VOLUME 28 · NUMBER 14 · MAY 10 2010

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Interactions Among Genetic Variants in Apoptosis Pathway Genes, Reflux Symptoms, Body Mass Index, and Smoking Indicate Two Distinct Etiologic Patterns of Esophageal Adenocarcinoma

Rihong Zhai, Feng Chen, Geoffrey Liu, Li Su, Matthew H. Kulke, Kofi Asomaning, Xihong Lin, Rebecca S. Heist, Norman S. Nishioka, Chau-Chyun Sheu, John C. Wain, and David C. Christiani

From the Departments of Environmental Health and Biostatistics, Harvard School of Public Health; Department of Medical Oncology, Dana Farber Cancer Institute, Harvard Medical School; Division of Hematology-Oncology, Massachusetts General Hospital Cancer Center, Thoracic Surgery Division, Department of Surgery, Pulmonary and Critical Care Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA; and Departments of Medicine, Medical Biophysics, and Epidemiology, Princess Margaret Hospital/Ontario Cancer Institute, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada.

Submitted September 21, 2009; accepted January 14, 2010; published online ahead of print at www.jco.org on April 12, 2010.

Supported by National Institutes of Health Grants No. CA109193, CA110822, CA074386, and CA092824, and by Flight Attendant Medical Research Institute Grant No. 062459-YCSA.

Presented in part at the 100th Annual Meeting of the American Association for Cancer Research, April 18-22, 2009, Denver, CO.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Rihong Zhai, MD, PhD; Department of Environmental Health, Harvard School of Public Health, 665 Huntington Ave, Boston, MA 02115; e-mail: rzhai@hsph.harvard.edu.

0732-183X/10/2814-2445/$20.00

DOI: 10.1200/JCO.2009.26.2790

ABSTRACT

Purpose
Apoptosis pathway, gastroesophageal reflux symptoms (reflux), higher body mass index (BMI), and tobacco smoking have been individually associated with esophageal adenocarcinoma (EA) development. However, how multiple factors jointly affect EA risk remains unclear.

Patients and Methods
In total, 305 patients with EA and 339 age- and sex-matched controls were studied. High-order interactions among reflux, BMI, smoking, and functional polymorphisms in five apoptotic genes (*FAS, FASL, IL1B, TP53BP,* and *BAT3*) were investigated by entropy-based multifactor dimensionality reduction (MDR), classification and regression tree (CART), and traditional logistic regression (LR) models.

Results
In LR analysis, reflux, BMI, and smoking were significantly associated with EA risk, with reflux as the strongest individual factor. No individual single nucleotide polymorphism was associated with EA susceptibility. However, there was a two-way interaction between *IL1B + 3954C>T* and reflux ($P = .008$). In both CART and MDR analyses, reflux was also the strongest individual factor for EA risk. In individuals with reflux symptoms, CART analysis indicated that strongest interaction was among variant genotypes of *IL1B + 3954C>T* and *BAT3S625P,* higher BMI, and smoking (odds ratio [OR], 5.76; 95% CI, 2.48 to13.38), a finding independently found using MDR analysis. In contrast, for participants without reflux symptoms, the strongest interaction was found between higher BMI and smoking (OR, 3.27; 95% CI, 1.88 to 5.68), also echoed by entropy-based MDR analysis.

Conclusion
Although a history of reflux is an important risk for EA, multifactor interactions also play important roles in EA risk. Gene-environment interaction patterns differ between patients with and without reflux symptoms.

J Clin Oncol 28:2445-2451. © 2010 by American Society of Clinical Oncology

INTRODUCTION

Esophageal adenocarcinoma (EA) is an aggressive cancer with 5-year overall survival rates < 15% despite significant advances in treatment strategies.[1] The poor prognosis of EA underlines the importance of prevention for this lethal disease. However, the incidence of EA has increased approximately 500% in Western countries during the past four decades.[2,3] Although a growing body of epidemiologic evidence indicates that gastroesophageal reflux symptoms (reflux), obesity, smoking, male sex, and certain genetic variations are individually associated with EA risk,[4,5] no single factor can explain the rising prevalence of EA. Previous reports suggest that the development of EA is likely the result of complex interactions and cross-talk between environmental factors and genetic variants.[6-9] These studies, however, have been limited by analyzing gene-environment interactions based on the calculation of the two risk factors' product term in a logistic regression model. It remains to be established whether multiple risk factor interactions play a role in the risk of EA.[10] A better understanding of EA etiology could aid in identification of effective preventive strategies against the disease.

Apoptosis is a process of programmed cell death that is essential for cell growth and the maintenance of homeostasis.[11] Impaired regulation of apoptosis might lead to uncontrolled cell growth

and to tumor formation.[12] Genes that participate in the initiation or suppression of apoptosis or induce apoptosis secondary to cell injury or inflammation include *FAS, FASL, IL1B, TP53*, and *BAT3*.[12] Functional single nucleotide polymorphisms (SNPs) in these apoptotic genes, such as *FAS* −1377G>A (rs2234767), *FASL* −844 T>C (rs763110), *IL1B* +*3954*C>T (rs1143634), *TP53BP1D353E* (rs560191), and *BAT3625P* (rs1052486) have been shown to modulate gene expression, protein expression and production, and apoptosis.[13-17] These genetic polymorphisms are also individually associated with susceptibility to cancer development at individual and multiple organ sites.[17,18-20] However, it is unclear how these polymorphisms contribute to EA development.

In this study, we hypothesized that functional apoptotic SNPs are associated with the development of EA. Since other known EA risk factors, such as reflux, higher body mass index (BMI), and smoking are also known to be involved in the apoptosis process,[21-23] our secondary hypothesis was that interactions among apoptotic SNPs, reflux, BMI, and smoking will confer an even greater risk of EA. We used several statistical approaches, including classification and regression tree (CART) and entropy-based multifactor dimensionality reduction (MDR) in addition to traditional multiple logistic regression (LR) to explore high-order gene-environment interactions in EA susceptibility.

PATIENTS AND METHODS

Study Population

Incident patients with newly diagnosed and histologically confirmed EA were recruited prospectively at Massachusetts General Hospital between 1999 and 2005 and at Dana-Farber Cancer Institute between 2004 and 2005.[6] All patients were age 18 years or older and were diagnosed within 6 months before study entry. Patients with gastroesophageal junction tumors, but not gastric cardia, were included as patients. All pertinent clinical data were reviewed by a team of consultants consisting of a gastroenterologist, medical oncologist, and thoracic surgeon to ensure the diagnosis. Controls were accrued from healthy friends and non–blood-related family members in the same hospitals in the same period, originally recruited for our parallel lung cancer study from 1993 to 1999 and for multiple studies since 2000. All controls had never received a diagnosis of cancer. Patients and controls were frequency matched for age and sex distribution. All participants provided informed written consent. The study was approved by the Human Subjects Committees of Massachusetts General Hospital, Dana-Farber Cancer Institute, and the Harvard School of Public Health.

Interview

Immediately after enrollment, a specially trained interviewer administered a questionnaire that collected clinical and demographic information. Patients were interviewed during their hospital/clinic visit. Because controls were recruited when they accompanied other patients to their hospital/clinic visits, interviews for the controls took place in the same area as those for the patients and by the same interviewers. The questions covered demographic variables (current weight, weight 1 year before diagnosis/interview, weight at early adult age of 18 years, adult height, age, sex, race/ethnicity, and so on), a detailed smoking exposure assessment, past medical history (including exposure to radiation for benign conditions), family history, and occupational and environmental exposure history. Lifetime reflux symptoms (including reflux frequency and intensity) were assessed up to 1 year before diagnosis (patients) or 1 year before interview (controls). Chronic reflux was defined as having heartburn or regurgitation symptoms at least once monthly for at least a 6-month continuous period in one's lifetime. We chose this definition to capture a broader set of symptoms and to ensure that individuals defined as not having reflux were a pure group.[7] Participants were considered reflux-free if they had less than one episode of reflux per month. Smoking status was defined on the basis of whether the patients and controls smoked 1 year before diagnosis/interview.

Genotyping

On the basis of previous functional and epidemiologic studies and common frequency in the population, we selected five functional SNPs in five apoptotic genes, including *FAS* −1377 G>A (rs2234767), *FASL* −844T>C (rs763110), *IL1B* +*3954* C>T (rs1143634), *TP53BP1D353E* (rs560191), and *BAT3625P* (rs1052486). Genomic DNA was isolated from peripheral blood with a Gentra Systems/Qiagen Kit (Qiagen, Valencia, CA). Genotyping was performed using the TaqMan method with a 7900HT sequence detection system (Applied Biosystems, Foster city, CA). The primer and probe sequences for each SNP are available on request. Patients and controls were genotyped blinded to all clinical information and patient-control status in mixed batches. Two coauthors checked all results independently and a third arbitrated discrepancies. Controls were included in each plate to ensure accuracy of the genotyping, and a random 15% of the samples were run in duplicates with 100% concordance.

Statistical Analysis

Demographic and genotype information for patients and controls were compared using χ^2 test, Fisher's exact test, *t* test, and nonparametric Wilcoxon rank sum test, where appropriate. Hardy-Weinberg equilibrium for genotypes was tested by a goodness-of-fit χ^2 test (Appendix Table A1, online only). Because previous studies have shown that biologic functions differed between genotypes with major alleles and minor alleles in the SNPs studied,[13-17] genotypes were coded as wild type (major-allele homozygote) and variant genotype (minor-allele homozygote + heterozygote). Unconditional multivariate LR was used to estimate odds ratios (ORs) and 95% CIs, adjusting for age, sex, smoking status, reflux, and BMI. Interaction was tested using a multiplicative interaction term included in the multivariate model. The independent effects of individual risk factors on EA susceptibility were analyzed in a stepwise manner. False-discovery rate (FDR) was assessed to account for multiple comparison.[24] Statistical significances of the interactions were assessed using likelihood ratio tests comparing the models with and without interaction terms.

CART Analysis

CART analysis was performed by rpart package in R (version 2.8.1, 2008) to build a decision tree via recursive partitioning. Outcome in this analysis was coded as binary values (patient or control). CART creates a decision tree that depicts how well each genotype or environmental factor variable predicts class (eg, patient-control status).[25] A splitting rule based on information index was used to stratify data into subsets of individuals, which are represented in the CART decision tree as nodes. This process continues until the classification reaches the lowest cross-validation error in the terminal node. The higher-order gene-environment interactions identified by the CART were further analyzed using the likelihood ratio test, comparing the models with or without interaction terms in the multiple LR and adjusting for covariates.

MDR Analysis

The MDR software (version 2.0 alpha) and MDR-permutation testing (MDRPT) software (version 0.4.7) were used to analyze gene-environment interactions. The MDR is a nonparametric (ie, no hypothesis about the statistical parameter is made), model-free (ie, it assumes no particular inheritance model) method that selects important combinations of variables on the basis of entropy measures for evaluating the information gain (IG) associated with attribute interactions. MDR collapses high-dimensional data into a single dimensional variable with two levels (high and low risk) using the ratio of the number of patients to the number of controls, thus permitting interactions to be detected in relatively small sample sizes. It has been shown that MDR can identify putative high-order interactions in the absence of any significant independent main effects in cancers.[26,27]

Table 1. Demographic Characteristics of EA Patients and Controls

Characteristic	Patients (n = 305) No.	Patients (n = 305) %	Controls (n = 339) No.	Controls (n = 339) %	P
Age, years					.9545
Mean	62.6		62.6		
SD	11.0		10.6		
Male sex	271	88.9	294	86.7	.4114
Smoking status, never	66	21.6	94	27.7	.0742
Smoking, pack-years					.0115
Median	25.0		18.0		
Range	0 to 218		0 to 212		
BMI at age 18 years, kg/m^2					.0057
Mean	23.5		22.7		
SD	3.5		3.3		
BMI ≥ 25 at age 18 years	90	29.5	62	18.3	< .001
Reflux, yes	149	48.9	94	27.7	< .001
Stage					
I-IIA	88	28.9			
IIB-IVB	217	71.1			

NOTE. Reflux is defined as gastroesophageal reflux symptom.
Abbreviations: EA, esophageal adenocarcinoma; SD, standard deviation; BMI, body mass index.

RESULTS

Demographics

The participation rates for eligible patients and controls were both > 85%. We restricted the analysis to white patients who comprised the majority (> 96%) of the study population. A total of 305 patients and 339 age- and sex-matched healthy controls were included in the analysis. Age and sex distribution were similar for patients and controls. BMI at age 18 years in patients was significantly higher than that in controls, and there were greater proportions of ever-smokers and reflux symptoms in patients than in controls (Table 1).

Associations of Individual Factor or Pair-Wise Factors With EA Risk (LR analysis)

The genotyping success rate for all selected SNPs ranged from 99.0% to 99.5%. All SNPs in this study population were consistent with Hardy-Weinberg equilibrium ($P > .05$) in the controls. In main effect analysis, reflux, BMI, and smoking were significantly associated with increased risk of EA, and reflux had the highest individual association (OR, 2.52; 95% CI, 1.81 to 3.52; $P < .001$). No individual SNP was associated with EA susceptibility (Table 2). In interaction analysis, however, there was a two-way interaction between *IL1B + 3954C>T* and reflux (OR, 0.06; 95% CI, 0.01 to 0.72; $P = .008$), even after adjusting for multiple comparisons (FDR $P = .042$; Table 3).

Association of Multiple Factor Interactions With EA Risk (CART analysis)

In the CART analysis, the initial split of the root node was reflux status, indicating that reflux was the strongest risk factor for EA among the factors considered. Further inspection of the classification tree structure suggested distinct interaction patterns for subjects with and without reflux symptoms (Fig 1). In the absence of reflux symptoms, BMI was the strongest risk factor, and the combination of ever-smoking and BMI ≥ 25 exhibited the highest risk of EA with a 62% patient rate (OR, 3.27; 95% CI, 1.88 to 5.68; $P < .001$). In individuals with reflux symptoms, however, the *IL1B + 3954* C>T polymorphism was the most significant risk factor, and carriers of *IL1B + 3954* C>T and *BAT3S625P* variant genotypes with BMI ≥ 25 in smokers had the highest risk of EA (patient rate = 74%; OR, 5.76; 95% CI, 2.48 to 13.38; $P < .001$; terminal node 8; Table 4).

Association of Multiple Factor Interactions With EA Risk (MDR analysis)

Table 5 summarizes entropy-based measures of IG for all candidate risk factors in the etiology of EA. Consistent with the results from

Table 2. Main Effects of Individual Risk Factors on Esophageal Adenocarcinoma Risk by Logistic Regression

Variable	Unadjusted OR	Unadjusted 95% CI	Unadjusted P	Adjusted OR_{adj}	Adjusted 95% CI	Adjusted P
Environmental factors						
Reflux, yes	2.49	1.80 to 3.45	< .001	2.52	1.81 to 3.52*	< .001
BMI ≥ 25 at age 18 years, yes	1.87	1.29 to 2.71	< .001	2.00	1.36 to 2.93†	< .001
BMI at age 18, kg/m^2	1.06	1.07 to 1.11	.0064	1.07	1.02 to 1.12†	.0058
Smoking status, ever	1.39	0.97 to 1.99	.0735	1.38	0.95 to 2.01‡	.0940
Smoking pack-years	1.006	1.001 to 1.011	.0120	1.007	1.002 to 1.012‡	.0064
SNPs§						
BAT3625P	1.04	0.72 to 1.50	.8375	1.07	0.73 to 1.56	.7425
FAS-1377G>A	1.08	0.33 to 3.58	.8983	1.34	0.38 to 4.68	.6468
FASL-844T>C	0.99	0.65 to 1.51	.9624	1.05	0.68 to 1.64	.8120
IL1B +3954C>T	1.15	0.51 to 2.57	.7310	1.14	0.49 to 2.61	.7626
TP53BP1D353E	0.83	0.50 to 1.38	.4732	0.86	0.51 to 1.46	.5751

Abbreviations: OR, odds ratio; BMI, body mass index; SNP, single nucleotide polymorphism.
*Adjusted for age, sex, smoking status, and BMI category.
†Adjusted for age, sex, smoking status, and reflux status.
‡Adjusted for age, sex, and reflux status.
§Effect of SNP was adjusted for age, sex, smoking status, reflux status, and BMI category.

Table 3. Two-Way Gene-Environmental Interactions Detected by Logistic Regression

SNP*	Interaction With Reflux Status				Interaction With Smoking Status				Interaction With BMI Category			
	OR	95% CI	P†	P for FDR	OR_{adj}	95% CI	P‡	P for FDR	OR_{adj}	95% CI	P§	P for FDR
BAT3625P	0.63	0.28 to 1.40	.2511	.6278	1.82	0.75 to 4.38	.1840	.3889	0.81	0.34 to 1.95	.6396	.6396
FAS-1377 G>A	0.33	0.02 to 5.56	.4247	.7078	1.50	0.07 to 34.32	.7991	.7991	4.92	0.27 to 87.95	.2682	.6179
FASL-844 T>C	0.98	0.40 to 2.39	.9627	.9627	2.48	0.89 to 6.89	.0816	.3889	1.82	0.66 to 4.99	.2461	.6179
*IL1B +3954*C>T	0.06	0.01 to 0.72	.0084	.0421	0.39	0.03 to 4.39	.4169	.5211	2.10	0.26 to 16.82	.4798	.6179
TP53BP1D353E	1.34	0.47 to 3.82	.5790	.7238	0.43	0.11 to 1.76	.2333	.3889	1.71	0.37 to 7.82	.4943	.6179

Abbreviations: SNP, single nucleotide polymorphism; BMI, body mass index; OR, odds ratio; FDR, false-discovery rate; OR_{adj}, adjusted odds ratio.
*SNPs were classified as wild type and variant genotype.
†Adjusted for age, sex, smoking status, and BMI category.
‡Adjusted for age, sex, reflux status, and BMI category.
§Adjusted for age, sex, smoking status, and reflux status.

LR and CART analyses, history of reflux symptom was the strongest one-factor risk in the MDR model, with the highest value of entropy (IG = 3.44) among all variables studied, followed by BMI (IG = 1.26), smoking (IG = 0.36), *TP53BP1D353E* (IG = 0.06), and *IL1B +3954* C>T (IG = 0.01). In individuals with reflux symptoms, the highest IG values were observed for *IL1B +3954*C>T and *BAT3625P*. In patients without reflux symptoms, BMI and smoking had the highest IG values among other factors.

Appendix Figure A1 illustrates the overall and stratified interaction maps of variables based on entropy measures among individual variables. The patterns of entropy recapitulate the main and/or interaction effects for each model. In overall analysis, stronger interaction effect was found among reflux, BMI, and smoking while reflux had a clear main effect. In participants with reflux symptoms, *IL1B + 3954* C>T and *BAT3625P* had the highest IG values among other factors, and strongest interaction was detected among *IL1B +3954*C>T, *BAT3625P, TP53BP1D353E,* and *FASL* −844T>C and BMI. In individuals without reflux symptoms, BMI had the highest IG value, and the strongest interaction was seen between higher BMI and smoking.

DISCUSSION

In this study, we used multianalytic strategies to systematically examine the interactions among apoptotic gene polymorphisms, reflux, smoking, and BMI and their associations with EA development. Although each strategy used different algorithms to define these interactions, results from LR, CART, and MDR consistently showed that (1) reflux was the most significant single risk for EA development, and (2) EA risk was substantially associated with multifactor interactions that differed in individuals with and without reflux symptoms.

Previous studies have found associations of *FAS* −1377G>A, *FASL* −844T>C, *IL1B* + *3954*C>T, *TP53BP1D353E*, and *BAT3625P* with risks of certain cancers, including esophageal squamous cell carcinoma and lung, cervical, bladder, breast, brain, and pancreatic cancers.[13-17] But there have been no reports published on the associations of these polymorphisms with EA risk in white populations. Unlike previous studies in which apoptosis gene polymorphisms showed main effects on the risk of cancers, we did not find any overall associations between SNPs studied and EA development. However, we observed that apoptosis SNPs were associated with EA risk only in analyses stratified by reflux status, suggesting that apoptosis gene variations may exert their effects on EA development in combination with certain environmental factors, such as reflux or higher BMI. Our results emphasize the importance of analyzing gene-environment interactions for revealing the molecular mechanisms of EA development.

Our results indicated that gene-environment interaction patterns in EA development may depend on reflux status. When reflux symptoms were present, EA risk was modulated more by genetic

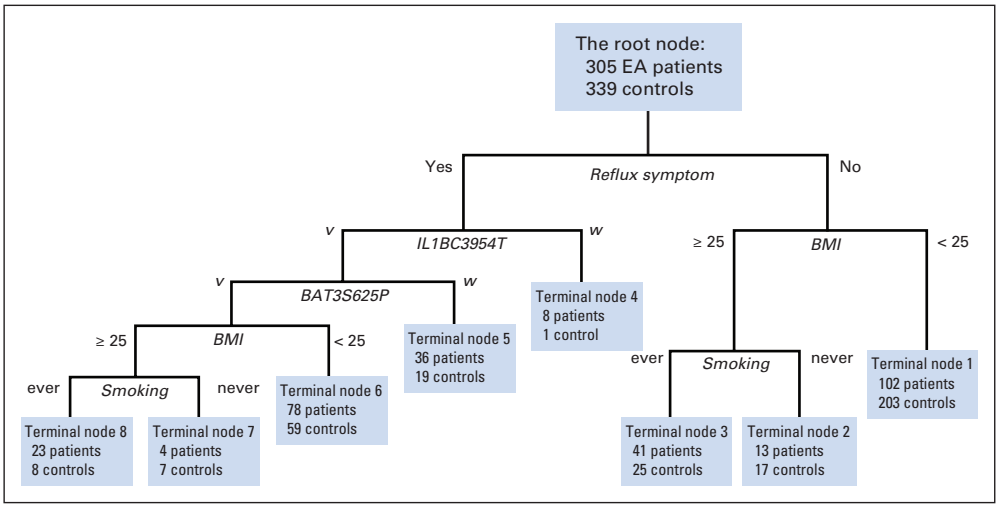

Fig 1. Classification and regression tree analysis of reflux, body mass index (BMI), smoking, and genetic polymorphisms in apoptosis pathway. Terminal nodes show number of esophageal adenocarcinoma (EA) patients/number of controls. Single nucleotide polymorphisms were classified as wild type (w) and variant genotype (v); smoking was defined as ever and never smoking; BMI was categorized as ≥ 25 and < 25; reflux was defined as yes and no.

Table 4. Risk Estimates of Classification and Regression Tree Terminal Nodes

Terminal Node	Combinations of Risk Factors				Patient:Control Ratio	OR	95% CI	*P**
	Reflux	Smoking*	BMI*	SNP*				
1	No		< 25		102:203	1.00 (reference)		—
2	No	Never	≥ 25		13:17	1.55	0.72 to 3.34	.2586
3	No	Ever	≥ 25		41:25	3.27	1.88 to 5.68	< .001
4	Yes			*IL1B + 3954C>T* (w)	8:1	15.64	1.93 to 26.95	.0101
5	Yes			*IL1B + 3954C>T* (v) *BAT3S625P* (w)	36:19	3.80	2.07 to 6.96	< .001
6	Yes		< 25	*IL1B + 3954C>T* (v) *BAT3S625P* (v)	78:59	2.64	1.74 to 3.99	< .001
7	Yes	Never	≥ 25	*IL1B + 3954C>T* (v) *BAT3S625P* (v)	4:7	1.12	0.32 to 3.91	.8633
8	Yes	Ever	≥ 25	*IL1B + 3954C>T* (v) *BAT3S625P* (v)	23:8	5.76	2.48 to 13.38	< .001

Abbreviations: BMI, body mass index; SNP, single nucleotide polymorphism; OR, odds ratio; w, wild type; v, variant genotype.
*Adjusted for age and sex.

variants in apoptosis genes than by environmental factors (BMI, smoking). In contrast, in the absence of reflux symptoms, EA risk was more attributable to higher BMI and smoking, two major environmental factors for EA risk. Although reflux symptoms have long been recognized as a major risk of EA, previous studies have also demonstrated that 40% or more of EA is found in patients without previous symptoms of reflux.[28,29] Additionally, higher BMI has been strongly associated with increased risk of EA, even after controlling for the severity of reflux symptoms.[30] Our current observations further suggest that mechanisms leading to EA in patients with reflux may be different from those in patients without reflux. It is therefore reasonable to consider that EA prevention strategies should be personalized according to individual's exposure to risk factor profiles.

Although statistical interactions do not necessarily imply biologic interactions, several lines of evidence suggest that our findings are biologically plausible. Reflux, smoking, and obesity have been proved to be individually associated with increased risks of EA and other cancers.[4] Obesity is a determinant of acid reflux.[31] Reflux is known to induce cyclooxygenase-2 (*COX-2*) gene expression that regulates apoptosis in the progression of Barrett's esophagus to adenocarcinoma.[21] Apoptosis gene polymorphisms have been consistently associated with cancer development at individual and multiple organ sites.[18] Cigarette smoking can induce apoptosis in a number of cell types.[23] Smoking is also known to be associated with obesity.[32] Obesity-related growth hormones, such as insulin and insulin-like growth factor (IGF) and its main binding protein, IGF binding protein-3 (IGFBP-3), regulate cell proliferation, differentiation, and apoptosis.[22] Previous studies by our group and others showed that the associations of *FASL* −844T>C and *IL1B + 3954*C>T SNPs with lung cancer were modified by smoking status.[21,33] Statistically significant interactions were found between polymorphisms in *FAS* and *FASL* and tobacco smoking in esophageal squamous cell carcinoma.[19] Interaction between the *FASL* −844T>C polymorphisms and smoking was also observed in pancreatic cancer predisposition.[34] These observations suggest that gene-environment interactions may be particularly important for apoptotic genes because their effect on cancer susceptibility is strongly determined by exposure to smoking and other environmental factors.

A major strength of our study is that gene-environment interactions were consistently identified by both parametric and nonparametric statistical models. LR analysis has the advantage of controlling for confounding variables simultaneously. When high-order interactions involving multidimensional factors are considered, there may be many sparse or empty cells, resulting in large standard errors (SEs) and an increased type I error.[35] The MDR method improves the power by converting multiple variables to a single attribute to efficiently identify potential gene-environment interactions in relatively small samples.[36] Furthermore, cross validation and permutation testing procedures in MDR reduce the chances of making type I errors as a result of multiple testing. CART analysis is an explorative, nonparametric approach that requires no assumption of a genetic model. It is a decision tree–based data mining approach to identify specific combinations of genetic and environmental factors associated with disease risk.[37] Recent studies have suggested that multiple complementary analytic strategies, including CART and MDR, improved statistical power to efficiently identify potential gene-gene and gene-environment interactions.[27] In this study, three analytic methods (LR, MDR, and CART) validate each other and emphasize the reproducibility of our findings.

Table 5. Entropy-Based Information Gain for Individual Attribute

Variable	Overall Analysis	In Subjects With History of Reflux	In Subjects Without History of Reflux
BAT3S625P	0.00	0.15	0.11
*FAS1377*G>A	0.00	0.02	0.15
*FASL844*T>C	0.00	0.00	0.00
IL1B + 3954C>T	0.01	1.06	0.57
TP53BP1D353E	0.06	0.00	0.04
Smoking	0.36	0.01	0.57
BMI	1.26	0.08	2.82
Reflux	3.44	—	—

NOTE. Entropy is a measure of the loss of information in a transmitted message or a measure of the uncertainty associated with the random variable. Information gain, also known as the relative entropy or the Kullback-Leibler entropy, is a measure of difference between two random variables. Single nucleotide polymorphisms were classified as wild type and variant genotypes. Smoking was defined as ever and never smoking. Body mass index (BMI) was categorized as ≥ 25 and < 25. Reflux was defined as yes (1) and no (0).

One of the limitations of this study is that we used a candidate polymorphism approach, which allowed us to compare our study with studies of other disease sites and focus on functional variants but therefore will not evaluate the entirety of polymorphic variation across these genes. Our study does, however, provide a guide for future studies that will replicate these findings, either through association designs or functional analysis. Second, the sample size of this study is relatively small because EA is still a rare cancer in whites. However, based on current sample size, genotype frequency, and prevalences of smoking, reflux symptoms, and higher BMI, we still have more than 80% power to detect a significant pair-wise interaction using LR or MDR analyses.[38,39] Further studies in larger populations are required to validate our findings. Third, reflux symptoms were collected on the basis of self-report only and were not validated by medical record review or with confirmatory studies such as 24-hour pH monitoring. In addition, dietary patterns and other environmental exposures were neither accounted for nor adjusted in the analyses because of missing or uncollected data. Given the strong interactions detected in this study, these potential confounders would probably have minor influence on the results.

In summary, these data confirm that reflux is strongly associated with EA risk and suggest that gene-environment interactions play more important roles than individual factors in EA development. More importantly, we observe two distinct reflux-related etiologies of EA, raising the prospect of research in personalized EA prevention strategies.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Rihong Zhai, Feng Chen, Geoffrey Liu, Xihong Lin, David C. Christiani
Financial support: Rihong Zhai, David C. Christiani
Administrative support: David C. Christiani
Provision of study materials or patients: Geoffrey Liu, Norman S. Nishioka, John C. Wain, David C. Christiani
Collection and assembly of data: Rihong Zhai, Feng Chen, Geoffrey Liu, Li Su, Kofi Asomaning, David C. Christiani
Data analysis and interpretation: Rihong Zhai, Feng Chen, Geoffrey Liu, Kofi Asomaning, Xihong Lin, Rebecca S. Heist, Chau-Chyun Sheu, David C. Christiani
Manuscript writing: Rihong Zhai, Feng Chen, Geoffrey Liu, Kofi Asomaning, Xihong Lin, Rebecca S. Heist, Chau-Chyun Sheu, David C. Christiani
Final approval of manuscript: Rihong Zhai, Feng Chen, Geoffrey Liu, Li Su, Matthew H. Kulke, Kofi Asomaning, Xihong Lin, Rebecca S. Heist, Norman S. Nishioka, Chau-Chyun Sheu, John C. Wain, David C. Christiani

REFERENCES

1. Crane SJ, Locke GR 3rd, Harmsen WS, et al: Survival trends in patients with gastric and esophageal adenocarcinomas: A population-based study. Mayo Clin Proc 83:1087-1094, 2008
2. Brown LM, Devesa SS, Chow WH: Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age. J Natl Cancer Inst 100:1184-1187, 2008
3. Pera M, Manterola C, Vidal O, et al: Epidemiology of esophageal adenocarcinoma. J Surg Oncol 92:151-159, 2005
4. Enzinger PC, Mayer RJ: Esophageal cancer. N Engl J Med 349:2241-2252, 2003
5. Cheung WY, Liu G: Genetic variations in esophageal cancer risk and prognosis. Gastroenterol Clin North Am 38:75-91, 2009
6. Zhai R, Liu G, Asomaning K, et al: Genetic polymorphisms of VEGF, interactions with cigarette smoking exposure and esophageal adenocarcinoma risk. Carcinogenesis, 29:2330-2334, 2008
7. Cheung WY, Zhai R, Kulke MH, et al: Epidermal growth factor A61G gene polymorphism, gastroesophageal reflux disease and esophageal adenocarcinoma risk. Carcinogenesis 30:1363-1367, 2009
8. Smith KJ, O'Brien SM, Smithers BM, et al: Interactions among smoking, obesity, and symptoms of acid reflux in Barrett's esophagus. Cancer Epidemiol Biomarkers Prev 14:2481-2486, 2005
9. Whiteman DC, Sadeghi S, Pandeya N, et al: Combined effects of obesity, acid reflux and smoking on the risk of adenocarcinomas of the oesophagus. Gut 57:173-180, 2008
10. Lagergren J: Controversies surrounding body mass, reflux, and risk of oesophageal adenocarcinoma. Lancet Oncol 7:347-349, 2006
11. Hanahan D, Weinberg RA: The hallmarks of cancer. Cell 100:57-70, 2000
12. McCabe ML, Dlamini Z: The molecular mechanisms of oesophageal cancer. Int Immunopharmacol 5:1113-1130, 2005
13. Sun T, Zhou Y, Li H, et al: FASL -844C polymorphism is associated with increased activation-induced T cell death and risk of cervical cancer. J Exp Med 202:967-974, 2005
14. Ferreira SB Jr, Trombone AP, Repeke CE, et al: An interleukin-1beta (IL-1beta) single-nucleotide polymorphism at position 3954 and red complex periodontopathogens independently and additively modulate the levels of IL-1beta in diseased periodontal tissues. Infect Immun 76:3725-3734, 2008
15. Rudd MF, Webb EL, Matakidou A, et al: Variants in the GH-IGF axis confer susceptibility to lung cancer. Genome Res 16:693-701, 2006
16. Sibley K, Rollinson S, Allan JM, et al: Functional FAS promoter polymorphisms are associated with increased risk of acute myeloid leukemia. Cancer Res 63:4327-4330, 2003
17. Zhang B, Sun T, Xue L, et al: Functional polymorphisms in FAS and FASL contribute to increased apoptosis of tumor infiltration lymphocytes and risk of breast cancer. Carcinogenesis 28:1067-1073, 2007
18. Zhivotovsky B, Orrenius S: Carcinogenesis and apoptosis: Paradigms and paradoxes. Carcinogenesis 27:1939-1945, 2006
19. Sun T, Miao X, Zhang X, et al: Polymorphisms of death pathway genes FAS and FASL in esophageal squamous-cell carcinoma. J Natl Cancer Inst 96:1030-1036, 2004
20. Ter-Minassian M, Zhai R, Asomaning K, et al: Apoptosis gene polymorphisms, age, smoking and the risk of non-small cell lung cancer. Carcinogenesis 29:2147-2152, 2008
21. Souza RF, Shewmake K, Pearson S, et al: Acid increases proliferation via ERK and p38 MAPK-mediated increases in cyclooxygenase-2 in Barrett's adenocarcinoma cells. Am J Physiol Gastrointest Liver Physiol 287:G743-G748, 2004
22. Khandwala HM, McCutcheon IE, Flyvbjerg A, et al: The effects of insulin-like growth factors on tumorigenesis and neoplastic growth. Endocr Rev 21:215-244, 2000
23. Zhong CY, Zhou YM, Joad JP, et al: Environmental tobacco smoke suppresses nuclear factor-kappaB signaling to increase apoptosis in infant monkey lungs. Am J Respir Crit Care Med 174:428-436, 2006
24. Benjamini Y, Yekutieli D: Quantitative trait Loci analysis using the false discovery rate. Genetics 171:783-790, 2005
25. Zhang H, Bonney G: Use of classification trees for association studies. Genet Epidemiol 19: 323-332, 2000
26. Duell EJ, Bracci PM, Moore JH, et al: Detecting pathway-based gene-gene and gene-environment interactions in pancreatic cancer. Cancer Epidemiol Biomarkers Prev 17:1470-1479, 2008
27. Chen M, Kamat AM, Huang M, et al: High-order interactions among genetic polymorphisms in nucleotide excision repair pathway genes and smoking in modulating bladder cancer risk. Carcinogenesis 28:2160-2165, 2007
28. Lagergren J, Bergstrom R, Lindgren A, et al: Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 340:825-831, 1999
29. Chow WH, Finkle WD, McLaughlin JK, et al: The relation of gastroesophageal reflux disease and its treatment to adenocarcinomas of the esophagus and gastric cardia. JAMA 274:474-477, 1995
30. Chow WH, Blot WJ, Vaughan TL, et al: Body mass index and risk of adenocarcinomas of the esophagus and gastric cardia. J Natl Cancer Inst 90:150-155, 1998

31. Delgado-Aros S, Locke GR 3rd, Camilleri M, et al: Obesity is associated with increased risk of gastrointestinal symptoms: A population-based study. Am J Gastroenterol 99:1801-1806, 2004

32. Saarni SE, Pietiläinen K, Kantonen S, et al: Association of smoking in adolescence with abdominal obesity in adulthood: A follow-up study of 5 birth cohorts of Finnish twins. Am J Public Health 99:348-354, 2009

33. Zhang X, Miao X, Sun T, et al: Functional polymorphisms in cell death pathway genes FAS and FASL contribute to risk of lung cancer. J Med Genet 42:479-484, 2005

34. Yang M, Sun T, Wang L, et al: Functional variants in cell death pathway genes and risk of pancreatic cancer. Clin Cancer Res 14:3230-3236, 2008

35. Peduzzi P, Concato J, Feinstein AR, et al: Importance of events per independent variable in proportional hazards regression analysis: II. Accuracy and precision of regression estimates. J Clin Epidemiol 48:1503-1510, 1995

36. Ritchie MD, Hahn LW, Roodi N, et al: Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138-147, 2001

37. Wu X, Gu J, Grossman HB, et al: Bladder cancer predisposition: A multigenic approach to DNA-repair and cell-cycle-control genes. Am J Hum Genet 78:464-479, 2006

38. Gauderman WJ: Sample size requirements for association studies of gene-gene interaction. Am J Epidemiol 155:478-484, 2002

39. Ritchie MD, Hahn LW, Moore JH: Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol 24:150-157, 2003

