

# Investigating Multiple Candidate Genes and Nutrients in the Folate Metabolism Pathway to Detect Genetic and Nutritional Risk Factors for Lung Cancer

Michael D. Swartz<sup>1,4\*</sup>, Christine B. Peterson<sup>2</sup>, Philip J. Lupo<sup>3</sup>, Xifeng Wu<sup>4</sup>, Michele R. Forman<sup>4</sup>, Margaret R. Spitz<sup>4,5</sup>, Ladia M. Hernandez<sup>4</sup>, Marina Vannucci<sup>2</sup>, Sanjay Shete<sup>4,6</sup>

**1** Division of Biostatistics, University of Texas School of Public Health, Houston, Texas, United States of America, **2** Department of Statistics, Rice University, Houston, Texas, United States of America, **3** Human Genetics Center, The University of Texas School of Public Health, Houston, Texas, United States of America, **4** Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, **5** Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas, United States of America, **6** Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America

## Abstract

**Purpose:** Folate metabolism, with its importance to DNA repair, provides a promising region for genetic investigation of lung cancer risk. This project investigates genes (*MTHFR*, *MTR*, *MTRR*, *CBS*, *SHMT1*, *TYMS*), folate metabolism related nutrients (B vitamins, methionine, choline, and betaine) and their gene-nutrient interactions.

**Methods:** We analyzed 115 tag single nucleotide polymorphisms (SNPs) and 15 nutrients from 1239 and 1692 non-Hispanic white, histologically-confirmed lung cancer cases and controls, respectively, using stochastic search variable selection (a Bayesian model averaging approach). Analyses were stratified by current, former, and never smoking status.

**Results:** Rs6893114 in *MTRR* (odds ratio [OR]=2.10; 95% credible interval [CI]: 1.20–3.48) and alcohol (drinkers vs. non-drinkers, OR=0.48; 95% CI: 0.26–0.84) were associated with lung cancer risk in current smokers. Rs13170530 in *MTRR* (OR=1.70; 95% CI: 1.10–2.87) and two SNP\*nutrient interactions [betaine\*rs2658161 (OR=0.42; 95% CI: 0.19–0.88) and betaine\*rs16948305 (OR=0.54; 95% CI: 0.30–0.91)] were associated with lung cancer risk in former smokers. SNPs in *MTRR* (rs13162612; OR=0.25; 95% CI: 0.11–0.58; rs10512948; OR=0.61; 95% CI: 0.41–0.90; rs2924471; OR=3.31; 95% CI: 1.66–6.59), and *MTHFR* (rs9651118; OR=0.63; 95% CI: 0.43–0.95) and three SNP\*nutrient interactions (choline\*rs10475407; OR=1.62; 95% CI: 1.11–2.42; choline\*rs11134290; OR=0.51; 95% CI: 0.27–0.92; and riboflavin\*rs8767412; OR=0.40; 95% CI: 0.15–0.95) were associated with lung cancer risk in never smokers.

**Conclusions:** This study identified possible nutrient and genetic factors related to folate metabolism associated with lung cancer risk, which could potentially lead to nutritional interventions tailored by smoking status to reduce lung cancer risk.

**Citation:** Swartz MD, Peterson CB, Lupo PJ, Wu X, Forman MR, et al. (2013) Investigating Multiple Candidate Genes and Nutrients in the Folate Metabolism Pathway to Detect Genetic and Nutritional Risk Factors for Lung Cancer. PLoS ONE 8(1): e53475. doi:10.1371/journal.pone.0053475

**Editor:** Yan Gong, University of Florida, United States of America

**Received:** July 23, 2012; **Accepted:** November 28, 2012; **Published:** January 23, 2013

**Copyright:** © 2013 Swartz et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors would like to acknowledge the funding sources National Cancer Institute (NCI) award K07-CA 123109, and R03 CA141998 (MDS PI), NCI award R01 CA55679 (Lung Cancer Study, MRS PI). The funders had no role in study design, data collection and analysis, decision to publish or preparation of this manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: Michael.D.Swartz@uth.tmc.edu

## Introduction

Lung cancer accounted for 15% of all cancer diagnoses in 2010 and 28% of all cancer deaths [1]. Furthermore, the overall 5-year survival rate remains at 15% for all stages of lung cancer combined [1]. While smoking cigarettes is the dominant risk factor for lung cancer, only a fraction of smokers ever develop the disease [2], suggesting that lung cancer development depends on other factors, most likely genetic and other environmental factors (e.g., diet) [2,3,4]. Research suggests that dietary folate status [5,6] and variation within genes that comprise the folate metabolic pathway [7,8,9,10,11] may be associated with lung cancer risk.

Folate has been shown to be an important nutrient for DNA synthesis, repair, and methylation [8] and therefore, may influence cancer risk. Studies have shown that low folate intake is associated

with increased DNA strand breaks, decreased DNA methylation, and reduced DNA repair capacity [12]. High dietary folate intake (defined as above the median control intake level) is associated with a 40% reduction in lung cancer risk among former smokers, after adjusting for age, sex, ethnicity, total energy intake, body mass index, family history of smoking, pack years smoked, and alcohol consumption [13]. The protective effect of increased folate intake also appears to hold among current smokers [14]. More recently, high blood serum levels of vitamin B<sub>6</sub> and methionine have been found to offer different levels of protection against lung cancer for never, former and current smokers [6].

Aside from dietary folate, genes in the folate metabolic pathway have also been associated with lung cancer risk. Folate genes implicated in lung cancer risk include methylenetetrahydrofolate reductase (*MTHFR*) [7]; thymidylate synthase (*TYMS*) [10,15];

serine hydroxymethyltransferase 1 (*SHMT1*) [11]; and cystathionine  $\beta$ -synthase (*CBS*) [8]. Other suspected genes include methionine synthase (*MTR*) and methionine synthase reductase (*MTRR*); however, results for these genes have been equivocal [7].

This study investigates the roles of folate status, nutrition, genes and gene-nutrient interactions in the folate metabolic pathway in lung cancer risk. Previous assessments of the association between folate status and lung cancer have focused on nutrition, without consideration of genetic polymorphisms, and studies that have assessed genes in the folate metabolic pathway included only small panels of single nucleotide polymorphisms (SNPs). The current study examines multiple SNPs in important folate genes, while also investigating dietary intake of B vitamins, folate, methionine, betaine, and choline, allowing for the joint assessment of the effects of diet and genetic status and pairwise gene-nutrient interactions on lung cancer risk.

This study uses a more powerful approach than standard methods for detecting genes and nutrients associated with lung cancer, known as stochastic search variable selection (SSVS) [16,17]. SSVS, a form of Bayesian model averaging, randomly searches through all possible models, guided by the data, to identify the most likely risk factors accounting for the uncertainty of variable selection [17]. We employed stochastic search here for multiple reasons. First, such stochastic search methods have been effective in analyzing SNP data, particularly in genetic association studies (e.g. [17,18,19,20,21]). Second, simulation studies using the case-control design have demonstrated that SSVS has a greater accuracy to recover the “true model” than standard variable selection methods, such as forward, backward or stepwise selection [17]. Third, other researchers have shown that SSVS outperforms penalized sparse regression [22] and standard lasso techniques [23], especially in problems investigating many SNPs where then number of SNPs and interactions is larger than the sample size [23,24,25].

## Methods

### Ethics Statement

This study was approved by both the MD Anderson and The University of Texas Health Science Center at Houston Institutional Review Boards (IRB). The University of Texas Health Science Center at Houston IRB is also the governing IRB for the UT School of Public Health, as a member school of the University of Texas Health Science Center at Houston. This study involved secondary analysis of de-identified data. The original data collection was consented by written informed consent that discussed such analyses.

### Study Population

The study population consisted of histologically confirmed lung cancer cases ( $n = 1239$ ) and controls ( $n = 1692$ ) diagnosed between 1995 and 2007 from an ongoing lung cancer case-control study conducted at The University of Texas MD Anderson Cancer Center. Details of the recruitment for the parent study have been published elsewhere [26,27]. Briefly, newly diagnosed cases of lung cancer were recruited from patients at MD Anderson Cancer Center. Controls (individuals without a previous diagnosis of cancer, except non-melanoma skin cancer) were recruited from Kelsey-Seybold clinics, the largest private physician group in Houston, Texas. The overall recruitment rate was about 75%.

### Dietary and Demographic Data

Dietary information, demographic factors, and smoking history were obtained through a personal interview. Trained interviewers

administered a food frequency questionnaire (FFQ) that is a modified version of the National Cancer Institute’s Health Habits and History Questionnaire [28]. The FFQ solicited usual intake for the year prior to the interview and included an open-ended food section and behavior-related dietary questions regarding restaurant dining and food preparation. The validity of the Block FFQ has been described across various populations [29,30,31]. The questionnaire was modified for the parent study to include ethnic foods commonly consumed in the greater metropolitan area of Houston, Texas. The estimated intake of several nutrients and beverages in controls was comparable to that reported by adults who participated in the National Health and Nutrition Examination Survey (NHANES), 1999–2000 [32,33,34].

In the current study, we limit our analysis to non-Hispanic whites to ensure a large enough sample size to stratify by smoking status and minimize confounding due to population stratification. We also focus on this population to keep our inference consistent with that of the earlier studied Spitz model for non-genetic risk of lung cancer [35].

Questions with multiple foods on the FFQ were weighted for each individual food item by the consumption of that food in the NHANES population (as in [36]). Then, the nutrient content of each food was derived from the U.S. Department of Agriculture National Nutrient Database for Standard Reference, Release 21 (USDA SR21) [37]. Thus, weighted averages of the individual foods in multiple food items as well as the items with one food were linked to the USDA SR21 to calculate nutrient intakes. We determined the daily nutrient intake of the following macro- and micronutrients: energy, carbohydrate, fat, protein, betaine, choline, methionine, folate, pantothenic acid (vitamin B<sub>5</sub>), niacin (vitamin B<sub>3</sub>), riboflavin (vitamin B<sub>2</sub>), thiamin (vitamin B<sub>1</sub>), vitamin B<sub>6</sub>, and vitamin B<sub>12</sub>. All nutrient values were adjusted for total energy intake per the method of Willett and Stampfer [38].

In our analysis, alcohol consumption reported on the FFQ was first dichotomized into non-drinkers (reported 0 drinks on the FFQ) versus drinkers (reported any drinking). After significance was assessed, we later categorized alcohol into: nondrinker, 0.1–4.9 g/day, 5.0–14.9 g/day, 15–29.9 g/day, and greater than 30 g/day as in [39] for comparative discussion purposes. We categorized smoking status as never (smoked fewer than 100 cigarettes in their lifetime), former (smoked at least 100 cigarettes in their lifetime and quit more than 1 year prior to study enrollment), or current (currently smoking or quit less than 1 year prior to enrollment) smokers [35]. Family history of smoking-related cancer in a first-degree relative was included in the analysis on the basis of a yes/no response. For all smokers, we computed pack years. For former smokers, we additionally computed years since cessation. For never and former smokers, we recorded exposure to environmental tobacco smoke, defined as exposure to someone else’s cigarette smoke at work or at home on a regular basis, as described in [35].

### SNP Selection and Genotyping

We selected 293 SNPs across the genes in the folate metabolism pathway. The full panel of SNPs genotyped, their function and location are given in Table S1. These SNPs consist of all those listed in the HapMap and National Institute of Environmental Health Sciences (NIEHS) SNP databases as members of the above-mentioned genes of the folate metabolism pathway. We consider a SNP to belong to a gene if it is located within 500 kilo base pairs (kb) of the gene. No other genes with known function in folate metabolism have been implicated in lung cancer risk, so we focus our analyses on this set. Thus, our custom chip was composed of SNPs from these genes in the folate pathway. The

selected SNPs were genotyped using the custom iSelect Infinium Beadchip design in conjunction with SNPs for other projects.

Participants' genomic DNA was extracted from peripheral blood lymphocytes and stored at  $-80^{\circ}\text{C}$  until use. We genotyped SNPs from case and control DNA samples using Illumina's BeadXpress platform according to the standard 3 day protocol. Genotypes were autocalled using the BeadStudio software. SNPs with genotype call rates of less than 95% or SNPs with a minor allele frequency less than 0.01, or more than 10% missing across our data set were omitted from our analysis (39 SNPs removed). A chi-square test confirmed that the pattern of missing observations for each SNP was independent of the affection status of the subjects. For the multivariable analysis, individuals missing SNPs were removed, and 2,225 subjects (1175 cases, 1050 controls) remained in the analysis.

Once the data set was reduced to genotypes not missing SNPs, we reduced the dimensionality and collinearity by empirically selecting tag SNPs using the method of Carlson et al. [40], with a threshold of  $r^2 = 0.8$ . We selected representative tag SNPs that were in exons, or previously mentioned in prior studies, when available. We examined Hardy-Weinberg proportions for the tag SNPs using PLINK [41], and all tag SNPs were found to be in Hardy-Weinberg equilibrium at the 0.001 level.

### Statistical Analysis

Continuous demographic variables were compared using two-sample t-tests; nutrient variables were compared using Wilcoxon rank-sum tests; and categorical demographic variables were compared with Pearson's chi-squared test. For model selection, we used a Bayesian model averaging method known as stochastic search variable selection (SSVS) [16] applied to logistic regression [17,20]. SSVS uses Markov Chain Monte Carlo (MCMC) methods to search through all possible models to identify joint genetic and dietary effects on lung cancer risk. These methods have been shown to be more powerful than traditional stepwise selection methods [17,20]. SSVS has two levels of prior distributions: the prior on the model coefficient or odds ratio, which includes a correlation matrix for genetic factors defined by linkage disequilibrium (LD), and the prior for probability of selection for each variable [17,20].

**Prior Distribution.** To conform with fully Bayesian methods, we modeled the prior correlation among the tag SNPs to be analyzed using the pairwise  $r^2$  values from the NIEHS Environmental Genome Project [42], external to our data. SNPs more than 400,000 base pairs apart or from different chromosomes were assumed to be independent. As a reliable estimate for the correlation of nutrient values could not be externally defined, the priors for the dietary coefficients were independent normal distributions centered at 0 [17]. We assumed gene-environment independence and modeled the priors for the coefficients for the genes and nutritional covariates as uncorrelated. We also used a prior probability of inclusion of 0.5 for each variable, which has been shown to best control for both false-positive and false-negative results when the prior information for all risk factors may not be available [17].

**Smoking Variables and Additional Covariates.** Because smoking is a well-established risk factor for lung cancer, cases and controls were frequency matched by smoking status in the original study design. Thus, we stratified subjects into three groups based on smoking history: never smokers, former smokers, and current smokers as in Spitz et al. [35]. Other non-genetic risk factors for lung cancer that have been established as significant were included in each model following the approach in Spitz et al. [35], and were not subject to variable selection. For never smokers, we included

sex, age, family history of cancer, and exposure to environmental tobacco smoke. For former smokers, we included sex, age, family history of cancer, and a factor indicating whether the subject stopped smoking before age 40, between ages 40 and 53, or at age 54 or later, selected for its stronger association with lung cancer risk than pack years smoked (as described in [35]). For current smokers, we included sex, age, family history of cancer, and a factor indicating whether pack years smoked were less than 27, between 27 and 53, between 54 and 82, or 83 or greater, as in [35]. Genotypes were coded additively, using homozygous for the major allele as the reference genotype.

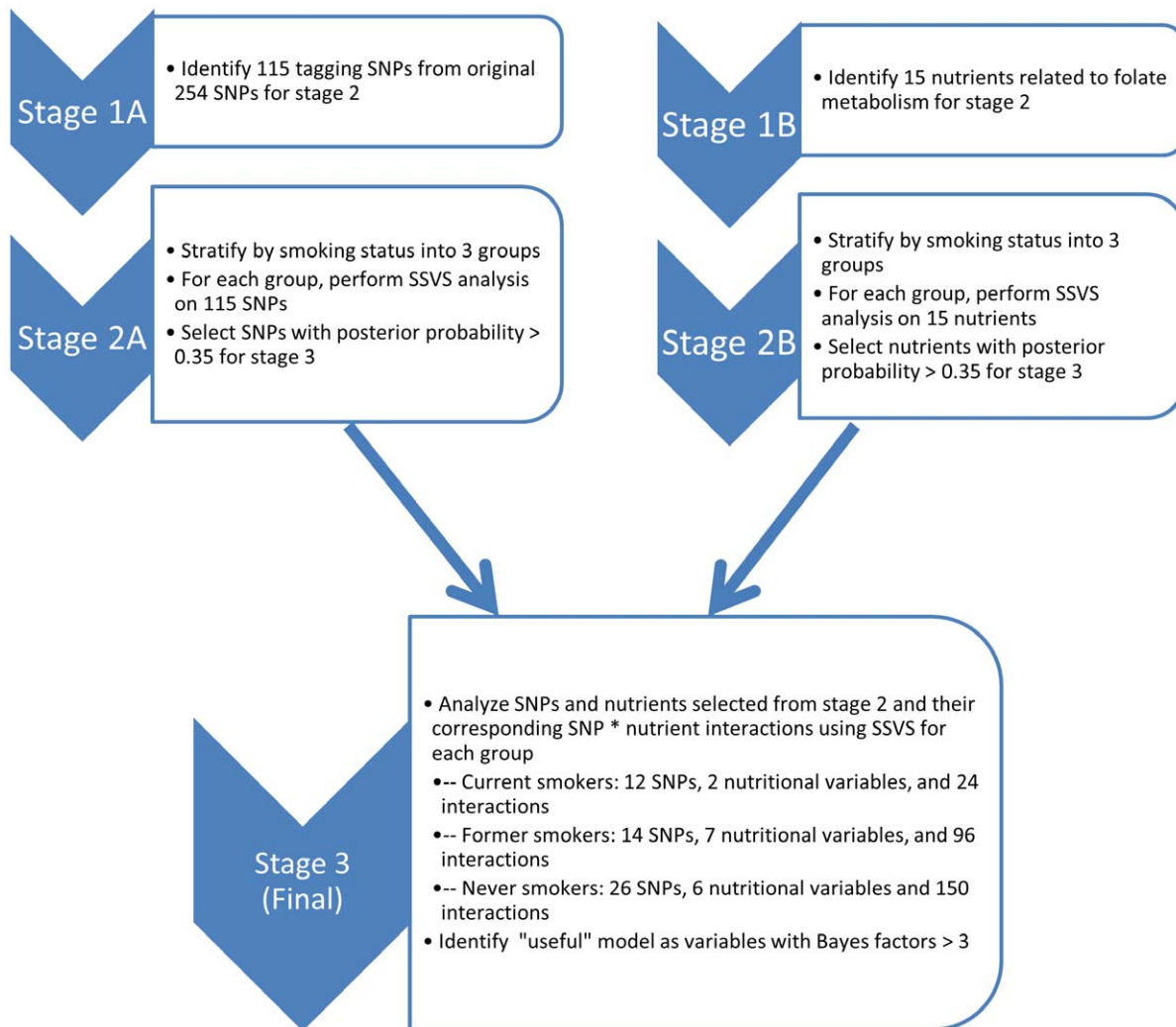
**Markov Chain Monte Carlo Analysis.** All MCMC computations were completed using WinBugs [43], R and the R2WinBUGS package [44] to prepare the data (categorization, clean missing, stratification) and to compute posterior inference. We ran two chains with distinct starting values for 300,000 iterations and used the last two thirds of the iterations to estimate our posterior quantities to ensure convergence to the stationary distribution, as described in [45]. The two chains for each stratum were found to have very high correlations, indicating that they converged to the same model, and were pooled for inference.

The statistical analysis proceeded in three stages (see Figure 1): In stage 1, we identified the tag SNPs and nutrients to be analyzed. Stage 2 consisted of a stochastic search of the SNPs and a separate stochastic search of the nutrients to screen for promising SNPs and nutrients. Any SNPs and nutrients with a posterior probability of inclusion greater than 0.35 proceeded to stage 3 [17,46]. In stage 3, we jointly searched through SNPs and nutrients and their corresponding SNP and nutrient interactions for those SNPs and nutrients that proceeded to stage 3. The gene-nutrient interactions consisted of the pairwise product of each SNP additively coded and the nutrient as a continuous variable. For the stage 3 model, we selected genes, nutrients and interactions with a marginal Bayes factor greater than 3 which indicates moderate evidence for association [47] (marginal Bayes factors were computed similar to the SNP specific Bayes factor in [48], based on the marginal probabilities of inclusion). In addition to computing the Bayes factor, we also computed the expected false discovery rate, as defined in [49]. The Bayes factor of 3 or greater also coincides with controlling the false discovery rate to less than 0.15. We estimated the odds ratios (ORs) using the posterior model averaged coefficients, conditional on inclusion and their 95% credible intervals (95% CI), as in [20].

**Sample Size for SNP Analysis.** There were 763 current smokers with complete genotype and covariate data: 406 cases and 357 controls. There were 719 former smokers: 453 cases and 266 controls, and 743 never smokers: 316 cases and 427 controls.

**Sample Size for Nutritional Analysis.** There were 545 current smokers with full nutritional and covariate data available: 250 cases and 295 controls. There were 547 former smokers: 319 cases and 228 controls, and there were 685 never smokers: 279 cases and 406 controls.

**Determining the final model.** The following sample sizes reflect the numbers of subjects with available genotype and nutritional data. Of the current smokers, there were 577 subjects, with 263 cases and 314 controls, in whom we investigated 12 SNPs and 2 nutritional variables and 24 interaction terms. Of the former smokers, there were 572 subjects, with 337 cases and 235 controls, in whom we investigated 14 SNPs and 7 nutritional variables and 96 interactions. For never smokers, the sample size was 743, with 304 cases and 439 controls, and we investigated 26 SNPs and 6 nutritional variables and 150 interactions. Highly collinear interactions were dropped from the selection process (2 interac-



**Figure 1. Analysis Flow Chart.** This figure depicts the flow of analysis. We analyzed SNPs and nutrients in parallel, using stochastic search methodology in stage 2. Then the most important SNPs and nutrients were jointly investigated along with the gene-nutrient interactions in stage 3, again using stochastic search methodology. doi:10.1371/journal.pone.0053475.g001

tions for former smokers, and 6 interaction terms for never smokers).

## Results

We summarize selected demographic variables and nutritional variables in our study population in Table 1. There were equal proportions of male and female cases (50.5% versus 49.5%, respectively) but slightly fewer male (46.7%) than female controls (53.3%). The mean age of cases was significantly higher than the mean age of controls (63.5 years versus 57.2 years,  $p < 0.001$ ). Our sample had a higher proportion of former and current smokers in the cases (72.4%) than in the controls (58.8%). Cases who smoked had significantly more pack years than controls (mean 78.3 pack years versus mean 59.0 pack years,  $p < 0.001$ ). (Our analyses were stratified by smoking status and adjusted for pack years.) More controls (41.1%) were never smokers than cases (27.5%). Current and former smokers reported exposure to environmental tobacco smoke in close to equal proportions of cases and controls (68.2% versus 68.5%, respectively), and significantly more cases than

controls reported having at least one relative with a smoking-related cancer (30.3% versus 21.2%, respectively).

Only a few dietary factors differed significantly across cases and controls. Considering the median grams of alcohol drunk, cases reported significantly less drinking (median = 0.27 g) than controls (median = 0.82 g). Cases reported eating significantly less protein (median = 73.58 g) than controls (76.43 g). Cases also reported significantly less betaine (median = 48.94 mg), methionine (median = 1480.22 mg), niacin (median = 21.99 mg), vitamin B<sub>6</sub> (median = 5.43 mg), and vitamin B<sub>12</sub> (median = 5.26 mcg) than controls (medians = 53.20 mg, 1554.95 mg, 22.67 mg, 2.07 mg, 5.42 mcg, respectively).

## Initial Screen for Genotypes

Our initial screen for SNPs associated with lung cancer stratified by smoking status is reported in Table S2. Those SNPs with a PPI greater than 35% were further analyzed in the final model that jointly considered genes and nutrition. Among current smokers, we identified 12 SNPs for further consideration: 2 SNPs from the *CBS* gene, 7 SNPs from *MTRR*, 2 SNPs from *SHMT1*, and 1 SNP from *TYMS*. In former smokers, we identified 14 SNPs: 3 SNPs

**Table 1.** Distribution of Epidemiologic/Demographic Variables and Nutrition Variables.

Variable <sup>a</sup>	Category	Cases (N = 1692)	Controls (N = 1239)	P-value <sup>b</sup>
<b>Gender, n (%)</b>	Male	854 (50.5)	579 (46.7)	
	Female	838 (49.5)	660 (53.3)	0.049
<b>Age, mean (SD)</b>		63.5 (11.0)	57.2 (13.2)	<0.001
<b>Smoking status, n (%)</b>	Never	466 (27.5)	510 (41.1)	
	Former	645 (38.1)	326 (26.3)	
	Current	581 (34.3)	403 (32.5)	<0.001
<b>Pack years<sup>c</sup>, mean (SD)</b>		78.3 (73.7)	59.0 (49.8)	<0.001
<b>ETS<sup>d</sup>, n (%)</b>	Exposed	236 (68.2)	322 (68.5)	
	Unexposed	110 (31.8)	148 (31.5)	0.97
<b>Family history, n (%)</b>	None	936 (69.7)	938 (78.8)	
	≥1 relative	406 (30.3)	253 (21.2)	<0.001
<b>Nutrition, median</b>	Total calories (kcal)	1734.91	1717.92	0.752
	Alcohol (g)	0.27	0.82	0.002
<b>Macronutrients<sup>e</sup></b>	Carbohydrate (g)	233.95	231.80	0.256
	Protein (g)	73.58	76.43	<0.001
	Total Fat (g)	67.82	67.98	0.876
<b>Micronutrients<sup>e</sup></b>	Betaine (mg)	48.94	53.20	0.003
	Choline (mg)	141.15	143.04	0.316
	Folate (mg)	501.19	514.48	0.344
	Methionine (mg)	1489.22	1554.95	<0.001
	Niacin (mg)	21.99	22.67	0.005
	Pantothenic acid (mg)	5.43	5.33	0.116
	Riboflavin (mg)	2.23	2.23	0.661
	Thiamin (mg)	1.47	1.47	0.801
	Vitamin B <sub>6</sub> (mg)	2.00	2.07	0.012
	Vitamin B <sub>12</sub> (mg)	5.26	5.42	0.005

<sup>a</sup>Totals may not equal total N due to missing data.

<sup>b</sup>P-value from the two-sided chi-squared test (for categorical variables), Student's t-test (for continuous demographic variables), or Wilcoxon rank sum test (for nutrition variables).

<sup>c</sup>Pack years computed for former and current smokers only.

<sup>d</sup>ETS = Environmental Tobacco Smoke exposure in never smokers.

<sup>e</sup>Macro- and micronutrients are adjusted by total calorie intake.

doi:10.1371/journal.pone.0053475.t001

from the *CBS* gene, 2 from *MTHFR*, 1 from *MTR*, 7 from *MTRR*, and 1 from *TYMS*. In never smokers, we identified 26 SNPs: 4 in *CBS*, 8 in *MTHFR*, 11 in *MTRR*, 1 in *SHMT1*, and 2 in *TYMS*.

### Initial Screen for Nutrients

The results from our initial screen for potential nutrients associated with lung cancer are reported in Table S3, stratified by smoking status. Those nutrients with a PPI greater than 35% were further analyzed in our final model that jointly considered genes and nutrition. In current smokers, only alcohol and vitamin B<sub>6</sub> were identified for consideration in the final model. In former smokers, we identified alcohol, carbohydrates, protein, betaine, methionine, thiamin, and vitamin B<sub>12</sub> for further consideration. In never smokers, the nutrients for further consideration were carbohydrates, protein, choline, folate, riboflavin, and thiamin.

### Final Models

The collinearity of variables in our final model was controlled using the tag SNP selection process described above. We include the LD matrix describing the LD between the final selected SNPs in Table S4 (max  $r^2 < 0.63$ ) and the correlations between the final

selected nutrients in Table S5 (max  $r^2 < 0.62$ ). Simulation studies have shown stochastic search to perform well in the presence of moderate collinearity of magnitude on the order of 0.6 [20].

**Current Smokers.** In current smokers, *MTRR* (rs6893114) and alcohol were associated with lung cancer risk, after adjusting for sex, age, pack years, and family history (Table 2). *MTRR* had the highest PPI, and the minor allele of rs6893114 conferred a twofold increase in lung cancer risk (OR = 2.10; 95% CI: 1.20–3.48). As alcohol drinking appeared to be protective among current smokers (OR = 0.48; 95% CI: 0.26–0.84), we further examined alcohol intake using more detailed categorization [39]: nondrinker, 0.1–4.9 g/day, 5.0–14.9 g/day, 15–29.9 g/day, and greater than 30 g/day. We computed an adjusted odds ratio for each level (see Table 3), using non-drinkers as the reference category and adjusting for age, sex, family history, pack years, and the *MTRR* variant. The two lowest drinking categories (light and moderate drinking) showed a decrease in risk from drinking: 0.1–4.9 g/day is associated with a 39% decrease in risk, (OR = 0.61; 95% CI: 0.40–0.93) and 5–14.9 g/day is associated with a 42% decrease in risk (OR = 0.58; 95% CI: 0.34–0.99). We did not find

**Table 2.** Final Model Stratified by Smoking Status.

Model	Variable	Gene <sup>a</sup>	OR <sup>b</sup>	95% CI	PPI	Bayes factor
<b>Current Smokers</b>	alcohol		0.48	0.26–0.84	0.90	8.81 <sup>c</sup>
	rs6893114	<i>MTRR</i>	2.10	1.20–3.48	0.88	7.25 <sup>c</sup>
<b>Former Smokers</b>	betaine*rs2658161	<i>MTRR</i>	0.42	0.19–0.88	0.82	4.62 <sup>c</sup>
	rs13170530	<i>MTRR</i>	1.70	1.10–2.87	0.81	4.21 <sup>c</sup>
	betaine*rs16948305	<i>TYMS</i>	0.54	0.30–0.91	0.76	3.10 <sup>c</sup>
	rs2658161	<i>MTRR</i>	2.23	0.96–4.98	0.72	2.22
	betaine		0.71	0.34–1.35	0.32	0.47
	rs16948305	<i>TYMS</i>	0.89	0.54–1.62	0.23	0.30
<b>Never Smokers</b>	rs2924471	<i>MTRR</i>	3.31	1.66–6.59	0.99	78.31 <sup>c</sup>
	rs13162612	<i>MTRR</i>	0.25	0.11–0.58	0.98	65.63 <sup>c</sup>
	rs10512948	<i>MTRR</i>	0.61	0.41–0.90	0.83	5.17 <sup>c</sup>
	choline*rs10475407	<i>MTRR</i>	1.62	1.11–2.42	0.82	4.83 <sup>c</sup>
	riboflavin*rs876712	<i>MTRR</i>	0.40	0.15–0.95	0.80	3.98 <sup>c</sup>
	choline*rs11134290	<i>MTRR</i>	0.51	0.27–0.92	0.78	3.48 <sup>c</sup>
	rs9651118	<i>MTHFR</i>	0.63	0.43–0.95	0.75	3.05 <sup>c</sup>
	rs11134920	<i>MTRR</i>	0.68	0.44–1.05	0.48	0.92
	riboflavin		1.57	0.65–4.03	0.45	0.81
	rs10475407	<i>MTRR</i>	0.77	0.57–1.04	0.38	0.62
	choline		0.75	0.35–1.78	0.36	0.55
	rs876712	<i>MTRR</i>	0.73	0.41–1.29	0.35	0.55

<sup>a</sup>SNPs located within 500 kb of given gene.

<sup>b</sup>Odds ratios for current smokers adjusted for sex, age, family history of smoking-related cancers, and pack years smoked; odds ratios for former smokers adjusted for sex, age, family history, and age at smoking cessation; odds ratios for never smokers adjusted for sex, age, family history, and exposure to secondhand tobacco smoke.

<sup>c</sup>Bayes factor greater than or equal to 3; included in the final model.

doi:10.1371/journal.pone.0053475.t002

any evidence of gene-nutrient interactions for current smokers in our data.

**Former Smokers.** For former smokers, one SNP in *MTRR* (rs13170530) showed evidence of an association with lung cancer risk after adjusting for age, sex, age at smoking cessation, and family history. We also found a significant interaction between betaine and a variant in *MTRR* (rs2658161) and a variant in *TYMS* (rs16948305). The minor allele in *MTRR* (rs13170530) confers a 70% increase in risk (OR = 1.70; 95% CI: 1.10–2.87). Although no nutrient main effects were significant, betaine was a part of two meaningful interactions that were both protective among former smokers: betaine\*rs2658161 (OR = 0.42; 95% CI: 0.19–0.88) and betaine\*rs16948305 (OR = 0.54, 95% CI: 0.30–0.91). Thus each copy of the minor allele of rs2658161 improves

the protective effect of ingesting more betaine. Zero copies of the minor allele results in no reduction in risk, 1 allele results in a 58% reduction in risk per mg increase in betaine intake, while 2 copies of the minor allele result in 0.83% reduction in risk. Individuals heterozygous for the minor allele of rs16948305 receive a 46% decrease in lung cancer risk per mg betaine consumed, while those homozygous for the minor allele have a 70% decrease in risk.

**Never Smokers.** For never smokers, 4 SNPs, three in *MTRR* (rs13162612, rs10512948, rs2924471) and one in *MTHFR* (rs9651118), were associated with lung cancer risk, after adjusting for environmental tobacco exposure, sex, age, and family history. The minor allele for rs13162612 was associated with a 75% reduction in lung cancer risk (OR = 0.25; 95% CI: 0.11–0.58), and rs10512948 was associated with a 39% decrease in lung cancer risk (OR = 0.61; 95% CI: 0.41–0.90). The third SNP from *MTRR*, rs2924471, was associated with a 3-fold increased risk (OR = 3.31; 95% CI: 1.66–6.59). In addition to the genetic main effects, three nutrient by gene interactions were selected: choline\*rs10475407, choline\*rs11134290 and riboflavin\*rs876712. The choline\*rs10475407 interaction conferred risk (OR = 1.62; 95% CI: 1.11–2.41). Being heterozygous for the minor allele of rs10475407 is associated with a 62% increased risk, while being homozygous for the minor allele is associated with a 2.6 fold increase in risk for each mg increase of choline intake. On the other hand, individuals heterozygous for the minor allele of rs11134290 had a 49% decrease in risk (OR = 0.51; 95% CI: 0.27–0.92) for increased choline intake, while those homozygous for the minor allele had a 74% decreased risk per mg increased intake in choline. Individuals heterozygous for the minor allele of rs876712 have a 60% decreased risk per increased intake in riboflavin, which increases to

**Table 3.** Further Examination of Alcohol from the Final Model for Current Smokers (263 Cases/314 Controls).

Alcohol Intake Category	Cases	Controls	OR <sup>a</sup>	95% CI
Non-drinkers	116	86	1	Reference
0.1–4.9 g/day	72	111	0.61	0.40–0.93
5–14.9 g/day	31	58	0.58	0.34–0.99
15–29.9 g/day	21	30	0.73	0.38–1.37
>30 g/day	23	29	0.75	0.40–1.40

<sup>a</sup>Odds ratios adjusted for age, sex, family history, pack years, and *MTRR* mutants.  
doi:10.1371/journal.pone.0053475.t003



84% for those homozygous for the minor allele (riboflavin\*rs876712 OR = 0.40; 95% CI: 0.15–0.95).

## Discussion

This study identified various nutritional factors and genetic factors related to folate metabolism that jointly play a role in lung cancer risk. Performing analyses stratified by smoking status (current, former, and never), we found folate-related dietary and genetic factors, and gene\*nutrient interactions were associated with lung cancer risk in current, former, and never smokers. Alcohol was associated with lung cancer risk in current smokers, while gene-nutrient interactions were associated with varying risk in former and never smokers. SNPs in *MTRR* were associated with lung cancer risk in current, former and never smokers, while a variant in *MTHFR* was associated with lung cancer risk in never smokers. An additional SNP in *TYMS* was found to interact with betaine to influence lung cancer risk in former smokers.

Recent research has shown mixed results regarding the association between alcohol drinking and lung cancer risk in non-Hispanic whites [39,50,51,52]. Evidence is more consistent in never smokers for no association between alcohol and lung cancer risk [50], which aligns with the findings of the current study. For current smokers, there is less evidence regarding the association between alcohol intake and lung cancer risk [51]. Recent studies suggest that the influence of alcohol may depend on the type of alcohol consumed, citing a possible protective effect for wine and increased risk for beer [52]. Other studies show a marginal, non-linear relationship between alcohol intake and lung cancer risk, with moderate drinking having a protective effect [39]. The strong posterior probability for alcohol seen here suggests that alcohol may be associated with lung cancer risk in current smokers; however, this study does not provide any definitive resolution regarding the mediating effects of smoking on the relationship between alcohol and lung cancer risk. Possible explanations for this apparent association are that cases stopped drinking recently relative to their diagnosis or simply under-report their drinking because of recent diagnosis.

Our analysis also identified several polymorphisms associated with lung cancer risk. For all smoking statuses, different SNPs in *MTRR* exhibited strong evidence for association with lung cancer risk. In current smokers, we identified rs6893114 with increased risk, for former smokers, we identified rs13170530 with increased risk, and for never smokers we identified multiple *MTRR* SNPs: rs13162612 and rs10512948 were associated with decreased risk; and rs2924471 was associated with increased risk for lung cancer. There is very little information regarding the association of *MTRR* with lung cancer, with most studies focusing on *MTRR A66G* (rs1801394) [7,9,53]. Two previous studies found no association for *MTRR A66G* [7,53], while a third found increased risk [9]. All three studies mentioned an interaction between smoking status and *A66G* alleles. The fact that the current study also found polymorphisms in *MTRR* provides further evidence of an association between *MTRR* and lung cancer.

In never smokers, an additional polymorphism in *MTHFR*, rs9651118, is associated with decreased lung cancer risk. Over the past decade, many researchers have focused their efforts on two particular polymorphisms of *MTHFR*, *C677T* and *A1289C*, owing to variants from wild-type at these loci resulting in altered serum folate levels [7,54,55]. However, results concerning these two loci and their association with lung cancer risk are often inconclusive [54,56]. In the current study, *C677T* was not selected as being associated with lung cancer. The SNP associated with lung cancer in never smokers, rs9651118 (T/C), has a borderline Bayes factor

(3.05), and moderate protective effect for lung cancer (37% decrease in risk). This SNP is in low LD with the *C677T* and *A1289C* polymorphisms for *MTHFR* ( $r^2 < 0.20$ ) and is located in an intronic region of *MTHFR*. (We do not use *D'* here, because by *D'*, all tag SNPs have  $D' > 0.9$  with *C677T*.) Given the findings in this study, further investigation of this SNP is encouraged.

Located at 1p36.3, the *MTHFR* gene codes for the methylenetetrahydrofolate reductase that converts 5,10-methylenetetrahydrofolate to 5-methylenetetrahydrofolate, which is the primary circulating form of folate and provides methyl groups for synthesis of methionine, an important factor for healthy DNA methylation. *MTRR* codes for methionine synthase reductase, which controls methionine synthase, which uses methionine as a methyl donor for DNA methylation. A disruption in any of these three metabolites can lead to chromosome instability and DNA under-methylation, and ultimately to cancer [57,58]. *TYMS* codes for thymidylate synthase, an enzyme that is key to a reaction providing thymidine, an important nucleotide used in DNA synthesis and repair. Increased activity is expected to be associated with healthier DNA, while decreased activity is expected to be associated with more DNA damage and thus higher cancer risk [10,53].

Our analysis did not detect any nutrient main effects; however, for each smoking status we did detect statistical interactions. In former smokers, we detected a statistical interaction between variants in *MTRR* and *TYMS* and betaine, and in never smokers we detected interactions between variants in *MTRR* and choline and riboflavin. The Bayes factors indicated no evidence for any associations between lung cancer and the main effects of betaine, choline or riboflavin or the SNPs involved in the interactions, but as the intake of betaine and allele dose of rs2658161 (*MTRR*) and rs16948305 (*TYMS*) increased, our model indicated a reduction in lung cancer risk for former smokers. With never smokers, a statistical interaction with choline and rs10475407 (*MTRR*) lead to an increased risk, while the interactions of choline with rs11134290 (*MTRR*) and riboflavin with rs876712 (*MTRR*) were modeled to decrease lung cancer risk. Researchers are just beginning to investigate choline and betaine intake in human studies, due to the recently available database linkage to FFQs for betaine and choline [59]. Some human studies have linked breast cancer [60] and colon cancer [61] to choline and betaine intake levels, while other studies have found no association [62,63,64]. Riboflavin has been reported with mixed associations with lung cancer as well [6,65,66]. Therefore, the literature offers other studies that support many of the SNPs found by this Bayesian model averaging method. However, this is one of the first studies to jointly model these risk factors for lung cancer, and further validation of these findings is needed.

The findings of the current study need to be interpreted in the light of certain limitations especially for the nutrition data. First, because this study sample was restricted to non-Hispanic whites, our findings may not generalize to other ethnicities. Second, this study is a cross-sectional study and information on all variables was collected upon recruitment, and we cannot investigate any real change in behavior over time, such as a change in drinking behavior. Third, the controls were selected from an HMO in the greater Houston metropolitan area. Therefore, controls may not be fully representative of the general population. The fact that these individuals sought medical care might suggest a higher awareness of health and, perhaps, of the importance of proper nutrition. Therefore, the nutrition profiles may not accurately reflect intake in the general population. However, a previous study found the intake of various food items in this population to be comparable to those found by NHANES [32,33,34].

Additionally, the pattern of missing data is significantly different in smokers versus non-smokers, but since we stratify by smoking status, the bias will be minimal. The missing pattern between cases and controls are not significantly different. The nutrition data were collected using food frequency questionnaires, which have the well-known limitations of recall bias [67], minimized in this study by interviewer administration. Even though this bias was minimized by administration by trained interviewers, it may be a factor contributing to the difference in the findings here compared to results found using prospective data such as EPIC [6] and ATBC [5]. Once we removed the missing data, and stratified, the sample sizes are small. Yet using the Bayesian approach, we were able to control the false discovery rate to be less than 15%, which for the number of findings of the study, comes to one expected false positive per model. Even though the false discovery rate was controlled, and recall bias was minimized, an important next step is to externally validate these findings with independent, prospectively collected data sets.

We would also like to discuss our independence assumptions. When constructing our priors, we modeled genetic covariance using linkage disequilibrium, but assumed nutrition variables and gene-nutrition interactions to be independent. Prior definitions are not rigid assumptions, but rather reflection of the prior belief of the modeler [68]. Previous simulation studies involving LD as a prior showed that it can reduce false positives from multicollinearity in the presence of high LD [20,69,70]. This covariance argument can generalize to correlation between any covariates. As a secondary precaution we computed the false discovery rate as described in [49], and it was controlled at around 15%.

To our knowledge, this is one of the first studies to jointly assess the association between lung cancer and a comprehensive panel of candidate genes in the folate pathway and nutrients related to folate metabolism, and nutrient-gene interactions. Furthermore, we used a novel Bayesian model averaging method to explore these associations. Strengths of this study include a sample size large enough to stratify by smoking status and jointly investigate multiple factors. Jointly modeling gene and nutrient factors allowed us to comprehensively assess the impact of folate metabolism and lung cancer risk. Through our stratified models, we also show that the genetic and nutritional impact on lung

cancer risk differs by smoking status. These preliminary findings suggest that the impact of dietary interventions for lung cancer risk may be modified by genotypes in key folate metabolism genes. These findings mark a first step toward more personalized interventions to reduce cancer risk. In developing dietary interventions to reduce lung cancer risk, we not only need to consider smoking status, but also potentially, the genotypes of folate metabolism genes, and how they interact with the nutrient intake levels.

## Supporting Information

**Table S1** SNP location and Function. Lists all SNPs analyzed and their minor allele frequency, gene, and function or location by RS number.

(DOCX)

**Table S2** SNPs selected from initial screening stratified by smoking status. Lists all snps that passed the first screen of association using PPI greater than 0.35 for each smoking status.

(DOCX)

**Table S3** Nutrients Selected from Initial Screening Stratified by Smoking Status. Table listing nutrients that passed the first screen of association using PPI greater than 0.35, for each smoking status.

(DOCX)

**Table S4** Linkage Disequilibrium Magnitudes Between SNPs in the Final Model. Table listing the absolute LD magnitudes for those SNPs identified in any final model.

(DOCX)

**Table S5** Correlation Between Nutrients in Final Model. Table listing the correlation between nutrients identified in any final model.

(DOCX)

## Author Contributions

Conceived and designed the experiments: MRS XW MDS MRF. Performed the experiments: MRS MRF XW. Analyzed the data: MDS SS LMH CBP MV. Contributed reagents/materials/analysis tools: XW. Wrote the paper: MDS CBP PJJ LMH MRS MRF SS MV.

## References

- American Cancer Society (2010) Cancer Facts and Figures 2010. Atlanta, Georgia.
- World Cancer Research Fund/the American Institute for Cancer Research (2007) Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective. Washington DC: AICR.
- Alberg AJ, Brock MV, Samet JM (2005) Epidemiology of lung cancer: looking to the future. *Journal of Clinical Oncology* 23: 3175–3185.
- Sakoda LC, Loomis MM, Doherty JA, Neuhauser ML, Barnett MJ, et al. (2011) Chromosome 15q24–25.1 variants, diet, and lung cancer susceptibility in cigarette smokers. *Cancer Causes Control* 22: 449–461.
- Hartman TJ, Woodson K, Stolzenberg-Solomon R, Virtamo J, Selhub J, et al. (2001) Association of the B-vitamins pyridoxal 5'-phosphate (B(6)), B(12), and folate with lung cancer risk in older men. *Am J Epidemiol* 153: 688–694.
- Johansson M, Relton C, Ueland PM, Vollset SE, Midttun O, et al. (2010) Serum B vitamin levels and risk of lung cancer. *JAMA* 303: 2377–2385.
- Piskac-Collier AL, Monroy C, Lopez MS, Cortes A, Etzel CJ, et al. (2011) Variants in folate pathway genes as modulators of genetic instability and lung cancer risk. *Genes Chromosomes Cancer* 50: 1–12.
- Shen M, Rothman N, Berndt S, He X, Yeager M, et al. (2005) Polymorphisms in folate metabolic genes and lung cancer risk in Xuan Wei, China. *Lung Cancer*.
- Shi Q, Zhang Z, Li G, Pillow PC, Hernandez LM, et al. (2005) Polymorphisms of methionine synthase and methionine synthase reductase and risk of lung cancer: a case-control analysis. *Pharmacogenet Genomics* 15: 547–555.
- Shi Q, Zhang Z, Neumann A, Li G, Spitz MR, et al. (2005) Case-control analysis of thymidylate synthase polymorphisms and risk of lung cancer. *Carcinogenesis* 26: 649–656.
- Wang L, Lu J, An J, Shi Q, Spitz MR, et al. (2007) Polymorphisms of cytosolic serine hydroxymethyltransferase and risk of lung cancer: a case-control analysis. *Lung Cancer* 57: 143–151.
- Wei Q, Shen H, Wang LE, Duphorne CM, Pillow PC, et al. (2003) Association between low dietary folate intake and suboptimal cellular DNA repair capacity. *Cancer Epidemiol Biomarkers Prev* 12: 963–969.
- Shen H, Wei Q, Pillow PC, Amos CI, Hong WK, et al. (2003) Dietary Folate Intake and Lung Cancer Risk in Former Smokers: A Case-control analysis. *Cancer Epidemiology, Biomarkers & Prevention* 12: 980–986.
- Voorrips LE, Goldbohm RA, Brants HA, van Poppel GA, Sturmans F, et al. (2000) A prospective cohort study on antioxidant and folate intake and male lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 9: 357–365.
- Liu H, Jin G, Wang H, Wu W, Liu Y, et al. (2008) Association of polymorphisms in one-carbon metabolizing genes and lung cancer risk: a case-control study in Chinese population. *Lung Cancer* 61: 21–29.
- George EI, McCulloch RE (1993) Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association* 88: 881–889.
- Swartz MD, Yu RK, Shete S (2008) Finding factors influencing risk: Comparing Bayesian stochastic search and standard variable selection methods applied to logistic regression models of cases and controls. *Stat Med* 27: 6158–6174.
- Fridley BL (2009) Bayesian variable and model selection methods for genetic association studies. *Genet Epidemiol* 33: 27–37.
- Stephens M, Balding DJ (2009) Bayesian statistical methods for genetic association studies. *Nat Rev Genet* 10: 681–690.
- Swartz MD, Kimmel M, Mueller P, Amos CI (2006) Stochastic Search Gene Suggestion: A Bayesian Hierarchical Model for Gene Mapping. *Biometrics* 62: 495–503.



21. Swartz MD, Thomas DC, Daw EW, Albers K, Charlesworth JC, et al. (2007) Model selection and Bayesian methods in statistical genetics: summary of group 11 contributions to Genetic Analysis Workshop 15. *Genet Epidemiol* 31 Suppl 1: S96–102.
22. Srivastava S, Chen L (2009) Comparison between the stochastic search variable selection and the least absolute shrinkage and selection operator for genome-wide association studies of rheumatoid arthritis. *BMC Proc* 3 Suppl 7: S21.
23. Guan Y, Stephens M (2011) Bayesian variable selection regression for genome-wide association studies, and other large-scale problems. *Annals of Applied Statistics* 5: 1780–1815.
24. Kwon S, Wang D, Guo X (2007) Application of an Iterative Bayesian Variable Selection Method in a Genome-Wide Association Study of Rheumatoid Arthritis. *BMC Proceedings* 1: SX.
25. Schumacher FR, Kraft P (2007) A Bayesian Latent Class Analysis for Whole-Genome Association Analyses. *BMC Proceedings* 1: S112–S116.
26. Hudmon KS, Honn SE, Jiang H, Chamberlain RM, Xiang W, et al. (1997) Identifying and recruiting healthy control subjects from a managed care organization: a methodology for molecular epidemiological case-control studies of cancer. *Cancer Epidemiology, Biomarkers & Prevention* 6: 565–571.
27. Wu X, Zhao H, Amos CI, Shete S, Makin N, et al. (2002) p53 Genotypes and Haplotypes Associated With Lung Cancer Susceptibility and Ethnicity. *J Natl Cancer Inst* 94: 681–690.
28. Block G, Coyle LM, Hartman AM, Scoppa SM (1994) Revision of dietary analysis software for the Health Habits and History Questionnaire. *American Journal of Epidemiology* 139: 1190–1196.
29. Block G, Hartman AM (1994) DIETSYS Version 3.0 User's Guide, pp.10–16.
30. Block G, Thompson FE, Hartman AM, Larkin FA, Guire KE (1992) Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. *Journal of the American Dietetic Association* 92: 686–693.
31. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, et al. (1986) A data-based approach to diet questionnaire design and testing. *Am J Epidemiol* 124: 453–469.
32. Mahabir S, Forman MR, Barerra SL, Dong YQ, Spitz MR, et al. (2007) Joint effects of dietary trace metals and DNA repair capacity in lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 16: 2756–2762.
33. Mahabir S, Spitz MR, Barrera SL, Beaver SH, Etzel C, et al. (2006) Dietary zinc, copper and selenium, and risk of lung cancer. *International Journal of Cancer* 120: 1108–1115.
34. Mahabir S, Wei Q, Barrera SL, Dong YQ, Etzel CJ, et al. (2008) Dietary magnesium and DNA repair capacity as risk factors for lung cancer. *Carcinogenesis* 29: 949–956.
35. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, et al. (2007) A risk model for prediction of lung cancer. *J Natl Cancer Inst* 99: 715–726.
36. Forman MR, Lanza E, Yong LC, Holden JM, Graubard BI, et al. (1993) The correlation between two dietary assessments of carotenoid intake and plasma carotenoid concentrations: application of a carotenoid food-composition database. *Am J Clin Nutr* 58: 519–524.
37. U. S. Department of Agriculture ARS (2008) USDA National Nutrient Database for Standard Reference, release 21. Nutrient Data Laboratory Home Page.
38. Willett W, Stampfer MJ (1986) Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol* 124: 17–27.
39. Rohrmann S, Linseisen J, Boshuizen HC, Whittaker J, Agudo A, et al. (2006) Ethanol intake and risk of lung cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Am J Epidemiol* 164: 1103–1114.
40. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, et al. (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 74: 106–120.
41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575.
42. NIEHS SNPS (2010) NIEHS Environmental Genome Project. In: University of Seattle W, editor. Seattle, WA.
43. Spiegelhalter DJ, Thomas A, Best N, Lunn D (2007) WinBUGS. 1.4.2 ed. Cambridge.
44. Sturtz S, Ligges U, Gelman A (2005) R2WinBUGS: A package for Running WinBUGS from R. *Journal of Statistical Software* 12: 1–16.
45. Gelman A (2004) Bayesian data analysis. Boca Raton, Fla.: Chapman & Hall/CRC. xxv, 668 p.
46. Barbieri MM, Berger JO (2004) Optimal predictive model selection. *Annals of Statistics* 32: 870–897.
47. Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical Association* 90: 773–795.
48. Wilson MA, Iversen ES, Clyde MA, Schmidler SC, Schildkraut JM (2010) Bayesian Model Search and Multilevel Inference for Snp Association Studies. *Ann Appl Stat* 4: 1342–1364.
49. Newton MA, Nouciry A, Sarkar D, Ahlquist P (2004) Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* 5: 155–176.
50. Bagnardi V, Rota M, Botteri E, Scotti L, Jenab M, et al. (2011) Alcohol consumption and lung cancer risk in never smokers: a meta-analysis. *Ann Oncol*.
51. Bandera EV, Freudenheim JL, Vena JE (2001) Alcohol consumption and lung cancer: a review of the epidemiologic evidence. *Cancer Epidemiol Biomarkers Prev* 10: 813–821.
52. Benedetti A, Parent ME, Siemiatycki J (2006) Consumption of alcoholic beverages and risk of lung cancer: results from two case-control studies in Montreal, Canada. *Cancer Causes Control* 17: 469–480.
53. Suzuki T, Matsuo K, Hiraki A, Saito T, Sato S, et al. (2007) Impact of one-carbon metabolism-related gene polymorphisms on risk of lung cancer in Japan: a case control study. *Carcinogenesis* 28: 1718–1725.
54. Boccia S, Boffetta P, Brennan P, Ricciardi G, Gianfagna F, et al. (2009) Meta-analyses of the methylenetetrahydrofolate reductase C677T and A1298C polymorphisms and risk of head and neck and lung cancer. *Cancer Lett* 273: 55–61.
55. Parle-McDermott A, Mills JL, Molloy AM, Carroll N, Kirke PN, et al. (2006) The MTHFR 1298CC and 677TT genotypes have opposite associations with red cell folate levels. *Mol Genet Metab* 88: 290–294.
56. Mao R, Fan Y, Jin Y, Bai J, Fu S (2008) Methylenetetrahydrofolate reductase gene polymorphisms and lung cancer: a meta-analysis. *J Hum Genet* 53: 340–348.
57. Sharp L, Little J (2004) Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review. *Am J Epidemiol* 159: 423–443.
58. Zijno A, Andreoli C, Leopardi P, Marcon F, Rossi S, et al. (2003) Folate status, metabolic genotype, and biomarkers of genotoxicity in healthy subjects. *Carcinogenesis* 24: 1097–1103.
59. Zeisel SH, Mar MH, Howe JC, Holden JM (2003) Concentrations of choline-containing compounds and betaine in common foods. *J Nutr* 133: 1302–1307.
60. Xu X, Gammon MD, Zeisel SH, Lee YL, Wetmur JG, et al. (2008) Choline metabolism and risk of breast cancer in a population-based study. *FASEB J* 22: 2045–2052.
61. Cho E, Willett WC, Colditz GA, Fuchs CS, Wu K, et al. (2007) Dietary choline and betaine intake and the risk of distal colorectal adenoma in women. *J Natl Cancer Inst* 99: 1224–1231.
62. Cho E, Holmes M, Hankinson SE, Willett WC (2007) Nutrients involved in one-carbon metabolism and risk of breast cancer among premenopausal women. *Cancer Epidemiol Biomarkers Prev* 16: 2787–2790.
63. Cho E, Holmes MD, Hankinson SE, Willett WC (2010) Choline and betaine intake and risk of breast cancer among post-menopausal women. *Br J Cancer* 102: 489–494.
64. Lee JE, Giovannucci E, Fuchs CS, Willett WC, Zeisel SH, et al. (2010) Choline and betaine intake and the risk of colorectal cancer in men. *Cancer Epidemiol Biomarkers Prev* 19: 884–887.
65. Bassett JK, Hodge AM, English DR, Baglietto L, Hopper JL, et al. (2012) Dietary intake of B vitamins and methionine and risk of lung cancer. *Eur J Clin Nutr* 66: 182–187.
66. Kabat GC, Miller AB, Jain M, Rohan TE (2008) Dietary intake of selected B vitamins in relation to risk of major cancers in women. *Br J Cancer* 99: 816–821.
67. Willett W (1998) Nutritional epidemiology. New York: Oxford University Press. xiv, 514 p.
68. Kruschke JK (2011) Doing Bayesian data analysis: a tutorial with R and BUGS. Burlington, MA: Academic Press. xvii, 653 p.
69. Swartz MD (2004) Stochastic Search Gene Suggestion: Hierarchical Bayesian Model Selection Meets Gene Mapping [Dissertation]. Houston, TX.: Rice University. 182 p.
70. Swartz MD, Shete S (2007) The Null Distribution of Stochastic Search Gene Suggestion: A Bayesian Approach to Gene Mapping. *BMC Proceedings* 1: S113–S118.