Axonal guidance signaling pathway interacting with smoking in modifying the risk of pancreatic cancer: a gene- and pathway-based interaction analysis of GWAS data

Hongwei Tang, Peng Wei1, Eric J. Duell2, Harvey A. Risch3, Sara H. Olson4, H. Bas Bueno-de-Mesquita5, Steven Gallinger2, Elizabeth A. Holly7, Gloria Petersen6, Paige M. Bracci2, Robert R. McWilliams8, Mazda Jenab9, Elio Riboli10, Anne Tjonneland11, Marie Christine Boutron-Ruault12, Rudolph Kaaks13, Dimitrios Trichopoulou14,16, Salvatore Panico17, Malin Sund18, Petra H. M Peeters19, Kay-Tee Khaw21, Christopher I Amos22,23 and Donghui Li4

Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA, 2Division of Biostatistics and Human Genetics Center, School of Public Health, University of Texas Health Science Center, Houston, TX 77030, USA, 3Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain, 4School of Public Health, Yale University, New Haven, CT 06520, USA, 5Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA, 6The Netherlands and Department of Gastroenterology and Hepatology, National Institute for Public Health and the Environment (RIVM), Bilthoven, University Medical Center Utrecht, Utrecht, The Netherlands, 7Department of Epidemiology, Public Health and Primary Care, Julius Center for Health Sciences and Primary Care, Utrecht, the Netherlands, 8School of Public Health, University of California San Francisco, San Francisco, CA 94143, USA, 9Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA, 10Unit of Nutrition and Cancer, International Agency for Research on Cancer, Lyon, France, 11Division of Epidemiology, Public Health and Primary Care, Imperial College London, London, UK, 12Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark, 13INSERM, Centre for Research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women’s Health team, F-94805, Villejuif, France, 14Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany, 15Department of Epidemiology, Harvard School of Public Health, 677 Huntington, Avenue, Boston, MA 02115, USA, 16Bureau of Epidemiologic Research, Academy of Athens, 28 Panepistimiou Street, Athens, GR-106 79, Greece, 17Hellenic Health Foundation, Kaisareias 13 & Alexandroupoleos Street, GR-115 24, Athens, Greece, 18Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy, 19Department of Surgery, Umeå University Hospital, Umeå, Sweden, 20Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands, 21Division of Epidemiology, Public Health and Primary Care, Imperial College London, London, UK, 22School of Clinical Medicine, University of Cambridge, UK and 23Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

*Present address: Department of Community and Family Medicine Geisel School of Medicine, Dartmouth College, Lebanon, NH 03766, USA

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Abbreviations: AR, adrenergic receptors; B-H, Benjamini-Hochberg; BMI, body mass index; CI, confidence interval; FDR, false discovery rate; GO, gene ontology; GWAS, genome-wide association study; IPA, ingenuity pathway analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; LRT, likelihood ratio test; PCA, principal component analysis; SNP, single nucleotide polymorphism.

Introduction

Pancreatic cancer is the fourth leading cause of cancer death resulting in >37 600 human deaths each year in the USA (1). Epidemiological studies have identified cigarette smoking as the most consistent modifiable risk factor for this lethal disease with a relative risk of ~2.0 and contribution to 25% of the total cases (2). Because only a small portion of smokers is affected by pancreatic cancer, genetic factors that influence individual susceptibility to smoking-associated pancreatic cancer have previously not been examined at the genome-wide level. Although a few case–control studies have previously found some associations of carcinogen metabolic genes or DNA repair genes with risk of pancreatic cancer among smokers, these findings were either not replicated or could not be replicated (3). Genome-wide association studies (GWAS) have identified novel susceptible loci or chromosome regions in pancreatic cancer (4,5), and these data have also provided an opportunity to uncover genes that have not been previously thought to be related to this disease. However, genetic factors that influence individual susceptibility to smoking-induced pancreatic cancer have previously not been examined at the genome-wide level.

We have previously analyzed the interactions of two other known risk factors for pancreatic cancer, obesity and diabetes, at the pathway, gene and single nucleotide polymorphism (SNP) levels using the GWAS data and exposure information from the Pancreatic Cancer Case Control Consortium (6). A significant interaction of obesity and chemokine signaling pathway was observed. Analogous to scenarios in these analyses, we expect that cigarette smoking may significantly interact with a group of functionally related genes (pathway) in modifying pancreatic cancer risk. In this study, we explored hierarchal interactions of genetic factors with smoking at pathway/gene/SNP level using an agnostic approach. In addition, we used the candidate gene approach and examined GWAS top hits that have been identified in smoking-related cancers, i.e. cancers of the lung, bladder and head and neck. Furthermore, based on the previous observation that nicotine dependence genes may modify the risk of smoking-related diseases, e.g. lung cancer and chronic obstructive pulmonary disease (7,8), we also examined nicotine-dependence-related gene ontology (GO) pathways in this study.

We examined 172 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways, 3 manually curated gene sets, 3 nicotine dependency gene ontology pathways, 1791 genes and 468 114 SNPs. None of the individual pathway/gene/SNP showed significant interaction with smoking after adjusting for multiple comparisons. Six KEGG pathways showed nominal interactions (P < 0.05) with smoking, and the top two are the pancreatic secretion and salivary secretion pathways (major contributing genes: RAB8A, PLCB and CTRB1). Nine genes, i.e. ZBED2, EXO1, PSG2, SLC36A1, CLSTN1, MTHFD3, FAT2, IL10RB and ATXN2 had \( P_{\text{interaction}} < 0.00003 \). Five intergenic region SNPs and two SNPs of the EVC and KCNIP4 genes had \( P_{\text{interaction}} < 0.00003 \). In IPA analysis of genes with nominal interactions with smoking, axonal guidance signaling (\( P = 2.12 \times 10^{-7} \)) and -adrenergic signaling (\( P = 2.52 \times 10^{-8} \)) were significantly overrepresented canonical pathways. Genes contributing to the axon guidance signaling pathway included the SLIT/ROBO signaling genes that were frequently altered in pancreatic cancer. These observations need to be confirmed in additional data set. Once confirmed, it will open a new avenue to unveiling the etiology of smoking-associated pancreatic cancer.
**Materials and methods**

**Study population and data set**

The study population was a subset of seven studies participating in the previously conducted GWAS of the Pancreatic Cancer Case Control Consortium (PanScan) and the Pancreatic Cancer Case Control Consortium (4,5) containing six case–control studies conducted at MD Anderson Cancer Center, Yale University, Mayo Clinic, Memorial Sloan-Kettering Cancer Center, University of California at San Francisco and University of Toronto, and one nested case–control study from the European Prospective Investigation into Cancer and Nutrition cohort. Cases were patients pathologically diagnosed with pancreatic adenocarcinoma; controls were free of pancreatic cancer on recruitment and frequency matched to cases in each institution on sex, birth year and self-reported race/ethnicity. GWAS was conducted at the National Cancer Institute Genotyping Facility using the Illumina HumanHap550-Duo SNP arrays and Illumina Human 610-Quad arrays (4.5). We downloaded the genotype data of 562,000 and 621,000 SNPs for 4,195 study subjects (2,163 cases and 2,232 controls) from Database of Genotypes and Phenotypes (dbGaP) website (http://www.ncbi.nlm.gov/gap) with the approval of MD Anderson Institutional Review Board. The quality control process removed SNPs deviating from Hardy–Weinberg equilibrium (p-value < 0.001) or with minor allele frequency < 5%, resulting in a final data set containing 468,114 SNPs that are common to both data sets. With the International HapMap Project genotype data (phase 3 release #3, National Center for Biotechnology Information build 36, dbSNP b126, 2010-05-28, minor allele frequency > 5%) as references for CEU (population with ancestry from northern and western Europe), JPT/CHB (Japanese/Chinese) and YRI (Yoruba in Ibadan, Nigeria) (9), we seeded 10,155 high-quality markers (r² < 0.004) in STRUCTURE (10) and identified 4,137 individuals (2,028 cases and 2,109 controls) as the study subjects (0.75–1.00 similarity to CEU) in current analysis. Then the top five principal components for population stratification were derived from Caucasian subjects using the EIGENSTRAT (11).

**Definition of pathways and genes**

The pathways analyzed in this study were either defined by KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology) databases or manually created according to the information in the literature. As described in our previous study (12), we identified 197 pathways each with >500 genes from KEGG database (13). Prior to G × E analysis, we performed principal component analysis (PCA) to refine the contributing genes in each pathway. The genetic variation in each gene was decomposed into orthogonal components (eigenSNPs) through PCA. We tested association of each gene within a pathway with disease in likelihood ratio test (LRT) nested in logistic regression model based on principal components (eigenSNPs) accounting for at least 85% variation of a gene (12,14), and only those genes with marginal association with disease (P ≤ 0.10) were retained in the pathway for interaction analysis (PCA–LRT). This approach is in line with the two-step gene-environment interaction analysis approach by Kooperberg et al. (5). Finally, a total of 172 KEGG pathways each having ≥2 genes surviving PCA–LRT screening (Supplementary Table 1, available at Carcinogenesis Online) remained for pathway–smoking interaction analysis. From GO database (16), we retrieved three nicotine dependence pathways, i.e. ‘response to nicotine’ (GO:0035094), ‘cellular response to nicotine’ (GO:0071316) and ‘behavioral response to nicotine’ (GO:0053095; Supplementary Table 1, available at Carcinogenesis Online). Furthermore, we curated three additional gene sets using the Phenotype-Genotype Integrator (PhenGenI; http://www.ncbi.nlm.nih.gov/gap/phengi) based on GWAS top hits (P < 1 x 10⁻⁴) for smoking-related cancers: lung cancer, head and neck cancer and bladder cancer (Supplementary Table 1, available at Carcinogenesis Online). Due to relatively small pathway size, the PCA–LRT screening was not applied for the three GO pathways and the self-curated gene sets.

We retrieved 19,058 genes from the human genome database version 18 (hg18) using the University of California at Santa Cruz Genome Browser data retrieval tool (17). We further extended a gene region to include SNPs within 20kb upstream or downstream. There were 17,912 genes each genotyped for ≥2 SNPs in current GWAS data set, resulting in a total of 468,114 SNPs for the G × E analysis.

**Exposure variables**

Exposure information without personal identifiers was provided by each collaborating institution to MD Anderson under Institutional Review Board approvals and Material Transfer Agreement (MTA) agreements. Exposure variables included sex, age, race/ethnicity, history of cigarette smoking, pack-years, adulthood body mass index (BMI), history of diabetes and family history of cancer. All variables were coded following the same data coding dictionary. Missing values for pack-years in 228 smokers were imputed using the mean values of study-age-sex specific pack-years (18). In this G × E analysis, the exposure variable was defined as 0, <20 and >20 pack-years. Other exposure variables that are adjusted in the multivariable models included sex, age (continuous), BMI (categorical ≤25, 25–29.9 and ≥30 kg/m²) and diabetes (yes versus no). Due to a large number of missing values, family history of cancer was not adjusted in the model and the former and current smoking status was not examined in this analysis.

**Statistical methods**

A LRT nested in logistic regression model was applied throughout all G × E analyses. The full model included sex, age (continuous), study sites (categorical), five principal components (quantitative) capturing population stratification, diabetes (yes versus no), BMI (categorical), genetic factors (eigenSNPs), smoking (0, <20 versus >20 pack-years) and the interaction terms (the products of smoking and genetic factors). The reduced model was the same as the full model except excluding the interaction term(s). Varied smoking intensity may display different G × E spectrum (19). Balancing statistical power and exposure level, as an exploratory effort, we also examined the gene–smoking interactions using 30 pack-years as the cutoff (0, <30 versus >30 pack-years) or using pack-years as a continuous variable at pathway level.

We conducted G × E analysis at the pathway and gene level based on the eigenSNPs accounting for at least 85% genetic variation of a pathway or a gene. We identified contributing genes to a pathway or contributing SNPs to a gene using the criterion of P ≤ 0.05 at the gene or SNP level. SNPs were coded as 0, 1 or 2 based on the number of the minor allele. We further explored the marginal effects of genes or SNPs with nominal significance in a subgroup analysis stratified by 20 pack-years (≤20, including non-smokers), >20. In addition, we analyzed the genes with nominal interactions with smoking using the ingenuity pathway analysis (IPA) (Ingenuity® Systems www.ingenuity.com). The null hypothesis to be tested in the IPA is that nominally significant interacting genes are equally represented in a given pathway, compared with the set of all other genes not in the pathway, which can be tested using Fisher’s exact test based on the hypergeometric distribution. This type of pathway analysis belongs to the category of ‘competitive’ pathway-based tests; in contrast, the LRT-based pathway interaction test belongs to the category of ‘self-contained’ tests, whose null hypothesis is that none of the genes/ SNPs in a given biological pathway interacts with smoking on the disease risk (20). As the IPA and LRT pathway-based methods test different null hypotheses, they may not identify the same pathways and we employed both in our analysis to complement each other.

To control the false-positive findings incurred by multiple testing, we applied both the Bonferroni correction and q value method with false discovery rate (FDR) at 0.10 for G × E analysis at pathway/gene/SNP level (21). P values ≤2.81 x 10⁻⁴ (0.05/178), 2.79 x 10⁻⁴ (0.05/17912) and 1.07 x 10⁻⁷ (0.05/468,114) after Bonferroni correction were considered statistically significant for G × E at pathway, gene and SNP levels, respectively. q values < 0.10 were considered statistically significant at FDR = 10%. IPA automatically provides Benjamini–Hochberg (B-H) adjusted P values for overrepresented pathways, and B-H adjusted P values < 0.10 were considered statistically significant at FDR = 10%.

**Results**

The demographics and exposure variables for the study population is described in Table I. The distributions of age, race and sex across case and control groups were balanced (all P > 0.10). Self-reported non-Hispanic whites made up >99% of the study population. Case–control association plot did not imply presence of population stratification (genomic control lambda = 0.99) (22). Smoking, obesity and diabetes were positively associated with increased risk of pancreatic cancer after adjusting for other factors (Ps < 0.001). A clear dose–response relationship was observed for pack-years smoked and risk of pancreatic cancer.

G X E interactions at pathway level

Ten out of 172 KEGG pathways but none of the six candidate pathways showed nominal interactions with smoking (P < 0.05). After adjustment for multiple testing, none of them reached the significance level (Table II). The top two pathways were pancreatic secretion (P = 0.0054) and salivary secretion (P = 0.012) pathways, with major contributing genes RAB5A, PLCB and CTRB1 (Table II). When
The results remained the same when the IPA analysis was restricted to genes with $P$ values of ≤0.04 or ≤0.03 (data not shown). Furthermore, removing 13 smoking- or nicotine-associated genes defined in previous studies (23–25) and three nicotine-response GO pathways (16) from the gene list did not significantly change the results of IPA analysis, excluding the possibility that the significance was inflated by these genes. Notably, one of the three nicotine dependency pathways ‘behavioral response to nicotine’ (GO:0035095) showed a nominal association with heavy smoking (30 or more pack-years) while the possibility that the significance was inflated by these genes. ICP analysis identified two significant pathways: axonal guidance signaling ($P = 2.12 \times 10^{-7}$) and $\alpha$-adrenergic signaling pathway ($P = 2.52 \times 10^{-5}$) (Table III). The B-H adjusted $P$ values were, respectively, $8.78 \times 10^{-5}$ and $5.20 \times 10^{-5}$. The results remained the same when the IPA analysis was restricted to genes with $P$ values of ≤0.04 or ≤0.03 (data not shown). Furthermore, removing 13 smoking- or nicotine-associated genes defined in previous studies (23–25) and three nicotine-response GO pathways (16) from the gene list did not significantly change the results of IPA analysis, excluding the possibility that the significance was inflated by these genes. Notably, one of the three nicotine dependency pathways ‘behavioral response to nicotine’ (GO:0035095) showed a nominal association with heavy smoking (30 or more pack-years) ($P = 0.046$) in combined data set of cases and controls.

As sensitivity analysis, we conducted G × E analysis at pathway level based on the data set with complete pack-years only (without imputed pack-years). The LRT results were almost identical to those obtained from analysis including the imputed pack-years.
We also analyzed marginal effects for the axon guidance (Supplementary Table 1 available at Carcinogenesis Online). Subgroup analysis by smoking status [<20 (including non-smokers) versus ≥20 pack-years] found that some genes had differential marginal associations (P ≤ 0.05) with risk of pancreatic cancer. For example, ZBED2 and CLSTN1 had P_{G × E} ≤ 0.0005 in those with ≥20 pack-years of smoking but P > 0.1 in those with <20 pack-years. In contrast, ATXN2 had a P_{G × E} value of 0.0029 and 0.21 among those with <20 or ≥20 pack-years of smoking, respectively (Table IV). We also analyzed marginal effects for the axon guidance genes by smoking status (≥20 versus <20 pack-years). Eleven genes (PLCB1, GNG12, WNT10A, PLXNA2, PRKCCZ, ITSN1, PIK3CD, ADAMTS6, ROBO1, KLC1 and BMP5) in individuals with ≥20 pack-years, four genes (Ngef, WNT4, BCAR1 and PRKAR2B) in those with <20 pack-years, and one gene (ADAM9) in both subgroups showed nominal effect on risk of pancreatic cancer (P < 0.05; Supplementary Table 3, available at Carcinogenesis Online). These observations are in line with the interaction of axon guidance pathway and smoking in modifying the risk of pancreatic cancer.

### Discussion

This study for the first time comprehensively explored gene–smoking interactions in modifying the risk of pancreatic cancer at the pathway, gene and SNP levels using GWAS data. We identified significant interactions of the axonal guidance and α-adrenergic signaling pathways with smoking in modifying the risk of pancreatic cancer by IPA analysis. We also observed a possible interaction of pancreatic secretion pathway with smoking using LRT test. We did not find supporting evidences that nicotine dependence genes/pathways or genes identified in other smoking-related cancers modify the risk of pancreatic cancer. These preliminary findings, if confirmed, may reveal novel molecular mechanisms underlying the development of smoking-related pancreatic cancer.

#### Table III. Top overrepresented canonical pathways interacting with smoking

<table>
<thead>
<tr>
<th>Biological process</th>
<th>P valuea</th>
<th>Ratiob</th>
<th>Contributing genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axonal guidance signaling</td>
<td>2.12 × 10^{-7}</td>
<td>44/469 (0.094)</td>
<td>KLC1 ITSN1 SLT1 MAPK3 BMP3 SEMA6B KRAS EPHA4 PLXNA2 ROBO1 BCAR1 PRKCCZ EPHA8 LNPEP TUBA8 SUFU ABLIM3 ADAM19 FIGF PLCB1 WNT4 ARPC1A CHMP1A GNG12 PRKD1 WNT5B NGEF PAK6 BMP5 PDGFBB GNG10 EFN1 A5GAP3 ADAMTS6 PRKAR2B WNT10A WAS RTN4 ADAM10 PRKAG2 BMP7 PIK3CD ADAM9 OPN1SW ADCY2 CAMK4 MAPK3 ADCY3 KRAS PRKCCZ GNG10 CALM1 PRKAR2B PRKAG2 ADCY7 PRKD1 GNG12 OPN1SW</td>
</tr>
<tr>
<td>α-Adrenergic signaling</td>
<td>2.52 × 10^{-5}</td>
<td>14/105 (0.13)</td>
<td></td>
</tr>
</tbody>
</table>

#### Table IV. Top genes interacting with smoking at P < 0.0005

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene description</th>
<th>No. of SNPs/ eigenSNPs</th>
<th>P_{G × E}</th>
<th>P value for marginal association</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZBED2</td>
<td>Zinc finger BED domain-containing protein 2</td>
<td>4/3</td>
<td>0.00013</td>
<td>0.00053</td>
</tr>
<tr>
<td>EKO1</td>
<td>Exonuclease 1</td>
<td>19/7</td>
<td>0.00017</td>
<td>0.17</td>
</tr>
<tr>
<td>PSG2</td>
<td>Pregnancy-specific beta-1-glycoprotein 2</td>
<td>3/3</td>
<td>0.00019</td>
<td>0.035</td>
</tr>
<tr>
<td>SLC36A1</td>
<td>Proton-coupled amino acid transporter 1</td>
<td>28/5</td>
<td>0.00022</td>
<td>0.089</td>
</tr>
<tr>
<td>CLSTN1</td>
<td>Calcyntenin-1</td>
<td>10/4</td>
<td>0.00023</td>
<td>0.00088</td>
</tr>
<tr>
<td>MTHFSD</td>
<td>Methyleneterahydrofolate synthase domain-containing protein</td>
<td>24/7</td>
<td>0.00036</td>
<td>0.03</td>
</tr>
<tr>
<td>FAT2</td>
<td>Proteodrherin Fat 2</td>
<td>35/8</td>
<td>0.00038</td>
<td>0.047</td>
</tr>
<tr>
<td>IIL10RB</td>
<td>Interleukin-10 receptor subunit beta</td>
<td>17/7</td>
<td>0.00043</td>
<td>0.016</td>
</tr>
<tr>
<td>ATTXN2</td>
<td>Ataxin-2</td>
<td>7/2</td>
<td>0.00047</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Note: Table IV includes a number of SNPs in the gene/number of principal components used for the analysis.

Obtained from LRT in logistic regression with adjustment for age, sex, study, diabetes, and five principal components for population substructure.

Including non-smokers.
Our observation on the axon guidance genes in smoking-related pancreatic cancer is novel and interesting. Axon guidance pathway is largely involved in neuronal extension and location during embryogenesis. Recently, there is increasing evidence supporting a role of axon guidance genes in cell proliferation, migration, adhesion, invasiveness, apoptosis, survival, metastasis and angiogenesis in various cancers including pancreatic cancer (26–28). In fact, a recently reported whole exome sequencing analysis of human pancreatic adenocarcinoma has found frequent alterations of the axon guidance genes including the SLIT/ROBO signaling genes (29). Careful examination of the reported data (Fig. 1 of ref. 29) revealed a higher frequency of the axon guidance gene alterations in neversmokers (26/49, 53%) than in smokers (14/46, 29%). Findings from other studies seem to suggest that these observations may not be simply made by chance. For example, SEMA5A, one of axon guidance genes, has also been identified as a novel biomarker for lung cancer in non-smokers (30). Although polymorphic variants of the axon guidance genes have been shown to influence the success of cigarette smoking cessation (31,32), we did not find any evidence that the nicotine addiction genes were involved in pancreatic cancer. Apparently, both the mutation data and GWAS data showing a possible interaction of smoking with axon guidance genes in cell proliferation, migration, adhesion, invasiveness, apoptosis, survival, metastasis and angiogenesis in various cancers including pancreatic cancer (26–28).

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Gene</th>
<th>Genotype</th>
<th>Odds ratio (95% CI)</th>
<th>Odds ratio (95% CI)</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1383180</td>
<td>EVC</td>
<td>CC</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TT</td>
<td>1.00 (0.94–1.51)</td>
<td>0.87 (0.75–1.03)</td>
<td>0.98 (0.86–1.12)</td>
</tr>
<tr>
<td>rs11248542</td>
<td>Inter-genic</td>
<td>CT</td>
<td>1.00 (1.20–2.38)</td>
<td>0.66 (0.52–0.83)</td>
<td>0.90 (0.74–1.09)</td>
</tr>
<tr>
<td>rs11195244</td>
<td>Inter-genic</td>
<td>CC</td>
<td>6/16 (0.44–1.17)</td>
<td>1.17 (0.96–1.43)</td>
<td>0.64 (0.28–1.45)</td>
</tr>
<tr>
<td>rs2043385</td>
<td>KCNP4</td>
<td>CC</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>rs7141385</td>
<td>Inter-genic</td>
<td>CT</td>
<td>619/751 (1.03–1.21)</td>
<td>0.94 (0.83–1.07)</td>
<td>0.000013</td>
</tr>
<tr>
<td>rs17256483</td>
<td>Inter-genic</td>
<td>AA</td>
<td>615/638 (1.35–1.78)</td>
<td>0.94 (0.75–1.17)</td>
<td>0.000014</td>
</tr>
<tr>
<td>rs2470697</td>
<td>Inter-genic</td>
<td>TT</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>rs1383180</td>
<td>EVC</td>
<td>TT</td>
<td>1.00 (0.94–1.51)</td>
<td>0.87 (0.75–1.03)</td>
<td>0.98 (0.86–1.12)</td>
</tr>
<tr>
<td>rs11248542</td>
<td>Inter-genic</td>
<td>CT</td>
<td>1.00 (1.20–2.38)</td>
<td>0.66 (0.52–0.83)</td>
<td>0.90 (0.74–1.09)</td>
</tr>
<tr>
<td>rs11195244</td>
<td>Inter-genic</td>
<td>CC</td>
<td>6/16 (0.44–1.17)</td>
<td>1.17 (0.96–1.43)</td>
<td>0.64 (0.28–1.45)</td>
</tr>
<tr>
<td>rs2043385</td>
<td>KCNP4</td>
<td>CC</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>rs7141385</td>
<td>Inter-genic</td>
<td>CT</td>
<td>619/751 (1.03–1.21)</td>
<td>0.94 (0.83–1.07)</td>
<td>0.000013</td>
</tr>
<tr>
<td>rs17256483</td>
<td>Inter-genic</td>
<td>AA</td>
<td>615/638 (1.35–1.78)</td>
<td>0.94 (0.75–1.17)</td>
<td>0.000014</td>
</tr>
<tr>
<td>rs2470697</td>
<td>Inter-genic</td>
<td>TT</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>rs1383180</td>
<td>EVC</td>
<td>TT</td>
<td>1.00 (0.94–1.51)</td>
<td>0.87 (0.75–1.03)</td>
<td>0.98 (0.86–1.12)</td>
</tr>
<tr>
<td>rs11248542</td>
<td>Inter-genic</td>
<td>CT</td>
<td>1.00 (1.20–2.38)</td>
<td>0.66 (0.52–0.83)</td>
<td>0.90 (0.74–1.09)</td>
</tr>
<tr>
<td>rs11195244</td>
<td>Inter-genic</td>
<td>CC</td>
<td>6/16 (0.44–1.17)</td>
<td>1.17 (0.96–1.43)</td>
<td>0.64 (0.28–1.45)</td>
</tr>
<tr>
<td>rs2043385</td>
<td>KCNP4</td>
<td>CC</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>rs7141385</td>
<td>Inter-genic</td>
<td>CT</td>
<td>619/751 (1.03–1.21)</td>
<td>0.94 (0.83–1.07)</td>
<td>0.000013</td>
</tr>
<tr>
<td>rs17256483</td>
<td>Inter-genic</td>
<td>AA</td>
<td>615/638 (1.35–1.78)</td>
<td>0.94 (0.75–1.17)</td>
<td>0.000014</td>
</tr>
<tr>
<td>rs2470697</td>
<td>Inter-genic</td>
<td>TT</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table V. Top SNPs interacting with smoking at P < 0.00003

*Including non-smokers.
*Odds ratio (95% CI) adjusted for age, sex, and five principal components for population substructure.
*In addition to the above covariates, BMI (categorical), diabetes and study was further adjusted.
*P value for interaction term obtained by LRT in logistic regression analysis of all study subjects.

Our observation on the axon guidance genes in smoking-related pancreatic cancer is novel and interesting. Axon guidance pathway is largely involved in neuronal extension and location during embryogenesis. Recently, there is increasing evidence supporting a role of axon guidance genes in cell proliferation, migration, adhesion, invasiveness, apoptosis, survival, metastasis and angiogenesis in various cancers including pancreatic cancer (26–28). In fact, a recently reported whole exome sequencing analysis of human pancreatic adenocarcinoma has found frequent alterations of the axon guidance genes including the SLIT/ROBO signaling genes (29). Careful examination of the reported data (Fig. 1 of ref. 29) revealed a higher frequency of the axon guidance gene alterations in neversmokers (26/49, 53%) than in smokers (14/46, 29%). Findings from other studies seem to suggest that these observations may not be simply made by chance. For example, SEMA5A, one of axon guidance genes, has also been identified as a novel biomarker for lung cancer in non-smokers (30). Although polymorphic variants of the axon guidance genes have been shown to influence the success of cigarette smoking cessation (31,32), we did not find any evidence that the nicotine addiction genes were involved in pancreatic cancer. Apparently, both the mutation data and GWAS data showing a possible interaction of smoking with axon guidance genes in pancreatic cancer need to be confirmed in future studies. Illustrating the biological mechanisms underlying these associations may shed new lights on the molecular mechanisms of pancreatic cancer in non-smokers.

One of the major components for cigarettes smoking is nicotine, which by itself does not cause cancer, but its metabolites, i.e. tobacco specific nitrosamines (NNK) are known pancreatic carcinogens (33). NNK not only covalently bind to DNA and contribute to K-ras gene mutations but also activate protumorigenic signaling pathways. For example, NNK interact with β-adrenergic receptors (β-ARs) to stimulate COX-2-mediated inflammatory response and activate epidermal growth factor receptor and its downstream Raf/mitogen-activated protein (MAP) kinase (MEK)/extracellular signal-regulated kinase pathway (34). Although the experimental evidence linking to β-ARs is missing, α-ARs are essential to activation of mitogen-activated protein kinase in vitro and α2-AR antagonists can completely suppress mitogen-activated protein kinase activity (35,36). In addition, α2-ARs may trans-activate PI3K and Akt pathways, which are very important for cell proliferation and survival (37).

The interaction of smoking with pancreatic secretion pathway was nominal but was reproducible. It remained as the top pathway when different smoking variables were used in the interaction analyses. The major contributing genes to this pathway include RAB8A, PLCB1 and CTRB1. The protein encoded by RAB8A gene is a member of the RAS superfamily, which may play a role in the transport of proteins from the endoplasmic reticulum to the Golgi and the plasma membrane. PLCB1 (phospholipase C, beta 1) was a major contributing gene to seven of the top ten pathways nominally interacting with smoking (Table II). The enzyme encoded by this gene catalyzes the formation of inositol 1,4,5-trisphosphate and diacylglycerol from phosphatidylinositol 4,5-bisphosphate. This reaction plays an important role in the intracellular transduction of many extracellular signals. CTRB1 (chymotrypsinogen) is a serine protease that is secreted into the gastrointestinal tract as an inactive precursor, which is activated by proteolytic cleavage with trypsin. Although we do not know how these genes contribute to smoking-related pancreatic carcinogenesis, the impact of cigarette smoking on exocrine and endocrine functions of the pancreas has been well documented (38). Long-term treatment with nicotine reduces pancreatic secretion (39) and increases acinar cell proliferation (40). Upon secretin stimulation, serum levels of secreted pancreatic proteins were significantly elevated in smokers but not in non-smokers (41). Among chronic pancreatitis patients, smokers had significantly lower levels of insulin and glucagon in pancreas than that of non-smokers (42). Further study on the role of pancreatic secretion pathway genes in smoking-induced pancreatic tumor models is needed.

Five intergenic region SNPs and two SNPs of the EVC (Ellis van Creveld syndrome) and KCNIP4 (Kv channel interacting protein 4) genes showed interactions with smoking at P < 0.00003. At gene level, EVC (P = 0.0034) but not KCNIP4 (P = 0.22) showed nominal interaction with smoking. Neither of the two genes has been linked to pancreatic secretion, axon guidance signaling, alpha-adrenergic signaling pathways or smoking-related disease. Whether the observed associations are due to chance alone or represents unknown...
mechanisms underlying smoking-induced pancreatic cancer remains to be investigated.

In this study, we employed both the Bonferroni procedure to control the family-wise error rate and q value/B-H methods to control the FDR. The family-wise error rate is the probability of making at least one false positive in all comparisons, whereas the FDR is the expected proportion of false positives among all tests determined to be significant. For large-scale genetic and genomic testing problems, controlling FDR is less conservative than controlling the family-wise error rate, i.e. the FDR method often leads to more true discoveries while allowing some false positives (21). In the current analysis, we performed interaction tests at the SNP, gene and pathway levels. We expected that the FDR method might lead to more true discoveries at the SNP and gene levels as there were, respectively, 468 114 and 17 912 interaction tests. It, however, turned out that no matter we employed the Bonferroni procedure or the q value/B-H-based FDR methods, the conclusions were the same for all levels of analyses. We provided both P values and FDR q values for pathway/gene/SNP interactions in Supplementary Tables 1 and 4, available at Carcinogenesis Online, to render the readers the freedom to choose between the two false-positive control methods.

The requirement of large statistical power remains a daunting challenge for G × E analysis of the GWAS data. Previous studies have shown that G × E analyses restricted to genes with marginal effect may increase the statistical power (15). However, we observed that some SNPs conferred differential effects between ever and never smokers but exhibited no marginal effect when smoker and non-smokers were combined. A similar scenario was observed in recent studies on the gene–alcohol interactions in esophageal squamous-cell carcinoma (43,44). Thereby, G × E analysis on genes with marginal effects only may miss those without main effects but truly interacting with an exposure. Therefore, we suggest that G × E analysis should make use of combined methodologies with complementary strengths, as used here and suggested by other investigators (45), to discover the missing heritability of pancreatic cancer (46).

This study has its strengths and limitations. This is by far the largest G × E analysis of all biological pathways defined by KEGG and IPA in pancreatic cancer employing an agnostic approach. We applied PCA approach to lower the magnitude of the GWAS data, increasing the sensitivity of finding true signals. Quality control was permeated into each step of genotyping and exposure measurement and data collection. Stringent statistical threshold was applied to reduce false-positive discovery. Nevertheless, relatively small sample size curbed the G × E GWAS analysis. Our analysis was based on pack-years rather than former and current smokers, which may limit generalization of our observations. Misclassification of pack-years due to imputation may impact our results. Unavailability of external data sets limited validation of our findings and the generalization of the results. Despite of those, the pathways identified in this study are highly relevant to pancreatic cancer and are supported by other studies. G × E analysis in GWAS provides us with an unprecedented opportunity to discover genetic factors bridging smoking and pancreatic cancer.

Supplementary material

Supplementary Tables 1–4 can be found at http://carcin.oxfordjournals.org/.

Funding

National Institutes of Health (RO1 CA98380-05 to D.L.); MD Anderson Cancer Center Support Grant CA016672; R01 grants (CA169122 and HL116720 to P.W.); PO1 grant (5P30CA023108-28) and U01 grant (7U19CA148127-03 to C.A.); Sheikh Ahmed Center for Pancreatic Cancer Research at The University of Texas MD Anderson Cancer Center Funds (to D.L.).

Conflict of Interest Statement: None declared.

References

G × E analysis of GWAS in pancreatic cancer


Received November 1, 2013; revised December 30, 2013; accepted January 4, 2014