2013), suggesting that folate might play a dual role in colorectal carcinogenesis (Kok et al. 2020; Cornelia M. Ulrich and Potter 2007; Mason and Tang 2017). Given the complexity of the relationship between CRC and folate, there is a need to elucidate the underlying biological mechanisms and possible differential risk based on individual genetics (Cornelia M. Ulrich and Potter 2007).

Increased folic acid consumption is known to lower circulating levels of homocysteine, a common amino acid that has been associated with numerous diseases (Hoey et al. 2007; Bailey et al. 2015; Hao et al. 2008). The absence of folic acid leads to impaired DNA synthesis and disturbances in red blood cell maturation (Kruman et al. 2002). Due to its role as a carrier of one-carbon groups and in folate-mediated one-carbon metabolism (FOCM), abnormal folate consumption has been implicated as a possible cause of cancer (Y. Kim 2004; Y.-I. Kim 2007; Zschäbitz et al. 2013). Consistent with this hypothesis previous studies have shown evidence that greater folate intake is associated with a reduced risk of colorectal adenomas and cancers (CRC) (Edward Giovannucci 2002a; Y.-I. Kim 2007; Figueiredo et al. 2013). A pooled analysis of 13 prospective studies in 2010 observed a modest effect, estimating a 2% risk reduction for CRC per 100 µg/day increase in total folate consumption (D.-H. Kim et al. 2010).

Candidate gene approaches targeting FOCM-related genes have shown associations with CRC risk (Figueiredo et al. 2013, 2014). This has raised interest in studying interactions between folate and genetic variants (Montazeri et al. 2019). As such, it has been hypothesized that germline mutations to the enzyme 5,10-methylenetetrahydrofolate reductase (MTHFR) would be a driver of the effects on folate on CRC risk (Ma et al. 1997; Edward Giovannucci 2002a). A common mutation, 677TT in MTHFR has been associated with a greater decreased risk of CRC in high consumers of folate and low alcohol consumption (Le Marchand et al. 2005; Ma et al. 1997; Slattery et al. 1999) compared to lower folate consumers. However, such analyses have relied on the assumption that FOCM-related genes are the driving genetic force on the pathway from folate consumption to CRC development. A genome-wide approach has the potential to identify novel genes that may modify the folate-CRC association.

To this end, we conducted a novel set-based genome-wide analysis to test interactions between genes and total folate intake on CRC risk. By using a set-based approach we may increase the power to detect associations, which is a common issue in traditional gene-environment interaction studies. We incorporate functional annotation based weights from PrediXcan, a transcriptome prediction tool (Gamazon et al. 2015).

Methods

Study participants

We used epidemiological and genetic data from studies included in three international CRC consortia: the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), the Colorectal Transdisciplinary Study (CORECT) and the Colon Cancer Family Registry (CCFR). Full details have been published previously (Hutter et al. 2012; Newcomb et al. 2007), and the demographic characteristics of study participants are summarized in Table 1. We describe the study designs in **Supplementary Table 1A** and present results for the study design specific effects of total folate on CRC for study designs in **Supplementary Table 1B**. Participants with non-European ancestry were excluded due to

small sample sizes among those with genetic data. Informed consent was given by all participants, and studies were approved by their respective Institutional Review Boards and complies with all relevant ethical regulations.

Table 1

Characteristics of participants included in the analysis from all studies by colorectal cancer case/control status and chi-square p-values for statistical differences.

	Cases	Controls	P-value
N	13,498	13,918	
Male (%)	6,190 (45.9%)	6,014 (43.2%)	<0.001*
Mean Reference Age (SD)	65.0 (10.5)	65.0 (9.9)	0.971
BMI in kg/m2 (%)			
Normal (18.5-24.9)	4,415 (34.0%)	5,260 (39.0%)	<0.001*
Overweight (25-30)	5,355 (41.2%)	5,451 (40.4%)	
Obese (>=30)	3,215 (24.8%)	2,782 (20.6%)	
Mean total energy consumption in kcal/day (SD)	1,938.3 (769.1)	1,911.3 (718.7)	0.003*
Sex-study specific quartile of total folate consumption in mcg/day (%)			
First quartile	3,399 (25.2%)	3,176 (22.8%)	<0.001*
Second quartile	3,089 (22.9%)	3,235 (23.2%)	
Third quartile	4,120 (30.5%)	4,162 (29.9%)	
Fourth quartile	2,890 (21.4%)	3,345 (24.0%)	
Alcohol consumption in g/day (%)			<0.001*
Nondrinker	7,031 (56.2%)	6,852 (52.6%)	
1-28 g/day	4,464 (35.7%)	5,250 (40.3%)	
>28 g/day	1,017 (8.1%)	924 (7.1%)	
Smoking status			0.626
Never smoker	6,452 (48.6%)	6,736 (49.4%)	
Former Smoker	5,271 (39.8%)	5,328 (39.1%)	
Smoker	1,530 (11.5%)	1,560 (11.5%)	
*Statistically significant difference between cases and controls			

Genotype Data

Details on genotyping and imputation have been reported previously (Peters et al. 2013). In brief, DNA was mostly obtained from blood samples, with some from buccal swabs. Several platforms (the Illumina HumanHap 300k, 240k, 550k and OncoArray 610k BeadChip Array system, or Affymetrix platform) were used for genotyping (Peters et al. 2012; Zanke et al. 2007). Samples were excluded on the basis of sample call rate $\leq 97\%$, heterozygosity,

unexpected duplicates or relative pairs, gender discrepancy and principal component analysis (PCA) outlier of HapMap2 CEU cluster. SNPs were excluded on the basis of inconsistency across platforms, call rate <98%, and out of Hardy-Weinberg equilibrium (HWE) in controls ($p < 0.0001$) (Peters et al. 2012). SNPs were imputed to the CEU population in Haplotype Reference Consortium (HRC version r1.0) if not directly genotyped (McCarthy et al. 2016), and restricted by imputation accuracy ($R^2 > 0.3$).

Genetically predicted gene expression

The sets of genetic variants and weights for predicting gene expression were downloaded from the publicly available PredictDB Repository (<https://hakyimlab.org/resource/predixcan/>). The weights for the predicted gene expression were obtained by an elastic net penalized regression approach using the genome-wide variant data and transcriptome data from 169 colon tissue samples from the GTEx project (GTEx v6) [29]. We restricted GTEx data to the transverse colon as it included the entire colonic wall and as such the epithelial layer in the mucosa most relevant to CC development while the GTEx sigmoid colon data only included the muscle layer. Genes with variation explained by SNPs ≥ 0.01 were selected for interaction analyses. A total of 4,839 genes were included.

Exposure assessment

Basic demographics and environmental risk factors were collected using in-person interviews and/or structured questionnaires (Newcomb et al. 2007; Slattery et al. 1997; Christen, Gaziano, and Hennekens 2000; Prorok et al. 2000; "Design of the Women's Health Initiative Clinical Trial and Observational Study. The Women's Health Initiative Study Group." 1998; Hoffmeister et al. 2009; Brenner et al. 2011; Küry et al. 2007; Colditz and Hankinson 2005; E Giovannucci et al. 1994). For these data, we carried out a multi-step data harmonization procedure, reconciling each study's unique protocols and data-collection instruments as discussed previously (Hutter et al. 2012). Folate and folic acid intake were assessed at the reference time using food frequency questionnaires (FFQs). For cohort studies, the reference time was time of enrollment or blood collection. Folate and folic acid intake in each study were determined based on micrograms per day (mcg/day) of folate from foods (i.e., dietary folate) and mcg/day of folic acid from supplements (single or multivitamins) when available. Only two of the 23 studies with dietary folate intake did not capture information regarding supplemental folate. To account for the higher bioavailability of synthetic folic acid vs. natural folate in foods, we calculated total folate intake as dietary folate equivalents (total mcg DFE = mcg of dietary folate + 1.7 x mcg folic acid from supplements) (Sutor and Bailey 2000). Because the time of enrollment for some studies overlapped or followed the period of folic acid fortification (1996-1998), these studies accounted for folic acid fortification when calculating dietary folate intake and entered dietary folate intake as mcg of natural food folate + 1.7 x mcg folic acid from fortified food (see **Supplementary table 1A**). Two studies (OFCCR, DALIS) entered supplement data as regular user vs. nonuser; for these, we assumed regular use was 400 mcg/day or 400 mcg/tablet (for multivitamins), which corresponds to the generic dose in supplements (D.-H. Kim et al. 2010; E Giovannucci et al. 1998). The primary analysis used sex-study specific quartiles of total folate using controls based on the calculated daily dietary and supplemental intake, if available. By using categorical sex-study specific quartiles we reduce the influence of outliers and skewed distributions and is consistent with the Cancer Cohort Pooling Project (Swerdlow et al. 2018). To further explore the differences in bioavailability, secondary analyses we explored sex-study specific quartiles of dietary folate and binary (yes/no) supplemental folate separately.

Statistical Analysis

We used MiSTi (Su et al. 2017), a mixed effects score test for gene-based interaction test with folate consumption on CRC risk. MiSTi modeled the gene-environmental interaction effect by two components. The fixed effects

component incorporates variant functional information from PrediXcan as weights with our genotype data to calculate the genetically predicted gene expression and then assess its interaction with folate consumption. The random effects component involves residual interaction effects that have not been accounted for by the fixed effects. We used sex- and study-specific quantiles of folate consumption. P-values were calculated separately for fixed and random effects interaction terms, after adjusting for age, sex, study, sex-study specific quantiles of total energy consumption in kcal, and principal components to account for population stratification. We used the MiSTi data-adaptive weighted combination approach to combining the fixed and random effects components.

Genes with p-values less than the Bonferroni correction ($0.05/4,839=1.03E-5$) were considered genome-wide statistically significant. P-values that reached false discovery rate (FDR) at 0.2 were considered having suggestive evidence of interaction. We further investigated the direction and magnitude of these interactions using the generalized linear model, which included all covariates in the original model, folate, standardized predicted gene expression, and an interaction term for folate and predicted gene expression. Due to some of the variants having high collinearity, we pruned variants by $R^2 < 0.9$. Genes for which the signal was driven by the random effects component were further investigated to identify individual variants of the gene set as drivers using the same approach with interactions for individual variants and folate.

All analyses were performed using R version 4.0.1 (Team, n.d.).

We performed these additional follow-up analyses for MTHFR, as prior candidate gene studies have shown variants, specifically the C677T mutation, alter the association between folate and CRC (Slattery et al. 1999; Dong et al. 2008; Levine et al. 2010; C M Ulrich et al. 1999; Le Marchand et al. 2005; Torre et al. 2014). We additionally include the results of the gene-environment interaction between rs1801133 (C677T mutation) per additional effect allele with sex-study specific quantiles of total folate consumption on colorectal cancer.

Results

The final sample included 13,498 cases and 13,918 controls with both folate and energy consumption measures available from 23 studies. We present demographic characteristics of all samples and reported measures for factors associated with CRC risk for study participants by case-control status in Table 1. Cases were more likely to be male, have higher BMI, and report consuming less folate daily and more calories daily than compared to controls. Multivariable logistic regression estimated a reduced risk of CRC per-quartile increase in total folate intake, adjusting for sex, age at reference, and total energy intake, and study (OR=0.91, 95% CI: 0.89, 0.93). Sensitivity analyses included further adjustment for smoking and alcohol consumption, which had little effect on the estimates for total folate and CRC risk.

We found no suggestion of interaction between predicted gene expression for the *MTHFR* gene and sex-study specific folate on risk of CRC in our analysis. **Supplementary tables 2A-C** present follow-up analyses conducted to test the interaction per standard deviation change in predicted gene expression within sex-study specific quantiles, allowing for a non-linear relationship between folate quantiles, as well as individual variant weights used in the modeling of predicted gene expression to capture the C677T mutation. In the snp-environment interaction analysis for the rs1801133 variant (C677T mutation), no interaction was shown between each additional effect allele with sex-study specific quantiles of total folate consumption on risk of CRC (ratio of odds ratio=1.02; 95% CI=0.98, 1.06; interaction p-value=0.235).

The median number of SNPs included in the gene sets was 25 (minimum: 1, inter-quartile range [IQR]: 13-43, maximum: 277). Figure 1 displays the quantile-quantile plot for the GxE test that combined both fixed and random effects using adaptive weight. While there was no GxE interaction that reached the Bonferroni threshold (0.05/4,839), three did surpass the false discover rate (FDR) of 0.2.

We present the findings with p-values that surpassed the FDR threshold for gene interactions with total folate consumption and CRC risk in Table 2. We observed suggestive evidence of interactions between total folate intake and 3 independent gene sets on risk of CRC at FDR<0.2., including Glutathione S-Transferase Alpha 1 (*GSTA1*; p=4.3E-4), Tonsuko Like, DNA Repair Protein (*TONSL*; p=4.3E-4), and Aspartylglucosaminidase (*AGA*; p=4.5E-4). In follow-up analyses for these three genes we observed positive interactions for *GST1A* and *AGA*, showing greater risk for CRC associated with higher gene expression and increasing folate consumption (Table 3). As the signal for *TONSL* primarily came from the random effects, indicating one or a few variants were drivers of the association, we investigated the individual interactions of variants with sex-study specific folate. We see two variants as possible drivers of the signal in our main analysis, 8:144964455_T/C and 8:144965104, as shown in Table 4.

Table 2
Top genes from MiSTi for interactions with total folate based on Adaptive Weight Test p-values with FDR < 0.2.

Gene	Chromosome	R ²	Number of SNPs	Fixed Effects	Random Effects	Adaptive Weight P
<i>GSTA1</i>	6	0.17	20	1.6E-4	0.11	4.3E-4
<i>TONSL</i>	8	0.06	28	0.048	1.3E-3	4.3E-4
<i>AGA</i>	4	0.37	39	1.2E-4	0.65	4.5E-4

Table 3
Estimated Odds Ratio (OR) per standard deviation change in predicted gene expression within sex-study specific quantiles of total folate consumption.

Gene	Sex-study specific quantile of total folate	OR	95% Confidence Interval	Interaction P-value*
<i>GSTA1</i>	1	1.29	0.86, 1.92	
	2	1.32	0.88, 1.98	0.48
	3	1.40	0.93, 2.1	0.01
	4	1.45	0.97, 2.17	6.5e-4
<i>AGA</i>	1	1.06	0.79, 1.43	
	2	1.16	0.85, 1.58	0.01
	3	1.15	0.85, 1.56	0.02
	4	1.21	0.89, 1.64	3.6e-4
*P-value tests for difference in quantile specific OR estimate for per standard deviation change in predicted gene expression and the lowest sex-study specific quantile of total folate quantile on colorectal cancer risk.				

Table 4
Variant specific interactions with sex-study specific quantiles of total folate consumption for variants included in TONSL*.

Variant*	Ratio of Odds Ratio	Standard Error	Interaction P-value
8:144757296_A/G	1.04	0.22	0.86
8:144801243_C/T	1.03	0.02	0.09
8:144801593_C/A	0.93	0.04	0.10
8:144856443_G/T	0.82	0.13	0.13
8:144964455_T/C	0.95	0.02	7.0e-3
8:144965104_G/A	0.95	0.02	5.1e-3
8:145651888_G/A	1.00	0.02	0.78
8:145653006_G/A	1.00	0.02	0.76
8:145668042_G/A	1.01	0.02	0.76
8:146216936_A/G	0.96	0.02	0.03
8:146280802_C/T	1.03	0.02	0.19
*Filtering to uncorrelated variants in TONSL ($r^2 < 0.9$)			
**Chromosomal position and reference/alternative alleles			

Discussion

In this sizable analysis including numerous studies we harmonized data on folate consumption and genome-wide genetic data to investigate interactions between folate intake and variants in genes on CRC risk. We observed an inverse association between folate intake and CRC risk across 23 studies. Using our novel statistical set-based GxE mixed effects score tests, MiSTi, we identified 3 genes with suggestive interactive effects with total folate consumption on CRC risk: *GSTA1*, *TONSL*, and *AGA*.

We observed a positive interaction between the predicted gene expression of *GSTA1* and folate for CRC risk. *GSTA1* located at 6p12.2 encodes for an enzyme that functions in cellular detoxification of electrophilic compounds through glutathione metabolism. Electrophilic compounds include carcinogens, therapeutic drugs, environmental toxins, and products of oxidative stress. Glutathione is a product of homocysteine metabolism, a key amino acid correlated with folate intake, and is bound to free radicals by *GSTA1* (Lushchak 2012). Mutations in *GSTA1* could feasibly alter the binding affinity of glutathione to carcinogenic compounds, leading to variation in cancer susceptibility. Of the six SNPs included in our analyses of *GSTA1*, three of the alternative alleles result in missense mutations to the gene (Stelzer et al. 2016). Compromised function of glutathione as an antioxidant due to mutations in *GSTA1* in conjunction with depleted levels of glutathione due to lower homocysteine levels may be a pathway to tumorigenesis (Hoey et al. 2007). Candidate gene studies have shown no association between *GSTA1* and colorectal cancer or adenoma risk (van der Logt et al. 2004; Economopoulos and Sargentanis 2010). However, previous studies have shown interactions between diet, such as cruciferous vegetable consumption, and *GSTA1* genotypes, supporting that associations between this gene and CRC are likely driven by dietary exposures (Coles et al. 2001; Sweeney et al. 2002; Tijhuis et al. 2005).

TONSL in the 8q24.3 region codes for a 1,378 amino acid protein component of the MMS22L-TONSL complex, which functions in recovery of damaged replication forks (O'Donnell et al. 2010). Numerous mutations in *TONSL* are considered pathogenic (Stelzer et al. 2016). Low levels of the MMS22L-TONSL complex result in increased frequency of DNA double-strand breaks and compromised DNA integrity (O'Donnell et al. 2010). In combination with increased DNA damage due to deficiencies in folic acid, impaired functionality of the MMS22L-TONSL due to functional mutations may be a pathway to increase tumorigenesis. Follow-up analyses further suggested that possible associations may be primarily driven by a small subset of variants included in our gene set in the main analysis.

We observed increasing risk of CRC per standard deviation increase in predicted gene expression of *AGA* with increasing folate consumption. The *AGA* gene is in the 4q34.3 region and codes for a 346 amino acid protein that functions in pathways related to the innate immune system and asparagine degradation (GeneCard n.d.). Once the protein is processed into the mature enzyme it takes part in the catabolism of N-linked oligosaccharides, cleaving asparagine from N-acetylglucosamines in one of the final steps in the lysosomal breakdown of glycoproteins. Mutations in the *AGA* gene are known to cause the lysosomal storage disease aspartylglycosaminuria, eventually resulting in neurodegeneration. Previous research has not indicated a link to cancer for this gene.

While we have numerous strengths in performing the largest investigation of gene-folate interactions to date using a powerful set-based approach that allows to account for functional prediction, some limitations should be considered when interpreting these findings. Our study population was limited to those of European descent. As gene expression levels may differ across populations of different ancestry, our results may not be generalizable to populations of non-European ancestry. The studies included in our analysis occurred over a range of time and geographic locations. Fortification with folate occurred in different places at different times and we used adjusted dietary equivalents to account for these differences (see **supplementary table 1A**). Study designs also varied. We looked at the effect size of folate on CRC by case/control versus cohort study designs and did not find a substantive difference to justify stratified analyses (see **supplementary table 1B**).

We utilized colon-specific gene expression data, specifically transverse colon tissue captured by the GTex Project (GTEx Consortium 2013). One limitation of this data is the diversity of cell types aside from epithelial cells of the mucosa of the colon, from which CRC derives given that the entire colonic wall was sampled. The impact of this would cause a dilution of gene expression for the tissue most relevant for CRC. However, we expect this to be an improvement over alternative tissue types including blood or sigmoid colon tissues in GTEx, which were collected from muscle tissues only and would not represent the gene expression profile of interest.

Although MiSTi is a powerful statistical tool, which accounts for both fixed- and random-effects of the gene-folate interaction, none of our findings reached the Bonferroni corrected threshold, which can be overly conservative as many genes are co-expressed. We did not perform independent replication and thus follow-up investigations are warranted, as a FDR of 0.2 should be considered liberal (Su et al. 2017). The previously suggested *MTHFR* gene was not identified in our analysis. However, in using the penalized elastic net to create our predicted gene expression the C677T was not included in the variant weights due to the insignificant contribution to regulation of gene expression. While it was also not seen in the gene-environment interaction analysis either, we believe these results to be representative of an agnostic approach which has not been shown before, as opposed to candidate gene studies.

Our analysis was conducted across numerous studies included in a large, well characterized and harmonized consortium with comprehensive genetic data which enabled a hypothesis-free genome-wide investigation of interactions with folate consumption on CRC risk. Numerous genes evaluated in prior candidate gene-folate

interaction studies, including *MTHFR*, were included among the 4,839 genes examined. However, none of those previously hypothesized genes were found to interact with folate consumption in our analysis (Le Marchand et al. 2005; C M Ulrich et al. 1999; Sharp and Little 2004). We conducted additional follow-up analysis for *MTHFR* using indicator terms for sex-study-specific folate quantiles and interaction terms for all quantiles with predicted gene expression were null (see Supplementary Table 1A). No previous study has agnostically tested for genetic interactions with folate for cancer. Our statistical approach was potentially improved by incorporating functional variant weights and testing gene-sets rather than individual SNPs reducing the penalty for multiple testing. In the end, we found three genes that were suggestive of interacting with folate consumption on risk of CRC, supporting the hypothesis that associations of folate with CRC may be modified by common genetic variation.

The biological functions of our top genes serve to primarily prevent or repair DNA damage. The combined effects of increased DNA damage due to folate deficiencies and compromised functionality of these genes may be an important pathway in CRC tumorigenesis. These findings, particularly for *GSTA1*, warrant follow-up in future studies with comprehensive genetic and data on folate intake in order to confirm the potential role of these genes in interacting with folate on CRC risk.

Declarations

Funding

This was supported by grant number T32 CA094880, R01-CA201407, and U01-CA164930 from the National Institutes of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NCI, NIH.

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA164930, U01 CA137088, R01 CA059045, R01201407). Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) (X01-HG008596 and X-01-HG007585). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268201200008I. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704.

The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

COLO2&3: National Institutes of Health (R01 CA60987).

The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). The CCFR Set-1 (Illumina 1M/1M-Duo) and Set-2 (Illumina Omni1-Quad) scans were supported by NIH awards U01 CA122839 and R01 CA143247 (to GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported by NIH award U19 CA148107 (to SBG) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN268201200008I. The SFCCR Illumina HumanCytoSNP array was supported through NCI award R01 CA076366 (to PAN). Additional funding for the OFCCR/ARCTIC was through award GL201-043 from the Ontario Research Fund (to BWZ), award 112746 from the Canadian Institutes of Health

Research (to TJH), through a Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society (to SG), and through generous support from the Ontario Ministry of Research and Innovation.

Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by the National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health and Human Services (grant numbers U19 CA148107, R01 CA81488, P30 CA014089, R01 CA197350,, P01 CA196569, R01 CA201407) and National Institutes of Environmental Health Sciences, National Institutes of Health (grant number T32 ES013678).

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. This study was conducted with Institutional Review Board approval.

DALS: National Institutes of Health (R01 CA48998 to M. L. Slattery).

Harvard cohorts (HPFS, NHS, PHS): HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, and R35 CA197735) and PHS by the National Institutes of Health (R01 CA042182).

Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8), NCI R01CA136726.

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 509348, 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database.

MEC: National Institutes of Health (R37 CA54281, P01 CA033619, and R01 CA063464).

MECC: This work was supported by the National Institutes of Health, U.S. Department of Health and Human Services (R01 CA81488 to SBG and GR).

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821), the National Institutes of Health, U.S. Department of Health and Human Services (U01 CA74783), and National Cancer Institute of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in the NFCCR samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute.

OFCCR/ARCTIC: Additional funding for the OFCCR/ARCTIC was through award GL201-043 from the Ontario Research Fund (to BWZ), award 112746 from the Canadian Institutes of Health Research (to TJH), through a Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society (to SG), and through generous support from the Ontario Ministry of Research and Innovation.

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council /Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute's Distinguished Professor Award to Alicja Wolk.

VITAL: National Institutes of Health (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

SPAIN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), and Junta de Castilla y León (grant LE22A10-2). Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncologia de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology.

PMH-SCCFR: National Institutes of Health (R01 CA076366 to P. Newcomb and U01 CA074794 to J. Potter).

Availability of data and material

Data will be made available upon request and approval.

Code availability

Please contact the corresponding author for code.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Acknowledgements

CPS-II: The authors thank the CPS-II participants and Study Management Group for their invaluable contributions to this research. The authors would also like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, and cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program.

The Colon CFR graciously thanks the generous contributions of their 42,500 study participants, dedication of study staff, and the financial support from the U.S. National Cancer Institute, without which this important registry would not exist. The content of this manuscript does not necessarily reflect the views or policies of the NIH or any of the

collaborating centers in the CCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR.

Harvard cohorts (HPFS, NHS, PHS): The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We would like to thank the participants and staff of the HPFS, NHS and PHS for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

Kentucky: We would like to acknowledge the staff at the Kentucky Cancer Registry.

PLCO: The authors thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible.

SFCCR: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer and Seattle Cancer Family Registry studies (CORE Studies).

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at:
<http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf>

Disclaimer: Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Conflict of interest: Authors do not have any conflicts of interest to disclose.

References

1. Bailey, Lynn, Patrick Stover, Helene McNulty, Michael Fenech, Jesse Gregory, James Mills, Christine Pfeiffer, et al. 2015. "Biomarkers of Nutrition for Development-Folate Review 1-5." *The Journal of Nutrition* 145 (7). <http://search.proquest.com/docview/1695233723/>.
2. Brenner, Hermann, Jenny Chang-Claude, Christoph M Seiler, Alexander Rickert, and Michael Hoffmeister. 2011. "Protection from Colorectal Cancer after Colonoscopy: A Population-Based, Case-Control Study." *Annals of Internal Medicine* 154 (1): 22–30. <https://doi.org/10.7326/0003-4819-154-1-201101040-00004>.
3. Choumenkovitch, Silvina F, Jacob Selhub, Peter W F Wilson, Jeanne I Rader, Irwin H Rosenberg, and Paul F Jacques. 2002. "Folic Acid Intake from Fortification in United States Exceeds Predictions." *The Journal of Nutrition* 132 (9): 2792–98. <https://doi.org/10.1093/jn/132.9.2792>.
4. Christen, W G, J M Gaziano, and C H Hennekens. 2000. "Design of Physicians' Health Study II—a Randomized Trial of Beta-Carotene, Vitamins E and C, and Multivitamins, in Prevention of Cancer, Cardiovascular Disease, and Eye Disease, and Review of Results of Completed Trials." *Annals of Epidemiology* 10 (2): 125–34. <http://www.ncbi.nlm.nih.gov/pubmed/10691066>.

5. Colditz, Graham A, and Susan E Hankinson. 2005. "The Nurses' Health Study: Lifestyle and Health among Women." *Nature Reviews. Cancer* 5 (5): 388–96. <https://doi.org/10.1038/nrc1608>.
6. Coles, B, S A Nowell, S L MacLeod, C Sweeney, N P Lang, and F F Kadlubar. 2001. "The Role of Human Glutathione S-Transferases (HGSTs) in the Detoxification of the Food-Derived Carcinogen Metabolite N-Acetoxy-PhIP, and the Effect of a Polymorphism in HGSTA1 on Colorectal Cancer Risk." *Mutation Research* 482 (1–2): 3–10. [https://doi.org/10.1016/s0027-5107\(01\)00187-7](https://doi.org/10.1016/s0027-5107(01)00187-7).
7. Crider, K. S., O. Devine, L. Hao, N. F. Dowling, S. Li, A. M. Molloy, Z. Li, J. Zhu, and R. J. Berry. 2014. "Population Red Blood Cell Folate Concentrations for Prevention of Neural Tube Defects: Bayesian Model." *Bmj* 349 (jul29 3): g4554–g4554. <https://doi.org/10.1136/bmj.g4554>.
8. "Design of the Women's Health Initiative Clinical Trial and Observational Study. The Women's Health Initiative Study Group." 1998. *Controlled Clinical Trials* 19 (1): 61–109. <http://www.ncbi.nlm.nih.gov/pubmed/9492970>.
9. Dong, Linda M, John D Potter, Emily White, Cornelia M Ulrich, Lon R Cardon, and Ulrike Peters. 2008. "Genetic Susceptibility to Cancer: The Role of Polymorphisms in Candidate Genes." *JAMA* 299 (20): 2423–36. <https://doi.org/10.1001/jama.299.20.2423>.
10. Economopoulos, Konstantinos P, and Theodoros N. Sergentanis. 2010. "GSTM1, GSTT1, GSTP1, GSTA1 and Colorectal Cancer Risk: A Comprehensive Meta-Analysis." *European Journal of Cancer* 46 (9): 1617–31. <https://doi.org/10.1016/j.ejca.2010.02.009>.
11. Figueiredo, Jane C, Li Hsu, Carolyn M Hutter, Yi Lin, Peter T Campbell, John A Baron, Sonja I Berndt, et al. 2014. "Genome-Wide Diet-Gene Interaction Analyses for Risk of Colorectal Cancer." *PLoS Genetics* 10 (4): e1004228. <https://doi.org/10.1371/journal.pgen.1004228>.
12. Figueiredo, Jane C, A Joan Levine, Jimmy W Crott, James Baurley, and Robert W Haile. 2013. "Folate-Genetics and Colorectal Neoplasia: What We Know and Need to Know Next." *Molecular Nutrition & Food Research* 57 (4): 607–27. <https://doi.org/10.1002/mnfr.201200278>.
13. Finglas, Paul M. 2000. *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin and Choline. Trends in Food Science & Technology*. Vol. 11. [https://doi.org/10.1016/S0924-2244\(01\)00010-3](https://doi.org/10.1016/S0924-2244(01)00010-3).
14. Food and Drug Administration. 1996. "Food Standards: Amendment of Standards of Identity for Enriched Grain Products to Require Addition of Folic Acid." *Fed Regist.*, no. 61: 8781–8807.
15. Gamazon, Eric R, Heather E Wheeler, Kaanan P Shah, Sahar V Mozaffari, Keston Aquino-Michaels, Robert J Carroll, Anne E Eyler, et al. 2015. "A Gene-Based Association Method for Mapping Traits Using Reference Transcriptome Data." *Nature Genetics* 47 (9): 1091–98. <https://doi.org/10.1038/ng.3367>.
16. GeneCard. n.d. "Cytochrome P450 Family 2 Subfamily D Member 6." Accessed May 12, 2018. http://www.genecards.org/cgi-bin/carddisp.pl?gene=CYP2D6#aliases_descriptions.
17. Giovannucci, E, E B Rimm, M J Stampfer, G A Colditz, A Ascherio, and W C Willett. 1994. "Aspirin Use and the Risk for Colorectal Cancer and Adenoma in Male Health Professionals." *Annals of Internal Medicine* 121 (4): 241–46. <http://www.ncbi.nlm.nih.gov/pubmed/8037405>.
18. Giovannucci, E, M J Stampfer, G A Colditz, D J Hunter, C Fuchs, B A Rosner, F E Speizer, and W C Willett. 1998. "Multivitamin Use, Folate, and Colon Cancer in Women in the Nurses' Health Study." *Annals of Internal Medicine* 129 (7): 517–24. <http://www.ncbi.nlm.nih.gov/pubmed/9758570>.
19. Giovannucci, Edward. 2002a. "Epidemiologic Studies of Folate and Colorectal Neoplasia: A Review." *The Journal of Nutrition* 132 (8 Suppl): 2350S–2355S. <https://doi.org/10.1093/jn/132.8.2350S>.

20. ———. 2002b. "Epidemiologic Studies of Folate and Colorectal Neoplasia: A Review." *The Journal of Nutrition* 132 (8): 2350S-2355S. <https://doi.org/10.1093/jn/132.8.2350S>.
21. GTEx Consortium. 2013. "The Genotype-Tissue Expression (GTEx) Project." *Nature Genetics* 45 (6): 580–85. <https://doi.org/10.1038/ng.2653>.
22. Hao, Ling, Quan He Yang, Zhu Li, Lynn B. Bailey, Jiang Hui Zhu, Dale J. Hu, Bo Lan Zhang, et al. 2008. "Folate Status and Homocysteine Response to Folic Acid Doses and Withdrawal among Young Chinese Women in a Large-Scale Randomized Double-Blind Trial." *American Journal of Clinical Nutrition* 88 (2): 448–57. <https://doi.org/88/2/448> [pii].
23. Hirsch, Sandra, Hugo Sanchez, Cecilia Albala, María Pía de la Maza, Gladys Barrera, Laura Leiva, and Daniel Bunout. 2009. "Colon Cancer in Chile before and after the Start of the Flour Fortification Program with Folic Acid." *European Journal of Gastroenterology & Hepatology* 21 (4): 436–39. <https://doi.org/10.1097/MEG.0b013e328306ccdb>.
24. Hoey, Leane, Helene McNulty, Nadina Askin, Adrian Dunne, Mary Ward, Kristina Pentieva, J. J. Strain, Anne M. Molloy, Cliona A. Flynn, and John M. Scott. 2007. "Effect of a Voluntary Food Fortification Policy on Folate, Related B Vitamin Status, and Homocysteine in Healthy Adults." *American Journal of Clinical Nutrition* 86 (5): 1405–13. <https://doi.org/86/5/1405> [pii].
25. Hoffmeister, M, E Raum, A Krtischil, J Chang-Claude, and H Brenner. 2009. "No Evidence for Variation in Colorectal Cancer Risk Associated with Different Types of Postmenopausal Hormone Therapy." *Clinical Pharmacology and Therapeutics* 86 (4): 416–24. <https://doi.org/10.1038/clpt.2009.134>.
26. Hutter, Carolyn M, Jenny Chang-Claude, Martha L Slattery, Bethann M Pflugeisen, Yi Lin, David Duggan, Hongmei Nan, et al. 2012. "Characterization of Gene-Environment Interactions for Colorectal Cancer Susceptibility Loci." *Cancer Research* 72 (8): 2036–44. <https://doi.org/10.1158/0008-5472.CAN-11-4067>.
27. Initiative, Food Fortification. n.d. "FFI - Global Progress." Accessed December 7, 2017. http://ffinetwork.org/global_progress/index.php.
28. Kennedy, Deborah A., Seth J. Stern, Myla Moretti, Ilan Matok, Moumita Sarkar, Cheri Nickel, and Gideon Koren. 2011. "Folate Intake and the Risk of Colorectal Cancer: A Systematic Review and Meta-Analysis." *Cancer Epidemiology* 35 (1): 2–10. <https://doi.org/10.1016/j.canep.2010.11.004>.
29. Kim, Dong-Hyun, Stephanie A Smith-Warner, Donna Spiegelman, Shiaw-Shyuan Yaun, Graham A Colditz, Jo L Freudenheim, Edward Giovannucci, et al. 2010. "Pooled Analyses of 13 Prospective Cohort Studies on Folate Intake and Colon Cancer." *Cancer Causes & Control: CCC* 21 (11): 1919–30. <https://doi.org/10.1007/s10552-010-9620-8>.
30. Kim, Young-in. 2004. "Folate and DNA Methylation: A Mechanistic Link between Folate Deficiency and Colorectal Cancer? Folate and DNA Methylation: A Mechanistic Link between Folate Deficiency and Colorectal Cancer?" 13 (April): 511–19.
31. Kim, Young-In. 2007. "Folate and Colorectal Cancer: An Evidence-Based Critical Review." *Molecular Nutrition & Food Research* 51 (3): 267–92. <https://doi.org/10.1002/mnfr.200600191>.
32. Kok, Dieuwertje E., Wilma T. Steegenga, Eddy J. Smid, Erwin G. Zoetendal, Cornelia M. Ulrich, and Ellen Kampman. 2020. "Bacterial Folate Biosynthesis and Colorectal Cancer Risk: More than Just a Gut Feeling." *Critical Reviews in Food Science and Nutrition* 60 (2): 244–56. <https://doi.org/10.1080/10408398.2018.1522499>.
33. Kruman, Inna I, T S Kumaravel, Althaf Lohani, Ward A Pedersen, Roy G Cutler, Yuri Kruman, Norman Haughey, Jaewon Lee, Michele Evans, and Mark P Mattson. 2002. "Folic Acid Deficiency and Homocysteine Impair DNA

- Repair in Hippocampal Neurons and Sensitize Them to Amyloid Toxicity in Experimental Models of Alzheimer's Disease." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 22 (5): 1752–62.
34. Küry, Sébastien, Bruno Buecher, Sébastien Robiou-du-Pont, Catherine Scoul, Véronique Sébille, Hélène Colman, Claire Le Houérou, et al. 2007. "Combinations of Cytochrome P450 Gene Polymorphisms Enhancing the Risk for Sporadic Colorectal Cancer Related to Red Meat Consumption." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 16 (7): 1460–67. <https://doi.org/10.1158/1055-9965.EPI-07-0236>.
 35. Lee, Jung Eun, Walter C Willett, Charles S Fuchs, Stephanie A Smith-Warner, Kana Wu, Jing Ma, and Edward Giovannucci. 2011. "Folate Intake and Risk of Colorectal Cancer and Adenoma: Modification by Time." *The American Journal of Clinical Nutrition* 93 (4): 817–25. <https://doi.org/10.3945/ajcn.110.007781>.
 36. Levine, A Joan, Jane C Figueiredo, Won Lee, Jenny N Poynter, David Conti, David J Duggan, Peter T Campbell, et al. 2010. "Genetic Variability in the MTHFR Gene and Colorectal Cancer Risk Using the Colorectal Cancer Family Registry." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 19 (1): 89–100. <https://doi.org/10.1158/1055-9965.EPI-09-0727>.
 37. Logt, E M J van der, S M Bergevoet, H M J Roelofs, Z van Hooijdonk, R H M te Morsche, T Wobbes, J B de Kok, F M Nagengast, and W H M Peters. 2004. "Genetic Polymorphisms in UDP-Glucuronosyltransferases and Glutathione S-Transferases and Colorectal Cancer Risk." *Carcinogenesis* 25 (12): 2407–15. <https://doi.org/10.1093/carcin/bgh251>.
 38. Lushchak, Volodymyr I. 2012. "Glutathione Homeostasis and Functions: Potential Targets for Medical Interventions." *Journal of Amino Acids* 2012: 736837. <https://doi.org/10.1155/2012/736837>.
 39. Ma, J, M J Stampfer, E Giovannucci, C Artigas, D J Hunter, C Fuchs, W C Willett, J Selhub, C H Hennekens, and R Rozen. 1997. "Methylenetetrahydrofolate Reductase Polymorphism, Dietary Interactions, and Risk of Colorectal Cancer." *Cancer Research* 57 (6): 1098–1102.
 40. Marchand, Loïc Le, Lynne R Wilkens, Laurence N Kolonel, and Brian E Henderson. 2005. "The MTHFR C677T Polymorphism and Colorectal Cancer: The Multiethnic Cohort Study." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 14 (5): 1198–1203. <https://doi.org/10.1158/1055-9965.EPI-04-0840>.
 41. Mason, Joel B., and San Yuan Tang. 2017. "Folate Status and Colorectal Cancer Risk: A 2016 Update." *Molecular Aspects of Medicine* 53 (February): 73–79. <https://doi.org/10.1016/j.mam.2016.11.010>.
 42. Mason, Joel B, Aaron Dickstein, Paul F Jacques, Paul Haggarty, Jacob Selhub, Gerard Dallal, and Irwin H Rosenberg. 2007. "A Temporal Association between Folic Acid Fortification and an Increase in Colorectal Cancer Rates May Be Illuminating Important Biological Principles: A Hypothesis." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 16 (7): 1325–29. <https://doi.org/10.1158/1055-9965.EPI-07-0329>.
 43. McCarthy, Shane, Sayantan Das, Warren Kretschmar, Olivier Delaneau, Andrew R Wood, Alexander Teumer, Hyun Min Kang, et al. 2016. "A Reference Panel of 64,976 Haplotypes for Genotype Imputation." *Nature Genetics* 48 (10): 1279–83. <https://doi.org/10.1038/ng.3643>.
 44. Montazeri, Zahra, Xue Li, Christine Nyiraneza, Xiangyu Ma, Maria Timofeeva, Victoria Svinti, Xiangrui Meng, et al. 2019. "Systematic Meta-Analyses, Field Synopsis and Global Assessment of the Evidence of Genetic Association Studies in Colorectal Cancer." *Gut*, December, gutjnl-2019-319313. <https://doi.org/10.1136/gutjnl-2019-319313>.

45. Newcomb, Polly A, John Baron, Michelle Cotterchio, Steve Gallinger, John Grove, Robert Haile, David Hall, et al. 2007. "Colon Cancer Family Registry: An International Resource for Studies of the Genetic Epidemiology of Colon Cancer." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 16 (11): 2331–43. <https://doi.org/10.1158/1055-9965.EPI-07-0648>.
46. O'Donnell, Lara, Stephanie Panier, Jan Wildenhain, Johnny M Tkach, Abdallah Al-Hakim, Marie-Claude Landry, Cristina Escribano-Diaz, et al. 2010. "The MMS22L-TONSL Complex Mediates Recovery from Replication Stress and Homologous Recombination." *Molecular Cell* 40 (4): 619–31. <https://doi.org/10.1016/j.molcel.2010.10.024>.
47. Peters, Ulrike, Carolyn M Hutter, Li Hsu, Fredrick R Schumacher, David V Conti, Christopher S Carlson, Christopher K Edlund, et al. 2012. "Meta-Analysis of New Genome-Wide Association Studies of Colorectal Cancer Risk." *Human Genetics* 131 (2): 217–34. <https://doi.org/10.1007/s00439-011-1055-0>.
48. Peters, Ulrike, Shuo Jiao, Fredrick R Schumacher, Carolyn M Hutter, Aaron K Aragaki, John A Baron, Sonja I Berndt, et al. 2013. "Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-Analysis." *Gastroenterology* 144 (4): 799-807.e24. <https://doi.org/10.1053/j.gastro.2012.12.020>.
49. Prorok, P C, G L Andriole, R S Bresalier, S S Buys, D Chia, E D Crawford, R Fogel, et al. 2000. "Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial." *Controlled Clinical Trials* 21 (6 Suppl): 273S-309S. <http://www.ncbi.nlm.nih.gov/pubmed/11189684>.
50. Sharp, Linda, and Julian Little. 2004. "GENOME OF EPIDEMIOLOGY Polymorphisms in Genes Involved in Folate Metabolism and Colorectal Neoplasia." 159 (5): 423–43.
51. Slattery, M L, J Potter, B Caan, S Edwards, A Coates, K N Ma, and T D Berry. 1997. "Energy Balance and Colon Cancer—beyond Physical Activity." *Cancer Research* 57 (1): 75–80. <http://www.ncbi.nlm.nih.gov/pubmed/8988044>.
52. Slattery, M L, J D Potter, W Samowitz, D Schaffer, and M Leppert. 1999. "Methylenetetrahydrofolate Reductase, Diet, and Risk of Colon Cancer." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 8 (6): 513–18. <http://www.ncbi.nlm.nih.gov/pubmed/10385141>.
53. Stelzer, Gil, Naomi Rosen, Inbar Plaschkes, Shahar Zimmerman, Michal Twik, Simon Fishilevich, Tsippi Iny Stein, et al. 2016. "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses." In *Current Protocols in Bioinformatics*, 1.30.1-1.30.33. Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/cpbi.5>.
54. Su, Yu-Ru, Chong-Zhi Di, Li Hsu, and Genetics and Epidemiology of Colorectal Cancer Consortium. 2017. "A Unified Powerful Set-Based Test for Sequencing Data Analysis of GxE Interactions." *Biostatistics (Oxford, England)* 18 (1): 119–31. <https://doi.org/10.1093/biostatistics/kxw034>.
55. Suitor, C W, and L B Bailey. 2000. "Dietary Folate Equivalents: Interpretation and Application." *Journal of the American Dietetic Association* 100 (1): 88–94. [https://doi.org/10.1016/S0002-8223\(00\)00027-4](https://doi.org/10.1016/S0002-8223(00)00027-4).
56. Sweeney, Carol, Brian F Coles, Susan Nowell, Nicholas P Lang, and Fred F Kadlubar. 2002. "Novel Markers of Susceptibility to Carcinogens in Diet: Associations with Colorectal Cancer." *Toxicology* 181–182 (December): 83–87. [https://doi.org/10.1016/s0300-483x\(02\)00259-7](https://doi.org/10.1016/s0300-483x(02)00259-7).
57. Swerdlow, Anthony J., Chinonye E. Harvey, Roger L. Milne, Camille A. Pottinger, Celine M. Vachon, Lynne R. Wilkens, Susan M. Gapstur, Mattias Johansson, Elisabete Weiderpass, and Deborah M. Winn. 2018. "The National Cancer Institute Cohort Consortium: An International Pooling Collaboration of 58 Cohorts from 20

- Countries." *Cancer Epidemiology Biomarkers & Prevention* 27 (11): 1307–19. <https://doi.org/10.1158/1055-9965.EPI-18-0182>.
58. Team, R Core. n.d. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. <http://www.r-project.org/>.
59. Tijhuis, Mariken J, Petra A Wark, Jac M M J G Aarts, Marleen H P W Visker, Fokko M Nagengast, Frans J Kok, and Ellen Kampman. 2005. "GSTP1 and GSTA1 Polymorphisms Interact with Cruciferous Vegetable Intake in Colorectal Adenoma Risk." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 14 (12): 2943–51. <https://doi.org/10.1158/1055-9965.EPI-05-0591>.
60. Torre, Maria Luisa, Giuseppina T Russo, Marta Ragonese, Annalisa Giandalia, Ernesto De Menis, Giorgio Arnaldi, Angela Alibrandi, et al. 2014. "MTHFR C677T Polymorphism, Folate Status and Colon Cancer Risk in Acromegalic Patients." *Pituitary* 17 (3): 257–66. <https://doi.org/10.1007/s11102-013-0499-8>.
61. Ulrich, C M, E Kampman, J Bigler, S M Schwartz, C Chen, R Bostick, L Fosdick, S A Beresford, Y Yasui, and J D Potter. 1999. "Colorectal Adenomas and the C677T MTHFR Polymorphism: Evidence for Gene-Environment Interaction?" *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 8 (8): 659–68. <http://www.ncbi.nlm.nih.gov/pubmed/10744125>.
62. Ulrich, Cornelia M., and John D. Potter. 2007. "Folate and Cancer—Timing Is Everything." *JAMA* 297 (21): 2408. <https://doi.org/10.1001/jama.297.21.2408>.
63. Zanke, Brent W, Celia M T Greenwood, Jagadish Rangrej, Rafal Kustra, Albert Tenesa, Susan M Farrington, James Prendergast, et al. 2007. "Genome-Wide Association Scan Identifies a Colorectal Cancer Susceptibility Locus on Chromosome 8q24." *Nature Genetics* 39 (8): 989–94. <https://doi.org/10.1038/ng2089>.
64. Zschäbitz, Stefanie, Ting-Yuan David Cheng, Marian L Neuhouser, Yingye Zheng, Roberta M Ray, Joshua W Miller, Xiaoling Song, et al. 2013. "B Vitamin Intakes and Incidence of Colorectal Cancer: Results from the Women's Health Initiative Observational Study Cohort." *The American Journal of Clinical Nutrition* 97 (2): 332–43. <https://doi.org/10.3945/ajcn.112.034736>.

Supplementary

Supplementary Tables are not available with this version.

Figures

Figure 1

MiSTi results for Adaptive Weight test of predicted gene expression interactions with total folate intake. Dashed line is the Bonferroni corrected threshold, solid is the false discovery rate <0.2 threshold for p-value significance.