Published in final edited form as:
Cancer Causes Control. 2014 February ; 25(2): 215–226. doi:10.1007/s10552-013-0324-8.

Genetic variation in multiple biologic pathways, flavonoid intake and breast cancer

Nikhil K. Khankari[1,*], **Patrick T. Bradshaw**[2], **Lauren E. McCullough**[1], **Susan L. Teitelbaum**[3], **Susan E. Steck**[4], **Brian N. Fink**[5], **Xinran Xu**[3], **Jiyoung Ahn**[6], **Christine B. Ambrosone**[7], **Katherine D. Crew**[8], **Mary Beth Terry**[9], **Alfred I. Neugut**[8,9], **Jia Chen**[3], **Regina M. Santella**[10], and **Marilie D. Gammon**[1]

[1]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[2]Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[3]Department of Preventive Medicine, Mount Sinai School of Medicine, New York, NY, 10029, USA

[4]Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, 29208, USA

[5]Department of Public Health and Preventive Medicine, University of Toledo, Toledo, OH, 43614, USA

[6]Department of Environmental Medicine, New York University, New York, NY, 10016, USA

[7]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, 14263, USA

[8]Department of Medicine, Columbia University, New York, NY, 10032, USA

[9]Department of Epidemiology, Columbia University, New York, NY, 10032, USA

[10]Department of Environmental Health, Columbia University, New York, NY, 10032, USA

Abstract

Purpose—We previously reported an inverse association between flavonoid intake and breast cancer incidence, which has been confirmed by others; but no studies have considered simultaneously potential interactions of flavonoids with multiple genetic polymorphisms involved in biologically-relevant pathways (oxidative stress, carcinogen metabolism, DNA repair, and one-carbon metabolism).

Methods—To estimate interaction effects between flavonoids and 13 polymorphisms in these four pathways on breast cancer risk, we used population-based data (N = 875 cases and 903 controls) and several statistical approaches, including conventional logistic regression and semi-Bayesian hierarchical modeling (incorporating prior information on the possible biological functions of genes), which also provides biologic pathway-specific effect estimates.

Results—Compared to the standard multivariate model, the results from the hierarchical model indicate that gene-by-flavonoid interaction estimates are attenuated, but more precise. In the hierarchical model, the average effect of the deleterious versus beneficial gene, controlling for average flavonoid intake in the DNA repair pathway, and adjusted for the three other biologically-

[*]To whom correspondence should be addressed. Tel: 919-966-7421; Fax: 919-966-2089; khankari@email.unc.edu.

Conflict of Interest Statement: None declared.

relevant pathways (oxidative stress, carcinogen metabolism, and one-carbon metabolism), resulted in a 27% increase risk for breast cancer [Odds Ratio (OR) = 1.27; 95% Confidence Interval (CI) = 0.70, 2.29]. However, the CI was wide.

Conclusions—Based on results from the semi-Bayesian model, breast cancer risk may be influenced jointly by flavonoid intake and genes involved in DNA repair, but our findings require confirmation.

Keywords

dietary flavonoid intake; breast cancer; genetic polymorphisms; hierarchical model

Introduction

We observed a reduction of the risk of breast cancer in relation to flavonoid intake in a population-based study of American women [1]. Others [2–4] have confirmed these findings, which increases the attractiveness of using these compounds as potential chemo-preventives. Intervention trials and public health messaging would be enhanced if we had a better understanding of whether there are subgroups of women who are potentially more susceptible to the beneficial effects of flavonoids. However, a comprehensive consideration of potential interactions between flavonoid intake and multiple, biologically plausible genetic pathways has not previously been reported.

Flavonoids are a subgroup of more than 4000 polyphenols [5] that are known to have an inhibitory effect on hormone-related cancers. With six classes of flavonoids (including flavonols, flavones, flavonones, flavan-3-ols, anthocyanidins, isoflavones, and the related lignans) the biologic activity of these compounds spans multiple biologic pathways. Phytoestrogens, a group of flavonoids including the isoflavones and lignans, possess anti-estrogenic properties, and have demonstrated effects on aromatase inhibition, competition with estradiol for binding to estrogen receptors, and reduce levels of bioavailable estrogen via increased production of serum sex hormone binding globulin [6–8]. In addition to affecting breast cancer risk via an estrogenic pathway, flavonoids also act through other biologic pathways including: carcinogen metabolism, oxidative stress, and DNA repair. For example, flavonoids inhibit the activation of carcinogens by cytochromes (Phase 1) and induce the inactivation of these carcinogenic metabolites (Phase 2) [9]. Flavonoids are also involved in the oxidative stress pathway via two different mechanisms: first, flavonoids can act as antioxidants and scavenge free radicals [10, 11] in order to reduce oxidative stress; and second, flavonoids have also been shown to alter the activity of antioxidant enzymes [12–15], thereby reducing reactive oxygen species. Flavonoids can also stimulate DNA repair enzymes thereby preventing tumor initiation [16, 17]. Finally, polyphenols are methylated via COMT using *S*-adenosylmethionine (SAM), a product of the one-carbon methylation pathway, thereby reducing the availability of SAM for DNA methylation [18]. The potential interaction between flavonoid intake and genes involved in the four biologically-relevant pathways of interest is not limited to a specific flavonoid subtype [13, 16–21]. Thus, it is possible that the effects of total flavonoid intake on breast cancer risk could be modified by genetic variation in genes involved in carcinogen metabolism, oxidative stress, DNA repair, and one-carbon methylation pathways.

Surprisingly, few investigators [22, 23] have considered gene-flavonoid interactions with breast cancer risk, and none have considered multiple polymorphisms in multiple pathways simultaneously. Conventional approaches to considering gene-environment interactions generally examine genetic variants in a single gene with varying levels of exposure. For example, using a conventional approach, we could analyze the varying levels of flavonoid

intake among myeloperoxidase (*MPO*) genetic subgroups on breast cancer risk. However, these conventional models fail to consider the biologic similarity of *MPO* with other oxidative stress genes (e.g. *MnSOD, CAT*) [24]. Another option involves including all exposures, genes, and interactions in a single model. This method is not statistically optimal, since the model will often be over-parameterized, yielding imprecise estimates [24]. Alternatively, a semi-Bayesian approach using hierarchical modeling, as proposed by Witte et al. [25], can help further elucidate the effects of multiple exposures, multiple genes, and their interactions on breast cancer incidence by incorporating prior biologic information. In this ancillary study, we utilized population-based data to examine both interactive effects of total flavonoid intake and multiple genes (random effects) and the effect of biologic pathways (fixed effects) and breast cancer incidence adopting the hierarchical modeling approach.

Materials and methods

Details of the parent study, the population-based case-control component of the Long Island Breast Cancer Study Project (LIBCSP), have been published previously [26]. Institutional Review Board approval was obtained for the parent study and this ancillary project.

Study Population

Cases were women newly diagnosed with either primary *in situ* or invasive breast cancer between 1 August 1996 and 31 July 1997, and were English-speaking residents of Long Island, New York (Nassau and Suffolk counties) at the time of diagnosis. Newly diagnosed cases were ascertained using a ‘super-rapid’ identification network where study personnel contacted the pathology departments from participating hospitals either 2–3 times per week or daily (for hospitals with the largest numbers of newly diagnosed cases). Permission to contact eligible case women was obtained via physicians. Control women were randomly sampled from the same two Long Island counties, using Waksberg’s method of random digit dialling [27] for those under 65 years of age, and the Health Care Finance Administration (HCFA) rosters for those 65 years and older. Controls were frequency matched by 5-year age group to the expected age distribution of the cases.

Sample Size

Respondents to the main case-control interview included 1508 cases and 1556 controls [23]. Among these, 98% of cases (N=1481) and controls (N=1518) also completed the self-administered 101-item modified Block food frequency questionnaire (FFQ), which had been previously validated [28–30]. The instrument was specifically modified to include additional food sources of flavonoids [31]. Approximately 73% of cases (N=1045) and controls (N=1098) completed the FFQ, donated a blood sample and had available genotyping data for this project. The following exclusions were made for this ancillary study: (1) missing data on total energy intake (n=31); (2) subjects with total energy intake ±3 standard deviations from the mean (n=28); and (3) missing genetic data on any of the 13 SNPs of interest (n=306). Thus, the final complete-case analysis included 1778 subjects (cases = 875, controls = 903).

Risk Factor Assessment

The main case-control study questionnaire was administered at each subject’s home by a trained interviewer. On average, study participants were interviewed within three months of their diagnosis date (cases) or within 5.5 months of identification (controls). Respondents were asked about their demographic characteristics, pregnancy history, menstrual history, hormone use, medical history, family history of cancer, body size changes, alcohol use, active and passive cigarette smoking, physical activity, occupational history, and other

environmental exposures previously [26]. For the 98% of participants who self-completed the FFQ, flavonoid intake was estimated by linking data from the FFQ responses to US Department of Agriculture databases [31].

Genotyping

For the 73% of participants who donated a blood sample, DNA was isolated in the laboratory of Dr. Regina Santella at Columbia University using standard phenol and chloroform isoamyl alcohol extraction and RNase treatment [32].

The 13 genes selected for this ancillary project were chosen to represent the four biologic pathways (carcinogen metabolism, oxidative stress, DNA repair, and one-carbon methylation) that could potentially interact with flavonoids to affect breast carcinogenesis [11, 13–15, 17, 19–21, 33–35]. In addition, our polymorphism selection was influenced by our previous findings of modest effect estimates for the associations between each of the 13 presumed functional polymorphisms of these genes and breast cancer incidence [36–43].

Genotyping for the oxidative stress genes (manganese superoxide dismutase –*MnSOD* rs1799725, myeloperoxidase – *MPO* rs2333227, catalase – *CAT* rs1001179, and catechol-O-methyltransferase – *COMT* rs4680) and for the phase two metabolism genes (glutathione *S*-transferases – *GSTA1* rs3957356, *GSTP1* rs1695) was conducted using BioServe Biotechnologies in Laurel, MD, using Sequenom's high throughput matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry [36–39, 41]. Genotyping for deletion of *GSTM1* and *GSTT1* was carried out by multiplex PCR [41]. The DNA repair enzymes (excision repair cross-complementing group 1 – *ERCC1* rs3212986, xeroderma pigmentosum group A – *XPA* rs1800975, xeroderma pigmentosum group D – *XPD* rs1799793) was conducted using either a fluorescence polarization method (*ERCC1*) or Taqman assays (*XPA*, *XPD*) [42]. Genotyping for polymorphisms for genes involved in the one-carbon methylation pathway (methylenetetrahydrofolate reductase - *MTHFR* C677T rs1801133, *MTHFR* A1298C rs1801131) were conducted using a modification of the PCR-RFLP method of Frosst et al. [44].

Tests for Hardy-Weinberg equilibrium (HWE) among the controls were conducted previously. Polymorphisms were reported to be in HWE for *MnSOD* (p=0.86) [36][38], *CAT* (p=0.85) [38], *GSTA1* (p=0.84) [40], *GSTP1* (p=0.38) [45], *ERCC1* (p=0.75) [42], *XPA* (p=0.68) [42], *XPD* (p=0.83) [42], *MTHFR* C677T (p=0.96) [43], *MTHFR* A1298C (p=0.84) [43], and *COMT* (p=0.85) [39]. The *MPO* polymorphism was not in HWE (p=0.03), however the observer agreement in 8% of the randomly selected sample was high (kappa statistic=0.91), and the failure rate of the assay was less than 1% [37]. Also, genotypes and frequencies for the *MPO* polymorphism were reported to be similar to those observed in other studies [37].

Statistical Analyses

We examined the effects of flavonoid-gene interactions on breast cancer risk focusing on presumed functional polymorphisms in 13 genes in four biologically relevant pathways, including: (1) increased oxidative stress (*MnSOD*, *MPO*, *GSTA1*, *COMT*); (2) reduced Phase 2 metabolism (*GSTA1*, *GSTM1*, *GSTP1*, *GSTT1*); (3) reduced DNA repair capability (*ERCC1*, *XPA*, *XPD*); and (4) reduced methyl group formation (*MTHFR* C677T, *MTHFR* A1298C, *COMT*). Gene alleles were dichotomized using a dominant model, with the proposed "high risk" genotype given a value of "1," and the "low risk" genotype assigned a value of 0; the one exception was *COMT*, where the genotype according to dominant model reduces oxidative stress, and therefore for the Z-matrix it was coded as " −1" (see below). High risk genotypes were determined for each SNP based on prior literature regarding its

function within the four biologic pathways of interest [44, 46–66]. In order to maintain adequate cell sizes for examining the interaction, total flavonoid intake was dichotomized at the median, based on the distribution among controls.

To examine the gene-by-flavonoid interactions on breast cancer risk, three modeling approaches were employed: a single interaction term within a logistic regression model; all 13 interaction terms within a multivariate logistic regression model; and hierarchical regression, that simultaneously estimates the 13 interaction effects, but considers them clustered within the four pathways. A single referent coding scheme was used for all three models. Those with low flavonoid intake and the variant genotype (for most SNPs this genotype is thought to have deleterious effect, with the exception of the *COMT*, as discussed above) were considered the highest risk subgroup; this group was considered the referent category for the analysis. Confidence Limit Ratios (CLR), a measure of precision for ratio effect measures [67], was calculated for all estimates by dividing the upper 95% confidence limit by the lower 95% confidence limit. All analyses were conducted using PROC GENMOD (single interaction and multivariate models) or PROC GLIMMIX (hierarchical model) in SAS version 9.2 (SAS Institute, Cary, NC).

Unconditional logistic regression was used for the single interaction and multiple interaction models [68]. All models included the frequency matching factor age. In addition, the single interaction models included the main effect for the gene, the main effect for total flavonoid intake, the multiplicative interaction term, and potential confounders (see below). The multivariate models also included the main effects of all 13 genes, the main effect of total flavonoid intake, all multiplicative interaction terms between the 13 genes and total flavonoid intake, and potential confounders. Potential confounders (including age, total energy intake, body mass index, hormone replacement therapy, oral contraceptive use, alcohol intake, physical activity, family history of breast cancer) for this analysis were identified using a Directed Acyclic Graph (DAG) [69] and 10% change in effect estimate. Only total energy intake was found to change the effect estimates by more than 10%. Therefore, the final covariates in the three models included the matching factor age, and total energy intake.

Hierarchical regression models were based on the indicator-term approach for assessing gene-environment interactions [24]. The following model represents the first stage of the hierarchical model:

$$\mathrm{logit}(\mathrm{p}|\mathrm{X},\mathrm{W})=\alpha+\mathrm{X}\beta+\mathrm{W}\gamma \quad (1)$$

where p = risk of breast cancer, X is the matrix of indicator variables for the gene-flavonoid interaction, β is the effects for the gene-flavonoid interaction, and W is the matrix of covariates (age and total energy intake). To improve our estimates of breast cancer risk, we used a hierarchical regression incorporating prior knowledge regarding biologic pathways. The second level of the hierarchical regression for the logistic coefficients β of the indicator-terms was:

$$\beta=\mathrm{Z}\pi+\delta \quad (2)$$

where Z is the second-stage design matrix (see Table 1), which includes our prior information about the indicator terms and their respective roles in the biologic pathways; π is the vector of coefficients corresponding to the effects of the second stage covariates on breast cancer (i.e. the independent effect of each biologic pathway on breast cancer risk), and δ is the residual association of each indicator term on breast cancer which is not explained by the Z-matrix. The residual association, δ, was assumed to be normally

distributed with mean = 0 and variance = τ^2. A semi-Bayes approach was used in the hierarchical regression analysis and τ^2 was set at a constant of 0.1225 ($\tau = 0.35$), which allows the corresponding 95% confidence intervals to have a range of $e^{(3.92*\tau)} \approx 4.0$ if $\tau = 0.35$ [25]. Setting $\tau = 0.35$ provides adequate 95% CI coverage for residual effects that are not accounted for by the first and second stage in this analysis of 39 first-stage covariates (interaction terms) and 5 second-stage covariates (genetic pathways and indicator of high flavonoid intake) [70].

The columns of the Z-matrix represented each of the four biologic pathways of interest (i.e. increased reactive oxygen species, reduced phase 2 metabolism, reduced DNA repair capability, and reduced methyl group formation), with a fourth column representing the exposure (i.e. high flavonoid intake). If the indicator term contained the deleterious allele then a score of "1" was given in the appropriate column of the Z matrix. A score of "−1" was given to those indicators where the high risk genotype was proposed to have an opposite effect within the same biologic pathway (e.g. COMT). If the indicator term contained high flavonoid intake, then a score of "1" was given in the flavonoid intake column. The first (equation 1) and second-stage (equation 2) estimates are then combined to provide posterior estimates of each indicator term's effect (according to which pathway they belong) and breast cancer risk:

$$\text{logit}(\text{p}|\text{X}, \text{Z}, \text{W}) = \alpha + \text{XZ}\pi + \text{X}\delta + \text{W}\gamma \quad (3)$$

where the X, Z, W matrices are the same as defined previously, the π and γ coefficients are fixed, and the δ coefficient is random.

Results

The effect estimates and 95% confidence intervals for the single interaction and multivariate logistic regression models, and the hierarchical models are presented in Table 2. Compared to the estimates from the single interaction model, the estimates of the multivariate model tend to be less precise, as measured by the CLR. For example, the estimates for the beneficial *MnSOD* allele and high flavonoid intake indicate a five percent risk reduction (OR = 0.95, 95% CI = 0.70, 1.30) when examined using the single interaction logistic regression model. This same estimate in the multivariate logistic regression model for *MnSOD* increases slightly (OR = 1.18, 95% CI = 0.48, 2.90), but is less precise (CLR = 6.01), in comparison to the estimate derived from the single interaction model (CLR = 1.86). In the hierarchical model, the effect estimates shows a seven percent risk reduction seen in this *MnSOD* grouping, and the estimates are more precise (CLR = 2.92), when compared to the multivariate model. As shown in Table 2, similar trends were seen across other SNPs involved in the oxidative stress pathway, with risk reductions across all interaction terms.

For phase II metabolism, the observed risk reduction for the interaction between the presence of *GSTM1* (which indicates increased detoxification) and high flavonoid intake is similarly reduced in both the single-interaction model (OR = 0.78, 95% CI = 0.59, 1.02) and the multivariate model (OR = 0.81, 95% CI = 0.34, 1.93). The hierarchical model reported a 6% risk reduction when the gene is present and high flavonoid intake (OR = 0.94, 95% CI = 0.55, 1.61). The hierarchical estimate, however, was more precise (CLR = 2.93) compared to the multivariate model (CLR = 5.69). This similar pattern in estimates was observed with *GSTA1*. For both *GSTP1* SNP and *GSTT1* deletion, the multivariate model for the beneficial genotype and high flavonoid intake indicator term resulted in increased risk for breast cancer. However, the hierarchical model reported effect estimates suggesting a risk reduction, and these estimates were more precise when compared to multivariate model.

For the genes involved in the one-carbon methylation pathway, we observed risk reductions of 12% for the *MTHFR* C677T (OR = 0.88, 95% CI = 0.51, 1.51) and *COMT* (OR = 0.88, 95% CI = 0.49, 1.58) SNPs in the hierarchical models for the indicator term for beneficial genotype and high flavonoid intake. There was null association reported for this same grouping for the *MTHFR* A1298C SNP (OR = 1.03, 95% CI = 0.60, 1.78).

For DNA repair, among those with the beneficial genotype for *ERCC1* and high flavonoid intake, risk for breast cancer was increased in the multivariate model (OR = 1.07, 95% CI = 0.44, 2.57), whereas the single interaction model estimates an approximately 20% risk reduction (OR = 0.79, 95% CI = 0.61, 1.04). The hierarchical model reports a null effect of high flavonoid intake and beneficial genotype on breast cancer risk. For the other SNPs involved in DNA repair pathways, risk reductions were evident in the hierarchical model for *XPA* and *XPD* for this same grouping. The deleterious and high flavonoid intake grouping reported increased risk (ranging from 10–24%) in the hierarchical model for all three SNPs involved in the DNA repair pathway.

Table 3 provides fixed effect estimates from the hierarchical model. The fixed effects from the hierarchical model provide an estimate of the relative importance of the individual biologic pathways. The strongest effect can be seen among the DNA repair genes. The effect of having a deleterious versus beneficial genotype averaged over total flavonoid intake in the DNA repair biologic pathway, while adjusting for the genes involved in the other three pathways, increases risk for breast cancer by 27% (OR = 1.27, 95% CI = 0.70, 2.29), although the confidence interval is wide. There are essentially no effects seen for the fixed effects for the three other biologic pathways. Additionally, fixed effects were estimated for gene-only hierarchical models stratified by high and low flavonoid intake. Similar patterns were seen for both sets of fixed effect estimates (Supplementary Tables 4–5). Also, stronger effects were seen for the fixed effect estimates for reduced DNA repair capability among postmenopausal women (Supplementary Tables 6–7).

Discussion

Using population-based data from the LIBCSP, we found that the hierarchical model incorporating multiple flavonoid-gene interactions improved the precision of the effect estimates for the associations with breast cancer risk compared to the multivariate model. The results from the hierarchical model suggest that high flavonoid intake along with beneficial genotypes for *MnSOD*, *MPO*, *CAT*, *GSTA1*, *GSTP1*, *XPA*, *MTHFR* C677T, and *COMT* SNPs, may reduce the risk for breast cancer, although confidence intervals were wide for most observations.

Interestingly, high flavonoid intake appeared to enhance the activity of the deleterious SNPs of DNA repair genes. Also, the average effect of deleterious genotypes (compared to beneficial genotypes) in the DNA repair pathway (independent of the other two pathways) was suggested to increase breast cancer incidence by 27% for average total flavonoid intake, though the CI was wide. A recent study [20] suggested that soy intake could inhibit DNA repair capability inducing a toxic environment for non-small cell lung cancer. Thus, it is possible that high flavonoid intake could increase the activity of a deleterious genotype for DNA repair genes. While soy is only one source of flavonoids, it is still possible that this effect persists when considering total flavonoid intake. This may provide some insight into the results observed for the interactive effect of high flavonoid intake and deleterious alleles on increasing risk for breast cancer among the SNPs involved in the DNA repair pathway. However, our results require confirmation.

The benefits of this analysis include examining gene-environment interactions in a large case-control study with data available on total flavonoid intake and SNPs that are involved in multiple biologic pathways. We considered three different statistical modeling approaches to examining gene-environment interactions (single interaction, multivariate, and hierarchical models). The benefit of the hierarchical modeling approach includes allowing us to consider multiple biologically related genes simultaneously along with the ability to incorporate prior information regarding the biologic similarity of different genes. In addition to the random effects – in other words the individual effects of each of the gene-by-flavonoid interactions, the hierarchical model also provides fixed effects estimates for the independent effects of biologic pathways on breast cancer risk.

Another benefit of our study is our use of population-based data that was collected using incidence density sampling [26]. This superior approach permits stronger inferences regarding the impact of our findings on breast cancer incidence, as compared to other case-control sampling designs, which have more limited interpretations [69].

This analysis also has limitations. Since the LIBCSP study population is predominately Caucasian women, reflecting the underlying population of the two counties on Long Island, NY studied [26], the examination of racial subgroups is not possible. Thus, the generalizability of the study results is limited to Caucasian American women for whom the risk of breast cancer is high [71]. Only 13 polymorphisms were examined in this analysis, and inclusion of more genes in the analysis could potentially change the estimates of both the random and fixed effects. Incorporating additional genes within each pathway may help to better represent the functional effect of each pathway on breast cancer. However, a larger sample size would be more optimal to estimate interactive effects of flavonoid intake and additional polymorphisms when utilizing either a conventional or hierarchical approaches. In addition, a larger, more racially heterogeneous sample size would permit examination of effects within breast cancer subtypes. An additional concern is our reliance on the FFQ responses to estimate flavonoid intake, rather than employment of a more objective biomarker, for example. However, the half-life of flavonoids in plasma is less than three hours [72], which limits the usefulness of this exposure assessment tool in a population-based study such as ours. Additionally, a more detailed specification of the Z-matrix in the hierarchical regression models could be achieved by incorporating information regarding enzyme kinetics for each genetic polymorphism. For example, within the oxidative stress pathway, distinguishing polymorphisms that may result in greater enzymatic activity compared to other polymorphisms within the Z-matrix may help to improve estimation of the effect of biologic pathways on breast cancer risk. However, the available laboratory data on enzyme kinetic parameters for the genetic polymorphisms of interest to our study was too limited for use in our current analysis. Future epidemiologic studies would benefit from an improved understanding of these genes in the laboratory.

In conclusion, the findings reported here – which suggest a possible interaction between flavonoid intake and the DNA repair pathway on breast cancer risk -- are derived from an innovative method for understanding the impact of flavonoid intake on genes involved in multiple pathways, although a null association could not be ruled out. The results from this ancillary study highlight the importance of examining the interaction of multiple genes with total flavonoid intake to better understand which combinations of genetic subgroups and flavonoid intake confer the greatest impact in breast cancer risk for women. Future studies that examine gene-environment interactions should consider larger, more racially diverse populations that would enhance examination of effects by breast cancer subtype. Additional information regarding the effect of genetic variation on resulting enzymatic activity could help to improve the specification of the Z-matrix, and thus may provide further insight into the effect of biologic pathways on breast cancer risk. Future researchers should also consider

alternative approaches for modeling complex gene-environment interactions. However, using a hierarchical model, as we did here, may help to elucidate the biologic pathways that modulate the effect of flavonoids on breast cancer risk. This information could help to identify subgroups of women that may be more susceptible to the effects of increasing total flavonoid intake.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge Drs. Duncan Thomas and Charles Poole for their comments on previous versions of this analysis.

Funding

National Institutes of Health (U01CA/ES66572, P30ES10126, P30ES0090809, P50CA58233, R01CA109753, K07CA102640, and T32CA09529); Department of Defense (BC031746 and W81XWH-06-1-0298); and private foundations (Breast Cancer Research Fund and the American Institute for Cancer Research).

References

1. Fink BN, Steck SE, Wolff MS, Britton JA, Kabat GC, Schroeder JC, Teitelbaum SL, Neugut AI, Gammon MD. Dietary flavonoid intake and breast cancer risk among women on Long Island. Am J Epidemiol. 2007; 165:514–523. [PubMed: 17158855]
2. Peterson J, Lagiou P, Samoli E, Lagiou A, Katsouyanni K, La Vecchia C, Dwyer J, Trichopoulos D. Flavonoid intake and breast cancer risk: a case--control study in Greece. Br. J. Cancer. 2003; 89:1255–1259. [PubMed: 14520456]
3. Bosetti C, Spertini L, Parpinel M, Gnagnarella P, Lagiou P, Negri E, Franceschi S, Montella M, Peterson J, Dwyer J, Giacosa A, La Vecchia C. Flavonoids and breast cancer risk in Italy. Cancer Epidemiol. Biomarkers Prev. 2005; 14:805–808. [PubMed: 15824147]
4. Rossi M, Bosetti C, Negri E, Lagiou P, La Vecchia C. Flavonoids, proanthocyanidins, and cancer risk: a network of case-control studies from Italy. Nutr. Cancer. 2010; 62:871–877. [PubMed: 20924962]
5. Dai Q, Franke AA, Jin F, Shu XO, Hebert JR, Custer LJ, Cheng J, Gao YT, Zheng W. Urinary excretion of phytoestrogens and risk of breast cancer among Chinese women in Shanghai. Cancer Epidemiol. Biomarkers Prev. 2002; 11:815–821. [PubMed: 12223424]
6. Parl FF, Dawling S, Roodi N, Crooke PS. Estrogen metabolism and breast cancer: a risk model. Ann. N. Y. Acad. Sci. 2009; 1155:68–75. [PubMed: 19250193]
7. Kumar NB, Cantor A, Allen K, Riccardi D, Cox CE. The specific role of isoflavones on estrogen metabolism in premenopausal women. Cancer. 2002; 94:1166–1174. [PubMed: 11920488]
8. Brownson DM, Azios NG, Fuqua BK, Dharmawardhane SF, Mabry TJ. Flavonoid effects relevant to cancer. J. Nutr. 2002; 132:3482S–3489S. [PubMed: 12421874]
9. Moon YJ, Wang X, Morris ME. Dietary flavonoids: effects on xenobiotic and carcinogen metabolism. Toxicol. In. Vitro. 2006; 20:187–210. [PubMed: 16289744]
10. Pahari B, Chakraborty S, Chaudhuri S, Sengupta B, Sengupta PK. Binding and antioxidant properties of therapeutically important plant flavonoids in biomembranes: Insights from spectroscopic and quantum chemical studies. Chem. Phys. Lipids. 2011
11. Pietta PG. Flavonoids as antioxidants. J. Nat. Prod. 2000; 63:1035–1042. [PubMed: 10924197]
12. Fernandez-Pachon MS, Berna G, Otaolaurruchi E, Troncoso AM, Martin F, Garcia-Parrilla MC. Changes in antioxidant endogenous enzymes (activity and gene expression levels) after repeated red wine intake. J. Agric. Food Chem. 2009; 57:6578–6583. [PubMed: 19722566]

13. Sun L, Zhang J, Lu X, Zhang L, Zhang Y. Evaluation to the antioxidant activity of total flavonoids extract from persimmon (Diospyros kaki L.) leaves. Food Chem. Toxicol. 2011; 49:2689–2696. [PubMed: 21802475]

14. Rohrdanz E, Ohler S, Tran-Thi QH, Kahl R. The phytoestrogen daidzein affects the antioxidant enzyme system of rat hepatoma H4IIE cells. J. Nutr. 2002; 132:370–375. [PubMed: 11880557]

15. Petacci F, Freitas SS, Brunetti IL, Khalil NM. Inhibition of peroxidase activity and scavenging of reactive oxygen species by astilbin isolated from Dimorphandra mollis (Fabaceae, Caesalpinioideae). Biol. Res. 2010; 43:63–74. [PubMed: 21157633]

16. Gao K, Henning SM, Niu Y, Youssefian AA, Seeram NP, Xu A, Heber D. The citrus flavonoid naringenin stimulates DNA repair in prostate cancer cells. J. Nutr. Biochem. 2006; 17:89–95. [PubMed: 16111881]

17. Min K, Ebeler SE. Quercetin inhibits hydrogen peroxide-induced DNA damage and enhances DNA repair in Caco-2 cells. Food Chem. Toxicol. 2009; 47:2716–2722. [PubMed: 19651184]

18. Lee WJ, Shim JY, Zhu BT. Mechanisms for the inhibition of DNA methyltransferases by tea catechins and bioflavonoids. Mol. Pharmacol. 2005; 68:1018–1030. [PubMed: 16037419]

19. Mitchell AE, Burns SA, Rudolf JL. Isozyme- and gender-specific induction of glutathione S-transferases by flavonoids. Arch. Toxicol. 2007; 81:777–784. [PubMed: 17503020]

20. Singh-Gupta V, Joiner MC, Runyan L, Yunker CK, Sarkar FH, Miller S, Gadgeel SM, Konski AA, Hillman GG. Soy isoflavones augment radiation effect by inhibiting APE1/Ref-1 DNA repair activity in non-small cell lung cancer. J. Thorac. Oncol. 2011; 6:688–698. [PubMed: 21325978]

21. Fang M, Chen D, Yang CS. Dietary polyphenols may affect DNA methylation. J. Nutr. 2007; 137:223S–228S. [PubMed: 17182830]

22. Inoue M, Robien K, Wang R, Van Den Berg DJ, Koh WP, Yu MC. Green tea intake, MTHFR/TYMS genotype and breast cancer risk: the Singapore Chinese Health Study. Carcinogenesis. 2008; 29:1967–1972. [PubMed: 18669903]

23. Wang Q, Li H, Tao P, Wang YP, Yuan P, Yang CX, Li JY, Yang F, Lee H, Huang Y. Soy isoflavones, CYP1A1, CYP1B1, and COMT polymorphisms, and breast cancer: a case-control study in southwestern China. DNA Cell Biol. 2011; 30:585–595. [PubMed: 21438753]

24. Hung RJ, Brennan P, Malaveille C, Porru S, Donato F, Boffetta P, Witte JS. Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer. Cancer Epidemiol Biomarkers Prev. 2004; 13:1013–1021. [PubMed: 15184258]

25. Witte JS, Greenland S, Haile RW, Bird CL. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. Epidemiology. 1994; 5:612–621. [PubMed: 7841243]

26. Gammon MD, Neugut AI, Santella RM, Teitelbaum SL, Britton JA, Terry MB, Eng SM, Wolff MS, Stellman SD, Kabat GC, Levin B, Bradlow HL, Hatch M, Beyea J, Camann D, Trent M, Senie RT, Garbowski GC, Maffeo C, Montalvan P, Berkowitz GS, Kemeny M, Citron M, Schnabe F, Schuss A, Hajdu S, Vincguerra V, Collman GW, Obrams GI. The Long Island Breast Cancer Study Project: description of a multi-institutional collaboration to identify environmental risk factors for breast cancer. Breast Cancer Res. Treat. 2002; 74:235–254. [PubMed: 12206514]

27. Waksberg J. Sampling methods for random digit dialing. J Amer Statistic Assoc. 1978; 40

28. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am. J. Epidemiol. 1986; 124:453–469. [PubMed: 3740045]

29. Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. J. Clin. Epidemiol. 1990; 43:1327–1335. [PubMed: 2254769]

30. Potischman N, Swanson CA, Coates RJ, Weiss HA, Brogan DR, Stanford JL, Schoenberg JB, Gammon MD, Brinton LA. Dietary relationships with early onset (under age 45) breast cancer in a case-control study in the United States: influence of chemotherapy treatment. Cancer Causes Control. 1997; 8:713–721. [PubMed: 9328193]

31. Fink BN, Steck SE, Wolff MS, Kabat GC, Gammon MD. Construction of a flavonoid database for assessing intake in a population-based sample of women on Long Island, New York. Nutr. Cancer. 2006; 56:57–66. [PubMed: 17176218]

32. Gammon MD, Santella RM, Neugut AI, Eng SM, Teitelbaum SL, Paykin A, Levin B, Terry MB, Young TL, Wang LW, Wang Q, Britton JA, Wolff MS, Stellman SD, Hatch M, Kabat GC, Senie R, Garbowski G, Maffeo C, Montalvan P, Berkowitz G, Kemeny M, Citron M, Schnabel F, Schuss A, Hajdu S, Vinceguerra V. Environmental toxins and breast cancer on Long Island. I. Polycyclic aromatic hydrocarbon DNA adducts. Cancer Epidemiol. Biomarkers Prev. 2002; 11:677–685. [PubMed: 12163319]
33. Borras C, Gambini J, Gomez-Cabrera MC, Sastre J, Pallardo FV, Mann GE, Vina J. Genistein, a soy isoflavone, up-regulates expression of antioxidant genes: involvement of estrogen receptors, ERK1/2, and NFkappaB. FASEB J. 2006; 20:2136–2138. [PubMed: 16966488]
34. Wiegand H, Boesch-Saadatmandi C, Regos I, Treutter D, Wolffram S, Rimbach G. Effects of quercetin and catechin on hepatic glutathione-S transferase (GST), NAD(P)H quinone oxidoreductase 1 (NQO1), and antioxidant enzyme activity levels in rats. Nutr. Cancer. 2009; 61:717–722. [PubMed: 19838946]
35. Manach C, Scalbert A, Morand C, Remesy C, Jimenez L. Polyphenols: food sources and bioavailability. Am. J. Clin. Nutr. 2004; 79:727–747. [PubMed: 15113710]
36. Gaudet MM, Gammon MD, Santella RM, Britton JA, Teitelbaum SL, Eng SM, Terry MB, Bensen JT, Schroeder J, Olshan AF, Neugut AI, Ambrosone CB. MnSOD Val-9Ala genotype, pro- and anti-oxidant environmental modifiers, and breast cancer among women on Long Island, New York. Cancer Causes Control. 2005; 16:1225–1234. [PubMed: 16215873]
37. Ahn J, Gammon MD, Santella RM, Gaudet MM, Britton JA, Teitelbaum SL, Terry MB, Neugut AI, Josephy PD, Ambrosone CB. Myeloperoxidase genotype, fruit and vegetable consumption, and breast cancer risk. Cancer Res. 2004; 64:7634–7639. [PubMed: 15492293]
38. Ahn J, Gammon MD, Santella RM, Gaudet MM, Britton JA, Teitelbaum SL, Terry MB, Nowell S, Davis W, Garza C, Neugut AI, Ambrosone CB. Associations between breast cancer risk and the catalase genotype, fruit and vegetable consumption, and supplement use. Am J Epidemiol. 2005; 162:943–952. [PubMed: 16192345]
39. Gaudet MM, Bensen JT, Schroeder J, Olshan AF, Terry MB, Eng SM, Teitelbaum SL, Britton JA, Lehman TA, Neugut AI, Ambrosone CB, Santella RM, Gammon MD. Catechol-O-methyltransferase haplotypes and breast cancer among women on Long Island, New York. Breast Cancer Res. Treat. 2006; 99:235–240. [PubMed: 16596327]
40. Ahn J, Gammon MD, Santella RM, Gaudet MM, Britton JA, Teitelbaum SL, Terry MB, Neugut AI, Eng SM, Zhang Y, Garza C, Ambrosone CB. Effects of glutathione S-transferase A1 (GSTA1) genotype and potential modifiers on breast cancer risk. Carcinogenesis. 2006; 27:1876–1882. [PubMed: 16624829]
41. Steck SE, Gaudet MM, Britton JA, Teitelbaum SL, Terry MB, Neugut AI, Santella RM, Gammon MD. Interactions among GSTM1, GSTT1 and GSTP1 polymorphisms, cruciferous vegetable intake and breast cancer risk. Carcinogenesis. 2007; 28:1954–1959. [PubMed: 17693660]
42. Crew KD, Gammon MD, Terry MB, Zhang FF, Zablotska LB, Agrawal M, Shen J, Long CM, Eng SM, Sagiv SK, Teitelbaum SL, Neugut AI, Santella RM. Polymorphisms in nucleotide excision repair genes, polycyclic aromatic hydrocarbon-DNA adducts, and breast cancer risk. Cancer Epidemiol. Biomarkers Prev. 2007; 16:2033–2041. [PubMed: 17932351]
43. Chen J, Gammon MD, Chan W, Palomeque C, Wetmur JG, Kabat GC, Teitelbaum SL, Britton JA, Terry MB, Neugut AI, Santella RM. One-carbon metabolism, MTHFR polymorphisms, and risk of breast cancer. Cancer Res. 2005; 65:1606–1614. [PubMed: 15735051]
44. Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat. Genet. 1995; 10:111–113. [PubMed: 7647779]
45. McCarty KM, Santella RM, Steck SE, Cleveland RJ, Ahn J, Ambrosone CB, North K, Sagiv SK, Eng SM, Teitelbaum SL, Neugut AI, Gammon MD. PAH-DNA adducts, cigarette smoking, GST polymorphisms, and breast cancer risk. Environ. Health Perspect. 2009; 117:552–558. [PubMed: 19440493]
46. Bastaki M, Huen K, Manzanillo P, Chande N, Chen C, Balmes JR, Tager IB, Holland N. Genotype-activity relationship for Mn-superoxide dismutase, glutathione peroxidase 1 and catalase in humans. Pharmacogenet Genomics. 2006; 16:279–286. [PubMed: 16538174]

47. Hansson M, Olsson I, Nauseef WM. Biosynthesis, processing, and sorting of human myeloperoxidase. Arch. Biochem. Biophys. 2006; 445:214–224. [PubMed: 16183032]
48. Forsberg L, Lyrenas L, de Faire U, Morgenstern R. A common functional C-T substitution polymorphism in the promoter region of the human catalase gene influences transcription factor binding, reporter gene transcription and is correlated to blood catalase levels. Free Radic. Biol. Med. 2001; 30:500–505. [PubMed: 11182520]
49. Morel F, Rauch C, Coles B, Le Ferrec E, Guillouzo A. The human glutathione transferase alpha locus: genomic organization of the gene cluster and functional characterization of the genetic polymorphism in the hGSTA1 promoter. Pharmacogenetics. 2002; 12:277–286. [PubMed: 12042665]
50. Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, Autrup JL, Baranova H, Bathum L, Benhamou S, Boffetta P, Bouchardy C, Breskvar K, Brockmoller J, Cascorbi I, Clapper ML, Coutelle C, Daly A, Dell'Omo M, Dolzan V, Dresler CM, Fryer A, Haugen A, Hein DW, Hildesheim A, Hirvonen A, Hsieh LL, Ingelman-Sundberg M, Kalina I, Kang D, Kihara M, Kiyohara C, Kremers P, Lazarus P, Le Marchand L, Lechner MC, van Lieshout EM, London S, Manni JJ, Maugard CM, Morita S, Nazar-Stewart V, Noda K, Oda Y, Parl FF, Pastorelli R, Persson I, Peters WH, Rannug A, Rebbeck T, Risch A, Roelandt L, Romkes M, Ryberg D, Salagovic J, Schoket B, Seidegard J, Shields PG, Sim E, Sinnet D, Strange RC, Stucker I, Sugimura H, To-Figueras J, Vineis P, Yu MC, Taioli E. Metabolic gene polymorphism frequencies in control populations. Cancer Epidemiol. Biomarkers Prev. 2001; 10:1239–1248. [PubMed: 11751440]
51. Lin HJ, Johansson AS, Stenberg G, Materi AM, Park JM, Dai A, Zhou H, Gim JS, Kau IH, Hardy SI, Parker MW, Mannervik B. Naturally occurring Phe151Leu substitution near a conserved folding module lowers stability of glutathione transferase P1-1. Biochim. Biophys. Acta. 2003; 1649:16–23. [PubMed: 12818186]
52. van der Put NM, Gabreels F, Stevens EM, Smeitink JA, Trijbels FJ, Eskes TK, van den Heuvel LP, Blom HJ. A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects? Am. J. Hum. Genet. 1998; 62:1044–1051. [PubMed: 9545395]
53. Lachman HM, Papolos DF, Saito T, Yu YM, Szumlanski CL, Weinshilboum RM. Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. Pharmacogenetics. 1996; 6:243–250. [PubMed: 8807664]
54. Syvanen AC, Tilgmann C, Rinne J, Ulmanen I. Genetic polymorphism of catechol-O-methyltransferase (COMT): correlation of genotype with individual variation of S-COMT activity and comparison of the allele frequencies in the normal population and parkinsonian patients in Finland. Pharmacogenetics. 1997; 7:65–71. [PubMed: 9110364]
55. Bagley PJ, Selhub J. A common mutation in the methylenetetrahydrofolate reductase gene is associated with an accumulation of formylated tetrahydrofolates in red blood cells. Proc. Natl. Acad. Sci. U. S. A. 1998; 95:13217–13220. [PubMed: 9789068]
56. Chen P, Wiencke J, Aldape K, Kesler-Diaz A, Miike R, Kelsey K, Lee M, Liu J, Wrensch M. Association of an ERCC1 polymorphism with adult-onset glioma. Cancer Epidemiol. Biomarkers Prev. 2000; 9:843–847. [PubMed: 10952103]
57. Sturgis EM, Dahlstrom KR, Spitz MR, Wei Q. DNA repair gene ERCC1 and ERCC2/XPD polymorphisms and risk of squamous cell carcinoma of the head and neck. Arch. Otolaryngol. Head. Neck. Surg. 2002; 128:1084–1088. [PubMed: 12220217]
58. Cheng L, Sturgis EM, Eicher SA, Spitz MR, Wei Q. Expression of nucleotide excision repair genes and the risk for squamous cell carcinoma of the head and neck. Cancer. 2002; 94:393–397. [PubMed: 11900225]
59. Butkiewicz D, Popanda O, Risch A, Edler L, Dienemann H, Schulz V, Kayser K, Drings P, Bartsch H, Schmezer P. Association between the risk for lung adenocarcinoma and a (−4) G-to-A polymorphism in the XPA gene. Cancer Epidemiol. Biomarkers Prev. 2004; 13:2242–2246. [PubMed: 15598786]

60. Wu X, Zhao H, Wei Q, Amos CI, Zhang K, Guo Z, Qiao Y, Hong WK, Spitz MR. XPA polymorphism associated with reduced lung cancer risk and a modulating effect on nucleotide excision repair capacity. Carcinogenesis. 2003; 24:505–509. [PubMed: 12663511]
61. Hu Z, Wei Q, Wang X, Shen H. DNA repair gene XPD polymorphism and lung cancer risk: a meta-analysis. Lung Cancer. 2004; 46:1–10. [PubMed: 15364127]
62. Vogel U, Hedayati M, Dybdahl M, Grossman L, Nexo BA. Polymorphisms of the DNA repair gene XPD: correlations with risk of basal cell carcinoma revisited. Carcinogenesis. 2001; 22:899–904. [PubMed: 11375896]
63. Rybicki BA, Conti DV, Moreira A, Cicek M, Casey G, Witte JS. DNA repair gene XRCC1 and XPD polymorphisms and risk of prostate cancer. Cancer Epidemiol. Biomarkers Prev. 2004; 13:23–29. [PubMed: 14744728]
64. Smith TR, Levine EA, Perrier ND, Miller MS, Freimanis RI, Lohman K, Case LD, Xu J, Mohrenweiser HW, Hu JJ. DNA-repair genetic polymorphisms and breast cancer risk. Cancer Epidemiol. Biomarkers Prev. 2003; 12:1200–1204. [PubMed: 14652281]
65. Jeon HS, Kim KM, Park SH, Lee SY, Choi JE, Lee GY, Kam S, Park RW, Kim IS, Kim CH, Jung TH, Park JY. Relationship between XPG codon 1104 polymorphism and risk of primary lung cancer. Carcinogenesis. 2003; 24:1677–1681. [PubMed: 12869423]
66. Kumar R, Hoglund L, Zhao C, Forsti A, Snellman E, Hemminki K. Single nucleotide polymorphisms in the XPG gene: determination of role in DNA repair and breast cancer risk. Int. J. Cancer. 2003; 103:671–675. [PubMed: 12494477]
67. Poole C. Low P-values or narrow confidence intervals: which are more durable? Epidemiology. 2001; 12:291–294. [PubMed: 11337599]
68. Selvin, S. Statistical analysis of epidemiologic data. Oxford ; New York: Oxford University Press; 2004.
69. Rothman, KJ.; Greenland, S.; Lash, TL. Modern epidemiology. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
70. Witte JS, Greenland S. Simulation study of hierarchical regression. Stat. Med. 1996; 15:1161–1170. [PubMed: 8804145]
71. Desantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics, 2011. CA Cancer. J. Clin. 2011
72. Spencer JP, Abd El Mohsen MM, Minihane AM, Mathers JC. Biomarkers of the intake of dietary polyphenols: strengths, limitations and application in nutrition research. Br. J. Nutr. 2008; 99:12–22. [PubMed: 17666146]

Table 1

Z-matrix for hierarchical logistic regression model for the flavonoid-gene interaction in association with breast cancer risk in LIBCSP using an indicator-term approach

No.	Indicator Variable[1]	Gene Function	Flavonoid Intake Low = 0 High = 1	↑ROS[2]	↓Phase 2 activity	↓DNA repair activity	↓Methyl group formation
1	mnsod_flav_ind1	Beneficial	0	0	0	0	0
2	mnsod_flav_ind2	Deleterious	1	1	0	0	0
3	mnsod_flav_ind3	Beneficial	1	0	0	0	0
4	mpo_flav_ind1	Beneficial	0	0	0	0	0
5	mpo_flav_ind2	Deleterious	1	1	0	0	0
6	mpo_flav_ind3	Beneficial	1	0	0	0	0
7	cat_flav_ind1	Beneficial	0	0	0	0	0
8	cat_flav_ind2	Deleterious	1	1	0	0	0
9	cat_flav_ind3	Beneficial	1	0	0	0	0
10	gsta1_flav_ind1	Beneficial	0	0	0	0	0
11	gsta1_flav_ind2	Deleterious	1	1	1	0	0
12	gsta1_flav_ind3	Beneficial	1	0	0	0	0
13	gstm1_flav_ind1	Beneficial	0	0	0	0	0
14	gstm1_flav_ind2	Deleterious	1	0	1	0	0
15	gstm1_flav_ind3	Beneficial	1	0	0	0	0
16	gstp1_flav_ind1	Beneficial	0	0	0	0	0
17	gstp1_flav_ind2	Deleterious	1	0	1	0	0
18	gstp1_flav_ind3	Beneficial	1	0	0	0	0
19	gstt1_flav_ind1	Beneficial	0	0	0	0	0
20	gstt1_flav_ind2	Deleterious	1	0	1	0	0
21	gstt1_flav_ind3	Beneficial	1	0	0	0	0
22	ercc1_flav_ind1	Beneficial	0	0	0	0	0
23	ercc1_flav_ind2	Deleterious	1	0	0	1	0
24	ercc1_flav_ind3	Beneficial	1	0	0	0	0
25	xpa_flav_ind1	Beneficial	0	0	0	0	0
26	xpa_flav_ind2	Deleterious	1	0	0	1	0

Cancer Causes Control. Author manuscript; available in PMC 2015 February 01.

No.	Indicator Variable[1]	Gene Function	Flavonoid Intake Low = 0 High = 1	↑ROS[2]	↓Phase 2 activity	↓DNA repair activity	↓Methyl group formation
27	xpa_flav_ind3	Beneficial	1	0	0	0	0
28	xpd312_flav_ind1	Beneficial	0	0	0	0	0
29	xpd312_flav_ind2	Deleterious	1	0	0	1	0
30	xpd312_flav_ind3	Beneficial	1	0	0	0	0
31	mthfr677_flav_ind1	Beneficial	0	0	0	0	0
32	mthfr677_flav_ind2	Deleterious	1	0	0	0	1
33	mthfr677_flav_ind3	Beneficial	1	0	0	0	0
34	mthfr1298_flav_ind1	Beneficial	0	0	0	0	0
35	mthfr1298_flav_ind2	Deleterious	1	0	0	0	1
36	mthfr1298_flav_ind3	Beneficial	1	0	0	0	0
37	comt_flav_ind1	Beneficial	0	0	0	0	0
38	comt_flav_ind2	Deleterious	1	−1	0	0	1
39	comt_flav_ind3	Beneficial	1	0	0	0	0

[1]Referent group for the single-referent analysis is “Deleterious” gene function, and “Low” flavonoid intake.

[2]ROS = reactive oxygen species

Table 2

Random effects odds ratios (OR), 95% confidence intervals (95% CI), and confidence limit ratios (CLR) from conventional models (Single GxE, Multivariate) and hierarchical logistic regression models for flavonoid-gene interaction in association with breast cancer risk in the LIBCSP (N=1778)

Gene Variant rs#	Variant Function	Flavonoid Intake[1]	Case	Control	Single G×E			Multivariate			Hierarchical		
					OR[2]	95% CI	CLR[3]	OR[2]	95% CI	CLR[3]	OR[2]	95% CI	CLR[3]
MnSOD	Deleterious	Low	112	105	1.00			1.00			1.00		
rs1799725	Beneficial	Low	334	347	1.09	0.80, 1.48	1.85	1.07	0.78, 1.47	1.87	1.06	0.80, 1.41	1.77
	Deleterious	High	112	107	0.98	0.67, 1.43	2.13	1.19	0.47, 3.01	6.37	0.93	0.54, 1.63	3.03
	Beneficial	High	345	316	0.95	0.70, 1.30	1.86	1.18	0.48, 2.90	6.01	0.93	0.54, 1.58	2.92
MPO	Deleterious	Low	259	284	1.00			1.00			1.00		
rs2333227	Beneficial	Low	187	168	0.82	0.63, 1.08	1.71	0.83	0.63, 1.09	1.73	0.85	0.66, 1.09	1.66
	Deleterious	High	276	256	0.84	0.65, 1.07	1.63	0.91	0.39, 2.11	5.40	0.93	0.54, 1.61	2.98
	Beneficial	High	181	167	0.82	0.63, 1.08	1.73	0.91	0.39, 2.14	5.55	0.93	0.55, 1.59	2.92
CAT	Deleterious	Low	160	177	1.00			1.00			1.00		
rs1001179	Beneficial	Low	286	275	0.86	0.66, 1.13	1.72	0.87	0.66, 1.15	1.74	0.89	0.69, 1.15	1.67
	Deleterious	High	162	164	0.89	0.66, 1.22	1.85	1.13	0.47, 2.71	5.82	1.00	0.58, 1.73	2.99
	Beneficial	High	295	259	0.78	0.59, 1.03	1.73	0.96	0.40, 2.31	5.77	0.87	0.51, 1.48	2.91
GSTA1	Deleterious	Low	274	310	1.00			1.00			1.00		
rs3957356	Beneficial	Low	172	142	0.76	0.57, 1.00	1.74	0.78	0.59, 1.04	1.76	0.81	0.62, 1.05	1.69
	Deleterious	High	305	273	0.79	0.63, 1.00	1.60	0.81	0.35, 1.87	5.31	0.89	0.51, 1.56	3.04
	Beneficial	High	152	150	0.86	0.65, 1.14	1.76	0.86	0.37, 2.01	5.43	0.95	0.55, 1.63	2.94
GSTM1	Deleterious	Low	189	225	1.00			1.00			1.00		
gene deletion	Beneficial	Low	257	227	0.72	0.55, 0.94	1.70	0.73	0.56, 0.96	1.72	0.76	0.59, 0.98	1.65
	Deleterious	High	229	206	0.73	0.55, 0.96	1.73	0.77	0.32, 1.82	5.68	0.89	0.51, 1.55	3.02
	Beneficial	High	228	217	0.78	0.59, 1.02	1.73	0.81	0.34, 1.93	5.69	0.94	0.55, 1.61	2.93
GSTP1	Deleterious	Low	245	218	1.00			1.00			1.00		
rs1695	Beneficial	Low	201	234	1.31	1.01, 1.71	1.70	1.30	0.99, 1.70	1.71	1.25	0.98, 1.61	1.65
	Deleterious	High	221	217	1.09	0.83, 1.42	1.70	1.60	0.69, 3.76	5.48	0.97	0.56, 1.68	3.01
	Beneficial	High	236	206	0.97	0.74, 1.26	1.70	1.43	0.60, 3.37	5.58	0.87	0.51, 1.49	2.94
GSTT1	Deleterious	Low	98	104	1.00			1.00			1.00		
gene deletion	Beneficial	Low	348	348	0.96	0.70, 1.31	1.88	0.97	0.70, 1.33	1.90	0.97	0.73, 1.30	1.79

Gene Variant rs#	Variant Function	Flavonoid Intake[1]	Case	Control	Single G×E			Multivariate			Hierarchical		
					OR[2]	95% CI	CLR[3]	OR[2]	95% CI	CLR[3]	OR[2]	95% CI	CLR[3]
	Deleterious	High	101	86	0.82	0.55, 1.23	2.23	0.96	0.38, 2.47	6.55	0.88	0.50, 1.54	3.09
	Beneficial	High	356	337	0.88	0.64, 1.21	1.88	1.06	0.43, 2.64	6.14	0.96	0.56, 1.65	2.94
ERCC1	Deleterious	Low	205	215	1.00			1.00			1.00		
rs3212986	Beneficial	Low	241	237	0.93	0.72, 1.22	1.69	0.97	0.72, 1.29	1.79	0.97	0.74, 1.26	1.70
	Deleterious	High	200	206	0.96	0.73, 1.27	1.69	1.17	0.47, 2.90	6.17	1.10	0.63, 1.93	3.07
	Beneficial	High	257	217	0.79	0.61, 1.04	1.70	1.07	0.44, 2.57	6.17	0.99	0.57, 1.72	3.00
XPA	Deleterious	Low	387	404	1.00			1.00			1.00		
rs1800975	Beneficial	Low	59	48	0.79	0.53, 1.19	2.26	0.75	0.50, 1.14	2.30	0.82	0.57, 1.16	2.04
	Deleterious	High	395	381	0.91	0.74, 1.11	1.50	1.20	0.61, 2.37	3.90	1.24	0.69, 2.24	3.24
	Beneficial	High	62	42	0.64	0.42, 0.97	2.31	0.83	0.38, 1.80	4.74	0.88	0.50, 1.54	3.06
XPD	Deleterious	Low	244	271	1.00			1.00			1.00		
rs1799793	Beneficial	Low	202	181	0.80	0.61, 1.04	1.71	0.78	0.58, 1.04	1.80	0.81	0.62, 1.06	1.71
	Deleterious	High	242	254	0.93	0.72, 1.19	1.65	1.12	0.46, 2.69	5.80	1.19	0.68, 2.09	3.07
	Beneficial	High	215	169	0.69	0.53, 0.90	1.71	0.86	0.36, 2.03	5.65	0.92	0.53, 1.59	3.00
MTHFR	Deleterious	Low	273	279	1.00			1.00			1.00		
C677T	Beneficial	Low	173	173	0.97	0.74, 1.28	1.72	1.01	0.76, 1.33	1.75	1.00	0.77, 1.30	1.68
rs1801133	Deleterious	High	271	264	0.94	0.74, 1.19	1.62	1.23	0.54, 2.78	5.10	0.97	0.55, 1.70	3.10
	Beneficial	High	186	159	0.82	0.62, 1.08	1.73	1.11	0.48, 2.57	5.36	0.88	0.51, 1.51	2.96
MTHFR	Deleterious	Low	224	221	1.00			1.00			1.00		
A1298C	Beneficial	Low	222	231	1.05	0.81, 1.37	1.69	1.11	0.84, 1.46	1.73	1.09	0.84, 1.40	1.66
rs1801131	Deleterious	High	234	191	0.81	0.62, 1.06	1.72	0.96	0.40, 2.26	5.59	0.82	0.47, 1.45	3.09
	Beneficial	High	223	232	1.03	0.79, 1.35	1.70	1.22	0.52, 2.90	5.62	1.03	0.60, 1.78	2.96
COMT	Deleterious	Low	113	128	1.00			1.00			1.00		
rs4680	Beneficial	Low	333	324	0.87	0.65, 1.17	1.81	0.84	0.62, 1.14	1.84	0.87	0.66, 1.14	1.74
	Deleterious	High	119	117	0.85	0.59, 1.23	2.07	1.01	0.41, 2.50	6.12	0.96	0.38, 2.45	6.48
	Beneficial	High	338	306	0.80	0.59, 1.07	1.82	0.93	0.38, 2.25	5.88	0.88	0.49, 1.58	3.20

[1] Low flavonoid intake ≤ 150 mg/day, high flavonoid intake > 150 mg/day.

[2] All models adjusted for age and total energy intake.

[3] CLR = Upper 95% Confidence Limit / Lower 95% Confidence Limit. Calculated using 95% CI estimates.

Table 3

Fixed effects odds ratios (OR) and 95% confidence intervals (95% CI) from hierarchical logistic regression models for flavonoid-gene interaction in association with breast cancer risk in the LIBCSP (N=1778)

Gene pathways	OR[1]	95% CI[2]
Increased oxidative stress	1.00	0.62, 1.63
Reduced Phase 2 Metabolism	0.97	0.58, 1.65
Reduced DNA repair capability	1.27	0.70, 2.29
Reduced methyl group formation	0.98[3]	0.53, 1.82

Note: Odds ratios and 95% confidence intervals estimated using fixed effects coefficients from hierarchical model: logit $(p|\mathbf{X, W, Z}) = \alpha + \mathbf{XZ}\pi + \mathbf{X}\delta + \mathbf{W}\gamma$

[1] $\text{OR} = \exp[\pi]$

[2] $95\%\ \text{CI} = \exp[\pi \pm 1.96*\text{SE}(\pi)]$

[3] Example interpretation for fixed effects OR for reduced methyl group formation: “The average effect of deleterious versus beneficial gene, controlling for average flavonoid intake in the reduced methyl group formation pathway, and adjusted for the three other pathways, results in a 2% reduced risk for breast cancer.”