A Large-scale genetic association study of esophageal adenocarcinoma risk

Chen-yu Liu¹, Michael C.Wu², Feng Chen¹, Monica Ter-Minassian¹, Kofi Asomaning¹, Rihong Zhai¹, Zhaoxi Wang¹, Li Su¹, Rebecca S.Heist¹,², Matthew H.Kulke³, Xihong Lin², Geoffrey Liu¹,⁵,⁶ and David C.Christiani¹,³,⁵

¹Environmental and Occupational Medicine and Epidemiology Program, Department of Environmental Health, Harvard School of Public Health, Boston, MA 02115, USA, ²Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA, ³Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA, ⁴Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA, ⁵Division of Medical Oncology and Haematology, Department of Medicine, Princess Margaret Hospital, Toronto, Ontario, Canada MSG 2M9 and ⁶Division of Applied Molecular Oncology, Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto, Ontario, Canada MSG 2M9

To whom correspondence should be addressed. Environmental and Occupational Medicine and Epidemiology Program, 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 3441; Email: dchris@hsph.harvard.edu

The incidence of esophageal adenocarcinoma (EA) has been increasing rapidly, particularly among white males, over the past few decades in the USA. However, the etiology of EA and the striking male predominance is not fully explained by known risk factors. To identify susceptible genes for EA risk, we conducted a pathway-based candidate gene association study on 335 Caucasian EA cases and 319 Caucasian controls. A total of 1330 single-nucleotide polymorphisms (SNPs) selected from 354 genes were analyzed using an Illumina GoldenGate assay. The genotyped common SNPs include missense and exonic SNPs, SNPs within untranslated regions and 2 kb 5′ of the gene, and tagSNPs for genes with little functional information available. Logistic regression adjusted for potential confounders was used to assess the genetic effect of each SNP on EA risk. We also tested gene–gender interactions using the likelihood ratio tests. We found that the genetic variants in the apoptosis pathway were significantly associated with EA risk after correcting for multiple comparisons. SNPs of rs3127075 in Caspase-7 (CASP7) and rs4661636 in Caspase-9 (CASP9) genes that play a critical role in apoptosis were found to be associated with an increased risk of EA. A protective effect of SNP rs572483 in the progesterone receptor (PGR) gene was observed among women carrying the variant G allele [adjusted odds ratio (OR) = 0.19; 95% confidence interval (CI) = 0.08–0.46] but was not observed among men (adjusted OR = 1.38; 95% CI = 0.95–2.00). In conclusion, this study suggests that the genetic variants of CASP7 and CASP9 in the apoptosis pathway may be important predictive markers for EA susceptibility and that PGR in the sex hormone signaling pathway may be associated with the gender differences in EA risk.

Introduction

The incidence of esophageal adenocarcinoma (EA) is increasing more rapidly than any other cancer in the USA (1,2) and is typically detected at an advanced stage with a 5-year mortality of >80% (3,4). EA is found seven times more often in men than in women in almost all developed countries. In addition, it is very unusual for women before menopause to develop EA, suggesting the possibility that female sex hormones may protect against the development of EA (5–7). However, the etiology of EA and the striking male predominance is not fully understood.

Both genetic and environmental factors are suspected contributors to cancer development. Barrett’s esophagus is a syndrome that develops among a subgroup of patients with chronic gastroesophageal reflux disease (GERD) (8). Although Barrett’s esophagus is the only known premalignancy of EA, only a small fraction (0.5–1% annually) of Barrett’s esophagus patients subsequently develop adenocarcinoma (9). The other potential risk factors include smoking, alcohol and obesity (10–13). However, only a fraction of individuals with these risk factors develop EA. For instance, the progression from GERD to Barrett’s and Barrett’s to carcinoma occurs among very few patients (14).

Systematic review of published common genetic variants and EA risk have been reported (15). While these studies have provided some promising results, a large-scale candidate gene analysis has not been performed to evaluate genetic susceptibility to EA risk. Therefore, we conducted a pathway-based candidate gene association study on 335 Caucasian EA cases and 319 Caucasian controls. A total of 1330 single-nucleotide polymorphisms (SNPs) in 354 genes were successfully genotyped using an Illumina GoldenGate assay.

Materials and methods

Study population and DNA sample collection

This study was approved by the Human Subjects Committee of Massachusetts General Hospital, Dana-Farber Cancer Institute and the Harvard School of Public Health (Boston, MA). Cases and controls were over the age of 18 years. Written informed consent was obtained from all subjects prior to study participation. Patients with incident, historically confirmed EA were recruited at Massachusetts General Hospital between 1999 and 2005 and at Dana-Farber Cancer Institute between 2004 and 2005; patients with secondary or recurrent tumors were excluded. There was no restriction on tumor stage. Both Massachusetts General Hospital and Dana-Farber Cancer Institute are networked hospitals with similar practice patterns. Controls were selected from healthy friends and non-blood-related family members (usually spouses) of a multi-cancer susceptibility study conducted between 1999 and 2003 at Massachusetts General Hospital. All controls never had any diagnosis of cancer (16). Details of the control population have been described previously (16,17).

An in-person interview was conducted by a trained interviewer immediately after enrollment. A modified questionnaire (18) was used to obtain information on subjects’ socio-demographic characteristics and a detailed smoking and alcohol exposure assessment.

Blood samples were collected from all participants at the time of recruitment. DNA was extracted from peripheral blood samples using the Puregene DNA Isolation Kit (Genta Systems/QIAGEN, Valencia, CA).

SNP selection and genotyping

The selected candidate genes were in the pathways of carcinogens/procarcinogens metabolism, DNA repair, cell cycle control, apoptosis, inflammation, cell growth, cell signaling, angiogenesis, metastasis, sex hormone signaling, immunity, molecular transport, DNA methylation and telomere maintenance. These genes were selected based on published evidence of their relationship to carcinogenesis (19–30).

The genotyped common SNPs include missense and exonic SNPs, SNPs within untranslated regions and 2 kb 5′ of the gene and tagSNPs for genes with little functional information available. The common non-synonymous SNPs were selected using SNP500Cancer Project (http://snp500cancer.nci.nih.gov) and International HapMap Project (http://hapmap.org). Potential functional non-synonymous SNPs from cancer-associated genes were selected from the PICS (Predicted Impact of Coding SNPs) database (http://www.irc.ucl.ac.uk/rgenetics/Gene/MolPortGen_PICS_database.htm) (31) and FASTSNP (http://fastsnp.ibms.sinica.edu.tw) (32). SNPs on the Illumina Cancer Panel were selected with priority (http://www.illumina.com/products/cancer_snp_panel.ilmm). TagSNPs were selected using the r2-based Tagger program (33) with pairwise $r^2 \geq 0.80$ and minor allele frequency $\geq 5\%$ in the HAPMAP Caucasian population (CEU). SNPs probably to be problematic for genotyping...
were eliminated based on the Illumina genotyping performance score for BeadArray assays.

Genotyping was performed on the Illumina GoldenGate assay at the Broad Institute, Cambridge, MA, by laboratory personnel without the knowledge of case-control status. Duplicate samples (n = 48) were randomly selected for quality control. The concordance rate for all replicate samples for all assays was >99%.

A total of 1536 SNPs were genotyped. After exclusion of 162 SNPs from unstable assays, 24 SNPs with low minor allele frequency (<0.01) and 20 SNPs with substantial deviation from Hardy–Weinberg equilibrium (P < 0.0001) and 1330 SNPs in 354 genes were left for analysis (Supplementary Table S1 is available at Carcinogenesis Online).

**Statistical analysis**

Among the study subjects genotyped, 31 (8%) of the EA cases and 21 (6%) of the controls were removed from the data analysis since they had >10% missing genotype information. We further restricted the analysis to Caucasians (>90% of the study subjects) and the study subjects with complete information on age, gender and smoking status (97% cases and 93% controls).

Genotype frequencies among controls were tested for departure from Hardy–Weinberg equilibrium by the Pearson χ² test (degree of freedom = 1). The selected principal characteristics were evaluated using χ², Fisher exact and r-tests as appropriate.

Several potential confounders were adjusted for in the multiple regression models: age, gender, smoking status, body mass index at 18 years old, alcohol intake and prior history of GERD. As no confounding was seen (changes in ‘estimated effects’ <0.2) by these variables alone or together on the results of interest, the results presented are only adjusted for age and gender.

To control the overall type I error rate while adjusting for correlation between SNPs, empirical P-values based on global random permutation tests were computed. Specifically, we jointly permuted the case-control status and the demographic covariates of each subject and recorded the minimal P-value for each permuted dataset. The distribution of the minimal P-values obtained from 10 000 permuted datasets was used to derive the empirical significance of the observed test statistic (Ppermutation). The adjusted global-wise P-values were determined as Padjusted = P(Ppermutation ≥ Ppermutation). All reported P-values are based on two-sided tests. An adjusted P-value <0.05 was considered to be statistically significance. Data were analyzed using the SAS® software version 9.1 (SAS Institute, Inc., Cary, NC) and R version 2.1.0 (The R Project for Statistical Computing, Vienna, Austria; http://www.r-project.org/).

Pathway-based analysis was performed by the logistic kernel machine regression test for genetic data (Michael, C., Wu,P.K., Michael,P, Epstein, Deanne, M.Taylor, Stephen,J,Chanock, David,J,Hunter and Xihong Lin, in preparation). Specifically, for each pathway, we analyzed all the SNPs within the pathway simultaneously and tested for their joint effect on the outcome, while adjusting for the potential confounders (age and gender). We applied the logistic kernel machine regression test of Liu et al. (34) using the identical-by-state kernel function (35) to measure genetic pathway similarity between subjects. This approach tests whether pairwise genetic similarity in the pathway (defined as the proportion of alleles shared identical-by-state across the pathway) is correlated with phenotypic similarity while adjusting for confounding covariates. Subjects with missing genotype values, for the SNPs in the pathway, were initially omitted from the pathway analysis. Pathway analyses in which the missing values were first imputed to the most common genotype (for each SNP) generated results that were concordant with the complete-case-only pathway analyses. The Bonferroni correction (36) was used to control for multiple comparisons.

Individual SNP analysis was performed among the 1330 SNPs by using multiple logistic regressions with additive genetic effect models. Dominant genetic effect models were used to test for interactions. Likelihood ratio tests were used to test the significance of the interaction terms.

**Results**

**Demographics**

A total of 335 cases and 319 controls were included in the final analyses. The distributions of selected characteristics among study subjects are summarized in Table I. Cases were older than controls and had a greater proportion of males. All cases and controls were Caucasians. Body mass index at 18 years was higher in cases than in controls. Not surprisingly, more smokers, alcohol drinkers and higher GERD prevalence were seen in the cases, as these are known risk factors for the development of EA.

**Pathway and SNP effects**

A total of 1330 SNPs in 354 genes were successfully genotyped. Under the pathway-based analysis, we categorized the 354 genes into 14 pathways (Table II): carcinogens/procarcinogens metabolism, DNA repair, cell cycle control, apoptosis, inflammation, cell growth, cell signaling, angiogenesis, metastasis, sex hormone signaling, immunity, molecular transport, DNA methylation and telomere maintenance. As expected, several genes overlap between pathways. Only the apoptosis pathway remained significant after correcting for multiple comparisons (nominal P = 0.0004, Bonferroni P = 0.006).

<table>
<thead>
<tr>
<th>Table I. Distribution of study subjects by demographic and risk factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Age (years) (range)</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Female (%)</td>
</tr>
<tr>
<td>Smoking</td>
</tr>
<tr>
<td>Ever (%)</td>
</tr>
<tr>
<td>Body mass index at age 18</td>
</tr>
<tr>
<td>&lt;25 (%)</td>
</tr>
<tr>
<td>≥30 (%)</td>
</tr>
<tr>
<td>Alcohol intake</td>
</tr>
<tr>
<td>Ever (%)</td>
</tr>
<tr>
<td>Prior history of GERD</td>
</tr>
<tr>
<td>Yes (%)</td>
</tr>
</tbody>
</table>

*Number and percent of group (%). Percents are rounded.

<table>
<thead>
<tr>
<th>Table II. Results of the pathway analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathway</td>
</tr>
<tr>
<td>Carcinogens/procarcinogens metabolism</td>
</tr>
<tr>
<td>DNA repair</td>
</tr>
<tr>
<td>Cell cycle control</td>
</tr>
<tr>
<td>Apoptosis</td>
</tr>
<tr>
<td>Inflammation</td>
</tr>
<tr>
<td>Cell growth</td>
</tr>
<tr>
<td>Cell signaling</td>
</tr>
<tr>
<td>Angiogenesis</td>
</tr>
<tr>
<td>Metastasis</td>
</tr>
<tr>
<td>Sex hormone signaling</td>
</tr>
<tr>
<td>Immunity</td>
</tr>
<tr>
<td>Molecular transport</td>
</tr>
<tr>
<td>DNA methylation</td>
</tr>
<tr>
<td>Telomere maintenance</td>
</tr>
</tbody>
</table>

*Pathway-based analysis assessed by the logistic kernel machine regression adjusted for age and sex. Significant pathway effect after correcting for multiple comparisons is indicated in bold.
We tested for possible joint effect between the polymorphisms rs3127075 in CASP7 and rs4661636 in CASP9 as caspase-7 is one of the downstream effectors activated by caspase-9 in the apoptosis pathway. Compared with the subjects with homozygous wild type of both CASP7 rs3127075 and CASP9 rs4661636, the subjects who carry the variant alleles of both CASP7 rs3127075 and CASP9 rs4661636 have 3.52 times higher risk of having EA (Table III).

Interaction effects

We conducted further analyses to evaluate the association of these SNPs with smoking and gender. There was no significant gene–smoking interaction effect after adjusting for multiple comparisons. Significant gene–gender interaction was observed among the polymorphism rs572483 in PGR (Table IV; nominal P < 0.0001, permutation P < 0.05). A protective effect was observed among women carrying the variant G allele (adjusted OR = 0.19; 95% CI = 0.08–0.46) but was not observed among men (adjusted OR = 1.38; 95% CI = 0.95–2.00).

Discussion

To our knowledge this is the first large-scale systematic pathway-based study to investigate the role of common genetic variation in susceptibility to EA. In this case–control study, we analyzed 14 pathways that are potentially important in carcinogenic processes and found that the genetic variants in the apoptosis pathway to be significantly associated with EA risk after correction for multiple comparisons. Polymorphisms of rs3127075 in CASP7 (nominal P = 0.0001, permutation P < 0.05) and rs4661636 in CASP9 (nominal P = 0.0004, permutation P = 0.2) are the top SNPs in the apoptosis pathway associated with EA risk.

Apoptosis, a genetically controlled process of programmed cell death, provides a protective mechanism by removing DNA-damaged cells that could either interfere with normal function or lead to neoplastic proliferation (37,38). Resistance to apoptosis is an important indicator related to esophageal carcinogenesis and chemotherapy resistance (38–42). Apoptosis activation occurs through the caspase family in both mitochondrial and death receptor pathways (43). CASP7 knockout mice were reported to be resistant to endotoxin-induced apoptosis (44). An inactivation mutation of the caspase genes, CASP7, has been found in esophageal cancer cells (42). It is plausible that SNPs in multiple genes in this pathway could affect cancer risk, as caspase-8, -9 and -10 initiate a cascade of caspase activation by cleaving downstream effectors such as caspase-7, which consequently causes cell death (45,46). Polymorphisms in CASP7 and CASP9 genes have found to be associated with several diseases including cancers (47–50) but have not been investigated in the EA risk. Polymorphisms of rs3127075 in CASP7 and rs4661636 in CASP9 are the top SNPs in the apoptosis pathway were associated with EA risk in our study. Our tagSNP CASP7 rs12416109 is in complete linkage disequilibrium with CASP7 rs2227309 (r2 = 1; D' = 1), which has been found to be associated with different CASP7 expression level and rheumatoid arthritis by Garcia-Lozano et al. (51), although the association of CASP7 rs2227309 with rheumatoid arthritis was not replicated in another study (52). The tagSNP CASP7 rs12416109 in our study did not show significant association with EA risk (data not shown).

Another interesting finding of the study is the gene–gender interaction effect of the PGR gene. A protective effect was observed among women carrying the variant G allele (adjusted OR = 0.19; 95% CI = 0.08–0.46) but was not observed among men (adjusted OR = 1.38; 95% CI = 0.95–2.00). Since the high male to female rate ratios are not probably to be explained by the different lifestyle risk factors among males and females (55), the different levels of hormones between genders may provide some further insights into the EA mechanisms. These results should be treated with caution due to the limited female cases in the current study. Further studies are needed to investigate the role of rs572483 in PGR gene in relation to the EA susceptibility.

The controls were selected to represent the general middle-aged adult distribution of Massachusetts population (56). While the case and control groups differed in several risk factors (Table I), we did not find the observed genetic associations confounded by these risk factors. No confounding bias was detected using multiple logistic regression models adjusting for these variables alone or in combination on the associations of interest.

It is important that the observed associations be confirmed in a larger, independent study. Although functionality is not known for all of the genotyped SNPs on our platform, our results are biologically plausible given the connections between the variants in apoptosis

### Table III. The frequencies and adjusted ORs of genotypes of CASP7 rs3127075 and CASP9 rs4661636 to the susceptibility of EA

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Case (%)</th>
<th>Control (%)</th>
<th>Adjusted OR (95% CI)a</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASP7 rs3127075</td>
<td>GC or CC</td>
<td>67 (20.0)</td>
<td>99 (31.0)</td>
</tr>
<tr>
<td>CASP7 rs4661636</td>
<td>GC or CC</td>
<td>51 (15.2)</td>
<td>41 (12.9)</td>
</tr>
<tr>
<td>CASP9</td>
<td>GC or CC</td>
<td>122 (26.4)</td>
<td>133 (41.7)</td>
</tr>
<tr>
<td></td>
<td>GC or CC</td>
<td>95 (19.8)</td>
<td>46 (14.4)</td>
</tr>
</tbody>
</table>

aORs assessed by logistic regression adjusted for age and gender.

P < 0.0001.

### Table IV. The frequencies and adjusted odds ratio of PGR rs572483 polymorphism stratified by gender

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case (%)</td>
<td>Control (%)</td>
</tr>
<tr>
<td>PGR rs572483p</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>30 (75.0)</td>
<td>55 (39.9)</td>
</tr>
<tr>
<td>AG or GG</td>
<td>10 (25.0)</td>
<td>83 (60.1)</td>
</tr>
</tbody>
</table>

aORs assessed by logistic regression adjusted for age.

Significant interaction effect (nominal P < 0.0001, permutation P < 0.05) according to the likelihood ratio tests.

P < 0.0005.
C.-y.Liu et al.

genes and EA risk. However, associations with any specific SNP should be interpreted with caution until functionality is identified and these results are replicated.

Supplementary material

Supplementary Table S1 can be found at http://carcin.oxfordjournals.org/

Funding

National Institutes of Health (R01CA109193, R03CA110822, R01CA074386, ES00002); Canadian Institutes of Health Research operating grant; Kevin Jackson Memorial Fund; Alan B. Brown Chair in Molecular Genomics; National Cancer Institute of Canada Dorothy Lamont Award; Posluns Family Fund Foundation; Flight Attendant Medical Research Institute Award No. 062459. The Broad Institute Lamont Award; Posluns Family Fund Foundation; Flight Attendant Medical Research Institute Award No. 062459. The Broad Institute

Acknowledgements

The authors thank the physicians, staff and patients from the Massachusetts General Hospital Thoracic Oncology Center and the Dana-Farber Cancer Institute, Gastrointestinal Oncology Center for their cooperation and participation and Dr Daniel Mirel of the Broad Institute. We thank Andrea Shafar and Salvatore Mucci for technical assistance.

Conflict of Interest Statement: None declared.

References


Received January 20, 2010; revised April 2, 2010; accepted April 29, 2010