Interactions between genetic polymorphisms in the apoptotic pathway and environmental factors on esophageal adenocarcinoma risk

I-Chen Wu1,2, Yang Zhao1, Rihong Zhai1, Chen-yu Liu1, Feng Chen1, Monica Ter-Minassian1, Kofi Asomaning1, Li Su1, Rebecca S. Heist1,3,4, Matthew H. Kulke5, Geoffrey Liu1,6,7 and David C. Christiani1,3,4,•

1Environmental and Occupational Medicine and Epidemiology Program, Department of Environmental Health, Harvard School of Public Health, Boston, MA 02115, USA, 2Division of Gastroenterology, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan, 3Department of Medicine, Massachusetts General Hospital, Boston, MA 02115, USA, 4Department of Medicine, Harvard Medical School, Boston, MA 02115, USA, 5Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA, 6Division of Medical Oncology and Hematology, Department of Medicine, Princess Margaret Hospital, Toronto, Ontario, Canada and 7Division of Applied Molecular Oncology, Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto, Ontario, Canada

•To whom correspondence should be addressed. Department of Environmental Health, Harvard School of Public Health, 665 Huntington Avenue, Building I Room 1401, Boston, MA 02115, USA. Tel: +1 617 432 3323; Fax: +1 617 432 5441; Email: dchris@hsph.harvard.edu

How genetic variations in apoptosis pathway interact with environmental factors to contribute to esophageal adenocarcinoma (EA) risk has not been comprehensively investigated. We conducted a case-only analysis in 335 Caucasian EA patients that were genotyped for 242 single nucleotide polymorphisms (SNPs) in 43 apoptotic genes. Gene–environment interactions were assessed using a two-step approach. First, random forest algorithm was used to screen for the potential interacting markers. Next, we used case-only logistic regression model to estimate the effects of gene–environment interactions on EA risk. Four SNPs (PERP rs468802; PIK3CA rs4855094, rs7644468 and TNFRSF1A rs4149579) had significant interaction with gastroesophageal reflux disease (GERD). The presence of variant alleles in TP53BP1 rs560191, CASP7 rs7907519 or BCL2 rs12454712 enhanced the risk of smoking by 2.08–2.58 times [interaction odds ratio (ORi) = 2.08–2.58, adjusted P-value (Padj) = 0.02–0.04]. Compared with patients carrying ≤1 risk genotype, the risk of GERD on EA was increased in persons with two ORi = 1.89, Padj = 0.016) or ≥3 (ORi = 4.30, Padj < 0.0001) risk genotypes. Compared with cases with ≤1 risk genotype, smoking-associated EA risk increased by 3.15 times when ≥2 risk genotypes were present (ORi = 3.15, Padj < 0.0001). In conclusion, interactions among apoptotic SNPs and GERD or smoking play an important role in EA development.

Introduction

Cumulative evidence has indicated that symptomatic gastroesophageal reflux disease (GERD), obesity, smoking and male gender are four major risk factors for developing esophageal adenocarcinoma (EA) (1,2). Although GERD is the strongest individual risk factor and subjects who have the most frequent reflux symptom carry a 4-fold EA risk, most of them follow an indolent course for the entire life (3). On the other hand, nearly 50% of EA patients do not experience GERD-associated symptoms (2,4,5). Therefore, challenges arise when considering the benefit and cost of endoscopic surveillance for EA among subjects with these risk factors, especially GERD (6). One of the explanations for this phenomenon stems from the complex interactions between lifestyle/environmental exposure and genetic factors. Since most EA cases are sporadic, genetic influences are more probably to be polymorphisms in multiple genes instead of single gene mutation. Evidence from pathway-based studies suggests that genetic variance may modulate the susceptibility of developing EA (7–9). Incorporating one’s genetic background in the clinical setting could be a better strategy for risk assessment and cancer surveillance, and possibly provide a better understanding about how complex factors contribute to EA risk jointly. However, fewer studies have assessed interactions using a more comprehensive approach.

Apoptosis, or programmed cell death, is essential for normal tissues to regulate cell number and to eliminate unwanted or aging cells as an organism develops. Mutations and single nucleotide polymorphisms (SNPs) in apoptotic pathway genes that alter the ability of the cell to undergo apoptosis may induce cancers by allowing transformed cells to keep accumulating rather than dying (10). Our recent studies indicated that genetic polymorphisms in the apoptosis pathway, by themselves or through interaction with environmental factors, play an important role in the carcinogenesis of EA (7,11). In a case–control study covering 1330 functional/tagging SNPs categorized into 14 cancer-related pathways, two apoptotic SNPs (Caspase 7 rs312707 and Caspase 9 rs4661636) were significantly associated with EA risk. Moreover, apoptosis pathway was found to be the most important one in pathway-based analysis (11). However, it is possible that other SNPs further down the significance list of individual effects exert their importance through gene–environment interaction. Here, we conducted a pathway-based case-only study to further explore the interactions between 242 apoptosis-related SNPs and those well-known EA risk factors. We applied a two-step approach to identify gene–environment interaction markers in EA cases. Our results showed that gene–environment interactions play important roles in the development of EA.

Materials and methods

Study population, interview and DNA preparation

Details of patient recruitment were described in our previous paper (11). In brief, they were Caucasian, >18 years old and histologically confirmed to have EA at Massachusetts General Hospital between 1999 and 2005 and at Dana Farber Cancer Institute between 2004 and 2005. All of them had a tumor center located at or above the gastroesophageal junction and had at least two-thirds of the bulk tumor located in the esophagus. Patients with secondary or recurrent cancers were excluded. The recruitment rate was 86%. Participants were interviewed using a modified questionnaire (12) immediately after enrollment to collect information of their demographic characteristics and smoking history. Smoking status was defined based on whether the patient smoked 1 year prior to diagnosis. Lifetime GERD-related symptoms were collected based on the questions described before (13). The presence of GERD was defined as having heartburn or regurgitation symptoms at least once per month for more than six continuous months in one’s lifetime (7).

Blood samples were collected at the time of recruitment and DNA was extracted using the PureGene DNA Isolation Kit (Gentra Systems/Quagen, Valencia, CA). This study was approved by the Human Subjects Committees of Massachusetts General Hospital, Dana Farber Cancer Institute and the Harvard School of Public Health (Boston, MA). All subjects signed the informed consent prior to study participation.

SNP selection, pathway categorization and genotyping

The criteria for SNP selection in the GoldenGate assay was described previously (11). The candidate SNPs selected in the apoptosis pathway were common SNPs with minor allele frequency (MAF) ≥ 5% in the HapMap Caucasian population (CEU). They were either missense/exonic SNPs, SNPs within untranslated regions and 2 kb 5’ of the gene or tagSNPs for genes with

Abbreviations: BMI, body mass index; CI, confidence interval; EA, esophageal adenocarcinoma; GERD, gastroesophageal reflux disease; MAF, minor allele frequency; MDA, mean decrease of accuracy; ORi, interaction odds ratio; Padj, adjusted P-value; SNP, single nucleotide polymorphism.

© The Author 2011. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

502

Downloaded from https://academic.oup.com/carcin/article-abstract/32/4/502/2463672 by guest on 03 March 2019
little functional information available. The common non-synonymous SNPs were selected using SNP500Cancer Project and International HapMap Project. Potential functional non-synonymous SNPs were selected from the Predicted Impact of Coding SNPs database (14) and FASTSNP (15). SNPs on the Illumina Cancer Panel were chosen with priority. TagSNPs were selected using the $r^2$-based Tagger program (16) with pairwise $r^2 \geq 0.80$ and MAF $\geq 5\%$ among Caucasians. Apoptosis pathway genes were chosen based on the literature reports and information in the following database, the National Center for Biotechnology Information, Genetic Association Database (17), National Institute for Environmental Health Science GeneSNPs (18) and Kyoto Encyclopedia of Genes and Genomes (19). More frequently cited genes were given greater priority.

Genotyping was performed on the Illumina GoldenGate assay at the Broad Institute (Cambridge, MA) by laboratory personnel blinded to the subjects’ clinical information. Forty-eight duplicate samples were randomly selected for quality control. The concordance rate of the replicate samples for all assays was \(>99\%\). As described previously (11), 1330 SNPs with MAF $\geq 0.01$ and passed Hardy–Weinberg equilibrium ($P \geq 0.0001$) in the control subjects were successfully genotyped. Among them, 242 SNPs (43 genes) in apoptosis pathway (supplementary Table S1 is available at Carcinogenesis Online) were used for interaction analysis in the case-only model.

Statistical analysis

We applied a two-step approach to identify the SNPs, which may have interactions with environmental factors (20). First, we used a random forest algorithm (21), with one individual environmental factor (GERD, smoking or body mass index [BMI] at age 18) as the outcome and apoptotic SNPs as the predictors, to impute the missing values ($<4\%$ in each variable) and then obtain the mean decrease of accuracy (MDA) for each SNP. The MDA measures the degree of loss if the corresponding SNP was ‘removed’ from the model and could be regarded as an ‘importance’ score for the SNP. The top important SNPs were selected based on the turning point of MDA in random forest importance plots. The turning points in the MDA plots were defined as the cut points to select important markers for subsequent interaction analyses.

In the second step, we used the ‘important SNPs’ to perform case-only logistic regression analysis using dominant genetic model (mutant carriers versus wild-type homozygotes). SNPs with adjusted $P$-value ($P_{adj}$) $< 0.05$ after adjustment for covariates and multiple comparison using false discovery rate were considered to be significant (22). To estimate cumulative risk effects of the significant SNPs, we combined the number of risk genotypes representing the degree of ‘genetic risk score’, which was included in the case-only logistic model, allowing each marker to have $\sim 100\%$ of probability to be tested for 500 times. Based on MDA plot curves, the most important SNPs that had interaction with GERD (22 SNPs), smoking (17 SNPs) and BMI (15 SNPs) were selected for subsequent case-only logistic regression analysis (Figure 1).

In case-only logistic regression analysis, we found that five SNPs (rs651662, rs648802, rs4855094, rs7644468 and rs4149579) in three genes (PERP, PIK3CA and TNFRSF1A) had significant interaction with GERD; these interactions remained significant even after adjusting for covariates and false discovery rate (Table II). Two SNPs (rs651662 and rs648802) were in the gene PERP and were in high linkage disequilibrium ($r^2 = 0.99$). Another two significant SNPs in PIK3CA (rs4855094 and rs7644468) were not tagged to each other. Subjects carrying the variant allele of rs651662, rs648802, rs4855094 or rs7644468 significantly enhanced the risk of GERD to develop EA compared with subjects carrying homozygous wild genotypes. On the contrary, the risk of GERD was significantly reduced when one carries the wide-type allele of rs4149579 ($O_R^{adj} = 0.42$, 95% CI = 0.22–0.81, $P_{adj} = 0.04$) (Table II). Since rs651662 and rs4858820 were in linkage disequilibrium ($r^2 = 0.99$) and the later was a non-synonymous SNP, whereas the former located in intron, only rs648802 was used to combine with the other three SNPs to calculate the joint effect. There was a significant cumulative risk when those four SNPs were considered jointly. The risk of GERD on EA was increased when one has two ($O_R^{adj} = 1.89$, 95% CI = 1.13–3.16) or three ($O_R^{adj} = 4.30$, 95% CI = 2.25–8.21) risk genotypes compared with those with $\leq 1$ risk genotype (Table II).

For the SNP-smoking interaction, 4 of the top 17 SNPs were identified to be significant in logistic regression models (Table III). The presence of variant alleles in TP53BP1 rs560191, rs2602141, CASP7 rs7907519 or BCL2 rs12454712 enhanced the risk of smoking by 2.08–2.58 times ($O_R^{adj} = 0.42$, 95% CI = 0.22–0.81, $P_{adj} = 0.04$), TP53BP1 rs560191 and rs2602141 were tagged to each other ($r^2 = 0.99$) and the former was non-synonymous; thus, we dropped the later in the cumulative risk analysis. Taking those three SNPs together, persons with $\geq 2$ risk genotypes had 3.15 times of interaction effect with smoking ($O_R^{adj} = 3.15$, 95% CI = 1.80–5.52, $P_{adj} < 0.0001$) than those with $\leq 1$ risk genotype (Table III). No SNPs were found to be significantly interacted with BMI in logistic regression analysis. The allele frequencies of the SNPs with significant interaction with environmental factors were shown in Supplementary Table S2, available at Carcinogenesis Online.

Results

The mean age of the EA cases was 62.9 years and 295 (88%) were males. About half of them had history of GERD symptoms and 80.3% were ever smokers. Other demographic characteristics were shown in Table I. In random forest analysis, MDA scores were obtained by SNP–GERD, SNP–smoking and SNP–BMI models, respectively. To minimize random bias, 100 000 trees were constructed in each RF model, allowing each marker to have $\sim 100\%$ of probability to be tested for 500 times. Based on MDA plot curves, the most important SNPs that had interaction with GERD (22 SNPs), smoking (17 SNPs) and BMI (15 SNPs) were selected for subsequent case-only logistic regression analysis (Figure 1).

Gene–environment interaction and EA risk

Discussion

This is a comprehensive study covering 242 apoptotic SNPs to evaluate the gene–environment interaction in EA. Case-only study with random forest analysis has been proven to be a reliable method to investigate gene–environment interaction in complex diseases (20). The EA patients in this study were incident cases; therefore, the ORs estimated the interaction under the assumption that the genetic and

<table>
<thead>
<tr>
<th>Table I. Characteristics of 335 EA cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics</td>
</tr>
<tr>
<td>Age (years, mean ± SD)</td>
</tr>
<tr>
<td>Gender, male (%)</td>
</tr>
<tr>
<td>GERD*, yes (%)</td>
</tr>
<tr>
<td>Smoking, yes (%)</td>
</tr>
<tr>
<td>Smoking (pack-years), mean ± SD, medium (minimum to maximum)</td>
</tr>
<tr>
<td>BMI at age 18, (kg/m²), mean ± SD</td>
</tr>
<tr>
<td>BMI ≥ 25 at age 18 (%)</td>
</tr>
<tr>
<td>Stage</td>
</tr>
<tr>
<td>I–IIA</td>
</tr>
<tr>
<td>IIB–IV</td>
</tr>
</tbody>
</table>

*7 (2%) missing and imputed.
environment factors were independent. In addition to identifying potential modifiers, we also demonstrated increased gene–GERD and gene–smoking interactions when those significant SNPs were considered jointly. Our results provide more evidence to how apoptotic genes modify the risk of GERD and smoking on EA risk. Dominant model of each SNP was used to calculate the OR because it was more convenient and the sample size was limited in this study. We also used case–control model to test the interaction between smoking and the four significant SNPs found in this case-only analysis. The control subjects were described in our previous study (11). Three of them had very consistent results (rs560191, rs260214 and rs12454712; ORi 5 2.11–2.28, P-value 5 0.02–0.04). Although we did not find significant interaction between CASP7 rs7907519 and smoking in case–control model, the trend was consistent (ORi 5 1.20, 95% CI, 0.58–2.48, P-value 5 0.61). Because GERD information was missing in 50% of our control subjects (11), we cannot perform case–control analysis for genotype–reflux interaction.

Evading apoptosis, a characteristic of transformed cells, has been well documented in Barrett’s carcinogenesis (23,24). Gastric acid and bile acid are two principle components of refluxate that cause chronic esophageal inflammation and Barrett’s transformation. Direct exposure of an EA cell line to acid leads to suppression of apoptotic genes, which may occur via p53-dependent mechanism (25). TP53 tumor suppressor induces the intrinsic apoptotic pathway and activates many downstream genes, including PERP (TP53 apoptosis effector related to PMP-22) (26). Previous study also suggests that there is a reciprocal regulation between p53 and the PI3K–AKT pathway (27). The PIK3CA family regulates many normal cell functions, including cell proliferation, survival and growth. It is also known to be anti-apoptotic in many different cell types (28). Amplification of PIK3CA, which encodes the 110 kDa subunit of PI3K, has been found in many human cancers (29,30). Treatment with a PI3K inhibitor decreases proliferation and increases apoptosis; therefore, it has become a target for cancer therapy (28). Genetic polymorphisms of PIK3CA have been studied in the susceptibility of ovarian cancer (31). In addition to PERP and PIK3CA, we also found a significant interaction between the TNFRSF1A polymorphism and GERD. TNFRSF1A is one of the cognate receptors that binds to tumor necrosis factor superfamily ligands to activate the extrinsic apoptotic pathway (32). TNFRSF1A rs149570 has been linked to the survival of early-stage non-small cell lung cancer (33). However, no studies investigated the association between PERP, PIK3CA or TNFRSF1A polymorphism and esophageal cancer risk.

Our data revealed that polymorphisms in CASP7, TP53BP1 and BCL2 modified the risk of smoking to develop EA. Long-term exposure to carcinogens in cigarette smoke such as polycyclic aromatic hydrocarbons leads to DNA damage, which accumulates if a cell evades cell cycle regulation and apoptotic mechanism. Caspase activation plays a primary role in the apoptotic cascade. A previous report has shown an inactivation mutation of CASP7 in esophageal cancer cells (34). Polymorphisms in CASP7 have been examined in several human cancers (35,36) but not in esophageal cancer until our recent report (11). In this study, we identified a significant interaction between CASP7 rs7907519 and smoking in EA cases. Such information adds more evidence to the importance of genes in caspase cascade in the susceptibility of EA. TP53BP1 binds to TP53 and plays a role in responses to DNA damage. A study in Japan revealed that TP53BP1 Asp353Glu (rs560191) and TP53 Arg72Pro polymorphism had significant interaction in lung cancer risk. However, no gene–smoking interaction was found (37). BCL2 protein is one of the products of BCL2 family and has anti-apoptotic function. It has been found to aberrantly expressed in Barrett’s carcinogenesis (38). Polymorphisms in BCL2 were reported in many human cancers such as ovarian cancer, leukemia and esophageal cancer (39–41). However, no study examined the association of TP53BP1 or BCL2 polymorphism and EA risk among Caucasians.
In conclusion, our data suggest that the interaction among apoptotic SNPs and GERD on EA risk

Table II. Case-only interaction between apoptotic SNPs and GERD on EA risk

<table>
<thead>
<tr>
<th>SNPs</th>
<th>GERD</th>
<th>ORI* (95% CI)</th>
<th>P_adj</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No, n (%)</td>
<td>Yes, n (%)</td>
<td></td>
</tr>
<tr>
<td>PERP rs651662</td>
<td>GG</td>
<td>68 (41.97)</td>
<td>43 (24.86)</td>
</tr>
<tr>
<td></td>
<td>GA + AA</td>
<td>94 (58.03)</td>
<td>130 (75.14)</td>
</tr>
<tr>
<td>PERP rs648802</td>
<td>CC</td>
<td>67 (41.36)</td>
<td>43 (24.86)</td>
</tr>
<tr>
<td></td>
<td>CG + GG</td>
<td>95 (58.64)</td>
<td>130 (75.14)</td>
</tr>
<tr>
<td>PIK3CA rs4855094</td>
<td>GG</td>
<td>151 (93.21)</td>
<td>145 (83.82)</td>
</tr>
<tr>
<td></td>
<td>GA + AA</td>
<td>11 (6.79)</td>
<td>28 (16.18)</td>
</tr>
<tr>
<td>TNFRSF1A rs4149579</td>
<td>AA</td>
<td>123 (75.93)</td>
<td>109 (63.00)</td>
</tr>
<tr>
<td></td>
<td>GA + GG</td>
<td>39 (24.07)</td>
<td>64 (37.00)</td>
</tr>
</tbody>
</table>

*Adjusted for age, sex, BMI at age 18 (also false discovery rate for P_adj).

Table III. Case-only interaction between apoptotic SNPs and smoking on EA risk

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Smoking</th>
<th>ORI* (95% CI)</th>
<th>P_adj</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No, n (%)</td>
<td>Yes, n (%)</td>
<td></td>
</tr>
<tr>
<td>TP53BP1 rs560191</td>
<td>CC</td>
<td>44 (66.67)</td>
<td>119 (44.24)</td>
</tr>
<tr>
<td></td>
<td>CG + GG</td>
<td>22 (33.33)</td>
<td>150 (54.76)</td>
</tr>
<tr>
<td>TP53BP1 rs2602141</td>
<td>AA</td>
<td>43 (65.15)</td>
<td>118 (43.87)</td>
</tr>
<tr>
<td></td>
<td>AC + CC</td>
<td>23 (34.85)</td>
<td>151 (56.13)</td>
</tr>
<tr>
<td>CASP7 rs7907519</td>
<td>CC</td>
<td>30 (45.45)</td>
<td>76 (28.25)</td>
</tr>
<tr>
<td></td>
<td>CA + AA</td>
<td>36 (55.45)</td>
<td>193 (71.75)</td>
</tr>
<tr>
<td>BCL2 rs12454712</td>
<td>TT</td>
<td>36 (54.55)</td>
<td>99 (36.80)</td>
</tr>
<tr>
<td></td>
<td>TC + CC</td>
<td>30 (45.45)</td>
<td>170 (63.20)</td>
</tr>
</tbody>
</table>

*Adjusted for age, sex, BMI at age 18 (<25 versus ≥25) and GERD (+/−) (also false discovery rate for P_adj).

Acknowledgements

The authors thank the physicians and patients from the Massachusetts General Hospital Thoracic Oncology Center and the Dana Farber Cancer Institute GI Oncology Center for their cooperation and participation. The authors also thank researchers from Excellence for cancer research center, Department of Health, Executive Yuan, Taiwan, ROC (grant, DOH100-TD-C-111-002).

Conflict of Interest Statement: All authors disclosed no conflict of interest; the institutes, which provided the grant supports did not participate in the study design, the collection, analysis or interpretation of data.

References


studies are needed to validate the results and elucidate how the genes interact with environmental factors to contribute to the cancer risk.


Received November 2, 2010; revised December 17, 2010; accepted December 20, 2010.