Genes–Environment Interactions in Obesity- and Diabetes-Associated Pancreatic Cancer: A GWAS Data Analysis

Hongwei Tang\textsuperscript{1}, Peng Wei\textsuperscript{3}, Eric J. Duell\textsuperscript{4}, Harvey A. Risch\textsuperscript{5}, Sara H. Olson\textsuperscript{6}, H. Bas Bueno-de-Mesquita\textsuperscript{7}, Steven Gallinger\textsuperscript{8}, Elizabeth A. Holly\textsuperscript{9,10}, Gloria M. Petersen\textsuperscript{11}, Paige M. Bracci\textsuperscript{10}, Robert R. McWilliams\textsuperscript{11}, Mazda Jenab\textsuperscript{12}, Elio Riboli\textsuperscript{16}, Anne Tjønneland\textsuperscript{18}, Marie Christine Boutron-Ruault\textsuperscript{13,14,15}, Rudolf Kaaks\textsuperscript{19}, Dimitrios Trichopoulou\textsuperscript{20,21,22}, Salvatore Panico\textsuperscript{23}, Malin Sund\textsuperscript{24}, Petra H.M. Peeters\textsuperscript{8,16}, Kay-Tee Khaw\textsuperscript{17}, Christopher I. Amos\textsuperscript{2}, and Donghui Li\textsuperscript{1}

Abstract

\textbf{Background:} Obesity and diabetes are potentially alterable risk factors for pancreatic cancer. Genetic factors that modify the associations of obesity and diabetes with pancreatic cancer have previously not been examined at the genome-wide level.

\textbf{Methods:} Using genome-wide association studies (GWAS) genotype and risk factor data from the Pancreatic Cancer Case Control Consortium, we conducted a discovery study of 2,028 cases and 2,109 controls to examine gene–obesity and gene–diabetes interactions in relation to pancreatic cancer risk by using the likelihood-ratio test nested in logistic regression models and Ingenuity Pathway Analysis (IPA).

\textbf{Results:} After adjusting for multiple comparisons, a significant interaction of the chemokine signaling pathway with obesity (\(P = 3.29 \times 10^{-8}\)) and a near significant interaction of calcium signaling pathway with diabetes (\(P = 1.57 \times 10^{-5}\)) in modifying the risk of pancreatic cancer were observed. These findings were supported by results from IPA analysis of the top genes with nominal interactions. The major contributing genes to the two top pathways include \textit{NGT2}, \textit{RELA}, \textit{TIAM1}, and \textit{GNAS}. None of the individual genes or single-nucleotide polymorphism (SNP) except one SNP remained significant after adjusting for multiple testing. Notably, SNP rs10818684 of the \textit{PTGS1} gene showed an interaction with diabetes (\(P = 7.91 \times 10^{-7}\)) at a false discovery rate of 6%.

\textbf{Conclusions:} Genetic variations in inflammatory response and insulin resistance may affect the risk of obesity- and diabetes-related pancreatic cancer. These observations should be replicated in additional large datasets.

\textbf{Impact:} A gene–environment interaction analysis may provide new insights into the genetic susceptibility and molecular mechanisms of obesity- and diabetes-related pancreatic cancer. \textit{Cancer Epidemiol Biomarkers Prev}; 23(1); 98–106. © 2013 AACR.

\textbf{Note:} Supplementary data for this article are available at Cancer Epidemiol Biomarkers Prevention Online (http://cebp.aacrjournals.org/).

Current address for Christopher I. Amos: Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire.

\textbf{Corresponding Author:} Donghui Li, Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 426, Houston, TX 77030. Phone: 713-834-6690; Fax: 713-834-6153; E-mail: dli@mdanderson.org

doi: 10.1158/1055-9965.EPI-13-0437-T

© 2013 American Association for Cancer Research.

Authors' Affiliations: Departments of 1Gastrointestinal Medical Oncology and 1Epidemiology, The University of Texas MD Anderson Cancer Center; 2Division of Biostatistics and Human Genetics Center, School of Public Health, University of Texas Health Science Center, Houston, Texas; 3Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain; 4Yale University School of Public Health, New Haven, Connecticut; 5Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York; 6National Institute for Public Health and the Environment (RIVM), Bilthoven and Department of Gastroenterology and Hepatology, University Medical Center Utrecht, Utrecht; 7Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands; 8Samuel Lunenfeld Research Institute, Toronto General Hospital, University of Toronto, Toronto, Canada; 9Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA; 10Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York; 11National Institute for Public Health and the Environment (RIVM), Bilthoven and Department of Gastroenterology and Hepatology, University Medical Center Utrecht, Utrecht, the Netherlands; 12International Agency for Research on Cancer, Lyon; 13Institut national de la sante et de la recherche medicale (INSERM), Centre for research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women’s Health team; 14Univ. Paris Sud, UMRs 1018; 15IGR, F-94805, Villejuif, France; 16Division of Epidemiology, Public Health, and Primary Care, Imperial College London, London; 17School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom; 18Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark; 19Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; 20Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts; 21Bureau of Epidemiologic Research, Academy of Athens; 22Hellenic Health Foundation, Athens, Greece; 23Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy; and 24Department of Surgery, Umea University Hospital, Umea, Sweden

Published OnlineFirst October 17, 2013; DOI: 10.1158/1055-9965.EPI-13-0437-T

Downloaded from cebp.aacrjournals.org on October 29, 2017. © 2014 American Association for Cancer Research.
Introduction

Pancreatic cancer is the fourth leading cause of cancer-related deaths, accounting for more than 37,600 deaths each year in the United States (1). Epidemiologic studies have identified cigarette smoking as the major modifiable risk factor for this disease. Obesity and long-term history of diabetes mellitus may also affect risk and are also modifiable (2, 3). Genetic factors are known to play a role in pancreatic cancer development. Although genome-wide association studies (GWAS) have identified a few loci and chromosome regions that are significantly associated with the risk of pancreatic cancer (4, 5), these findings explain only a portion of the heritability of this disease. Because of the limitations of single-marker analysis on GWAS data, there have been increasing efforts recently on GWAS pathway analysis, which uses prior biologic knowledge of gene function and aims at combining moderate signals of single-nucleotide polymorphism (SNP) and obtaining biologically interpretable findings (6, 7). Despite its great promise in providing insights into disease mechanisms, current GWAS pathway analysis has some caveats including being limited to enrichment of marginal genetic effects in biologic pathways without considering possible interactions between pathways and environmental factors (8). On the other hand, environmental factors are likely to interact with multiple genes through various biologic pathways, contributing to the susceptibility of complex human diseases. Although current GWAS top hits account for only limited heritability, gene–environment interactions may account for some of the missing heritability of pancreatic cancer (9).

We have previously conducted pathway analyses of the GWAS data in pancreatic cancer. Several novel pathways significantly associated with risk were identified (10, 11). For example, the pancreas development pathway [Mature Onset Diabetes of the Young (MODY) pathway] was identified as a top pathway in pancreatic cancer etiology. One possible mechanism related to this pathway is through obesity and diabetes (12–14). However, our previous study did not detect a significant interaction of obesity or diabetes with the NR5A2 gene, a GWAS top hit and a major contributing gene to the pancreas development pathway (MODY pathway), in modifying the risk of pancreatic cancer. On the other hand, we detected a strong interaction of the fat mass and obesity-associated (FTO) gene with overweight for pancreatic cancer risk, even though the gene did not show a marginal effect (15). A recent post-GWAS analysis of diabetes-related genes also failed to find strong evidence that common variants underlying type 2 diabetes or related phenotypes interact with diabetes in modifying the risk of pancreatic cancer (16). These observations suggest that there are unidentified genes contributing to obesity- and diabetes-related pancreatic cancer. Taking advantage of the existing GWAS data and exposure variables from the Pancreatic Cancer Case Control Consortium (PanC4; ref. 17), we conducted a comprehensive gene–environment (G × E) interaction analysis of genetic factors that may modify the associations of obesity and diabetes with pancreatic cancer.

Materials and Methods

Study population and dataset

The study population was drawn from seven studies participating in the previously conducted GWAS of the Pancreatic Cancer Cohort Consortium (PanScan) and the PanC4 Consortium (4, 5), including six case–control studies conducted at the MD Anderson Cancer Center (Houston, TX), Mayo Clinic (Rochester, MN), Yale University (New Haven, CT), University of California (San Francisco, CA), Memorial Sloan-Kettering Cancer Center (New York, NY), and University of Toronto (Toronto, ON, Canada), and one nested case–control study from the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. Cases were defined as patients diagnosed with primary adenocarcinoma of the pancreas; in each study center, controls were matched to cases according to birth year, sex, and self-reported race/ethnicity and were free of pancreatic cancer at the time of recruitment.

GWAS scanning was performed at the National Cancer Institute Core Genotyping Facility using the Illumina HumanHap550 and HumanHap550-Duo SNP arrays and Illumina Human 610-Quad arrays (4, 5). Genotype data covering 562,000 and 621,000 SNPs from 4,195 study subjects (2,163 cases and 2,232 controls) were downloaded from the Database of Genotypes and Phenotypes website (http://www.ncbi.nlm.nih.gov/gap) with the approval of the MD Anderson Institutional Review Board (IRB). We first conducted data cleaning and quality control by removing SNPs with minor allele frequency (MAF) less than 5%, deviating from the Hardy–Weinberg equilibrium (P < 0.001) or not genotyped in both SNP array platforms, resulting in a final dataset of 468,114 SNPs. According to the International HapMap Project genotype data (phase III release #3, NCBI build 36, dbSNP b126, 2010-05-28, MAF > 5%) for CEU (European), JPT (Japanese)/CHB (Chinese), and YRI (Yoruba) (18), we used 10,155 high-quality markers (r² < 0.004) in STRUCTURE (19) and identified a total of 4,137 individuals (2,028 cases and 2,109 controls) with 0.75–1.00 similarity to CEU. Then, we derived 2,109 controls) with 0.75–1.00 similarity to CEU as the study subjects in the current analysis. Then, we derived a final dataset of 468,114 SNPs. Accord-
were included. In total, 5,127 genes annotated in the 197 pathways, covering 82,881 SNPs, were tested for interactions with risk factors.

**Exposure variables**

Exposure variables without personal identifiers were provided by each participating institution to MD Anderson under IRB approvals and MTA (Material Transfer Agreement). Exposure variables included age, sex, race/ethnicity, adulthood body mass index (BMI, weight/height²), history of cigarette smoking, history of diabetes, and family history of cancer. All data were coded according to a uniform data dictionary. Missing pack-years of smoking were imputed based on study–age–sex means in 228 smokers. After merging and cleaning the data, we defined the variables in this G × E analysis as follows: obesity (BMI ≤30 kg/m² vs. >30 kg/m²) and diabetes (yes vs. no). Other exposure variables that are adjusted in the multivariable models included age (continuous), sex, and smoking (0, <20, and ≥20 pack-years). Because of a large number of missing values for family history of cancer, this variable was not considered in the model.

**Statistical analysis**

We used principal component analysis (PCA) to reduce the dimension of SNPs within a gene or pathway before the interaction analysis (11, 23). Briefly, PCA was performed to decompose the genetic variation in a gene into orthogonal components called eigenSNPs; the eigenvalues were calculated to identify principal components (eigenSNPs) that explained at least 85% of the observed genetic variation within a gene. Before pathway-by-environment interaction analyses, we used the global likelihood-ratio test (LRT) to determine if genes represented by the eigenSNPs were marginally associated with disease status, and only those genes with nominal P values ≤0.10 were retained in the pathway (PCA-LRT) screening. The eigenSNPs of genes with marginal effects were included in the pathway–environment interaction analyses, along the same line as the two-step approach for SNP × SNP/SNP × environment interaction analysis proposed in refs. (24) and (25). The gene/pathway and environment interaction was analyzed using LRT in nested logistic regression models. The full model included age (continuous), sex, study sites (categorical), five principal components (quantitative) capturing population structure, smoking (pack-years), genetic factors (eigenSNPs), the risk factor of interest, and the interaction terms (the products of risk factor of interest and eigenSNPs). The interaction terms were removed from the reduced model.

For G × E analysis at the pathway level, in total 172 pathways having at least two genes with marginal effect were identified through the PCA-LRT screening (Supplementary Table S1). Genes with a $P_{G \times E}$ value <0.05 in the interaction analysis were considered as the major contributing genes to the pathway. We also performed a simulation study to demonstrate that the LRT method can effectively control the Type I error for the interaction analysis (Supplementary Text).

For G × E analysis at the gene level, a total of 5,127 genes were tested using LRT and logistic regression. SNPs with a $P_{G \times E}$ value of $<0.05$ were defined as the contributing SNPs to a gene. After screening all 5,127 genes, we also took the "gene to pathway" approach by conducting Ingenuity Pathway Analysis (IPA) on the genes with a $P_{G \times E}$ value of $<0.05$ to identify overrepresented canonical pathways (Ingenuity Systems; www.ingenuity.com).

For G × E analysis at the SNP level, we analyzed the interactions of 82,881 SNPs with obesity or diabetes on the risk of pancreatic cancer using LRT in nested logistic regression model. SNPs were coded as 0, 1, or 2 for counts of the minor allele.

To control the problem of false-positive findings associated with multiple testing, we applied the Bonferroni correction for G × E interaction analysis at the pathway level. P values of $<1.45 \times 10^{-4}$ [0.05/(2 $\times$ 172)] were considered statistically significant at the pathway level. Because of the large number of genes/SNPs, we used the q value method with false discovery rate (FDR) at 0.10 as the significance threshold for G × E analysis at the gene/SNP level (26).

**Results**

The characteristics and exposure variables of the study populations are described in Table 1. There are no significant case–control differences in the distributions of age, race, and sex (all $P > 0.10$). More than 99% of participants were self-reported non-Hispanic Whites. The case–control association did not suggest any population stratification (adjusted $\lambda = 0.999$; ref. 27). The prevalence of obesity (BMI >30 kg/m²) was 21.1% versus 16.6%, and diabetes was 20.3% versus 9.5% in cases and controls, respectively. Obesity, diabetes, and smoking (≥20 pack-years) were significantly associated with increased risk of pancreatic cancer, with adjusted odds ratios (AOR) and 95% confidence intervals (95% CI) 1.22 (1.02–1.47), 2.35 (1.94–2.84), and 1.60 (1.38–1.86), respectively.

**G × E interactions at pathway level**

Among the 172 pathways tested, 40 pathways showed nominal interactions ($P < 0.05$) with obesity (Supplementary Table S2) and 18 with diabetes (Supplementary Table S3). One pathway (contributing genes) remained statistically significant and one nearly so after Bonferroni correction: the chemokine signaling pathway (GNGT2, RELA, and TIA1M) interacting with obesity ($P = 3.29 \times 10^{-5}$), the calcium signaling pathway with diabetes (GNAS; $P = 1.57 \times 10^{-4}$; Table 2). In addition, four additional top pathways, i.e., interaction of obesity with pathways in cancer, cytokine–cytokine receptor interaction pathway, as well as interaction of diabetes with mitogen-activated protein kinase (MAPK) signaling pathway, and pathways in cancer, are also shown in Table 2. We checked the sensitivity of the statistical method (LRT) to pathway size and found that the significance levels were unrelated to pathway size.
Furthermore, as a complementary approach to the above PCA-LRT analysis, we performed IPA analysis on nominally significant genes ($P < 0.05$) in $G \times C$ interactions at the gene level (next section). Several pathways that were highly significant at $P < 10^{-8}$ were identified: The role of RIG1-like receptors in antiviral innate immunity canonical pathway and the role of phosphoinositide 3-kinase (PI3K)/AKT signaling in the pathogenesis of influenza were most overrepresented in obesity-interacting genes, whereas molecular mechanisms of cancer pathway were most overrepresented in diabetes-interacting genes (Table 3).

G x E interactions at gene level
Among the 5,127 genes tested, 335 and 263 genes showed nominal interactions with obesity and diabetes, respectively ($P < 0.05$; Supplementary Tables S4 and S5). After adjusting for multiple comparisons, none of these interactions remained statistically significant. Twelve genes with the smallest $P$ values ($<0.001$) are listed in Table 4, including seven genes interacting with obesity and five genes interacting with diabetes. To overcome the reverse causality problem, we analyzed gene interactions with diabetes after excluding subjects with new onset diabetes ($<2$ years), but no significant change in the results was observed (data not shown).

### Table 1. Distribution of demographics and risk factors among cases and controls

<table>
<thead>
<tr>
<th>Variable</th>
<th>Case ($N = 2,028$)</th>
<th>Control ($N = 2,109$)</th>
<th>$P$ ($\chi^2$)</th>
<th>AOR (95% CI)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group, y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\leq 50$</td>
<td>199 (9.81)</td>
<td>236 (11.19)</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>51–60</td>
<td>563 (27.76)</td>
<td>575 (27.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61–70</td>
<td>710 (35.01)</td>
<td>713 (33.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$&gt;70$</td>
<td>556 (27.42)</td>
<td>585 (27.74)</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Race$^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic Whites</td>
<td>2,008 (99.26)</td>
<td>2,092 (99.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanics</td>
<td>8 (0.40)</td>
<td>13 (0.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blacks</td>
<td>0 (0)</td>
<td>2 (0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>7 (0.35)</td>
<td>2 (0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>920 (45.36)</td>
<td>968 (45.90)</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,108 (54.64)</td>
<td>1,141 (54.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking$^c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>801 (39.63)</td>
<td>1,008 (47.91)</td>
<td>$&lt;0.001$</td>
<td>1.43 (1.26–1.63)</td>
</tr>
<tr>
<td>Ever</td>
<td>1,220 (60.37)</td>
<td>1,096 (52.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pack-years$^c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0$</td>
<td>801 (39.63)</td>
<td>1,008 (47.91)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$&lt;20$</td>
<td>463 (22.91)</td>
<td>485 (23.05)</td>
<td>1.23 (1.04–1.45)</td>
<td></td>
</tr>
<tr>
<td>$\geq 20$</td>
<td>757 (37.46)</td>
<td>611 (29.04)</td>
<td>$&lt;0.001$</td>
<td>1.60 (1.38–1.86)</td>
</tr>
<tr>
<td>History of diabetes$^d$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,583 (79.71)</td>
<td>1,877 (90.50)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>403 (20.29)</td>
<td>197 (9.50)</td>
<td>$&lt;0.001$</td>
<td>2.35 (1.94–2.84)</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)$^e$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\leq 25$</td>
<td>764 (37.95)</td>
<td>885 (42.45)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>25–29.9</td>
<td>824 (40.93)</td>
<td>854 (40.96)</td>
<td>1.07 (0.93–1.24)</td>
<td></td>
</tr>
<tr>
<td>$\geq 30$</td>
<td>425 (21.11)</td>
<td>346 (16.59)</td>
<td>$&lt;0.001$</td>
<td>1.22 (1.02–1.47)</td>
</tr>
</tbody>
</table>

$^a$OR was adjusted for age, sex, smoking/pack-years, history of diabetes or BMI (categorical), and study sites.

$^b$Missing values from 5 cases.

$^c$Missing values from 7 cases and 5 controls.

$^d$Missing values from 42 cases and 35 controls.

$^e$Missing values from 15 cases and 24 controls.

(data not shown). Furthermore, as a complementary approach to the above PCA-LRT analysis, we performed IPA analysis on nominally significant genes ($P < 0.05$) in $G \times E$ interactions at the gene level (next section). Several pathways that were highly significant at $P < 10^{-8}$ were identified: The role of RIG1-like receptors in antiviral innate immunity canonical pathway and the role of phosphoinositide 3-kinase (PI3K)/AKT signaling in the pathogenesis of influenza were most overrepresented in obesity-interacting genes, whereas molecular mechanisms of cancer pathway were most overrepresented in diabetes-interacting genes (Table 3).
There are a total of 120 SNPs genotyped for GNGT2 (8 SNPs), RELA (5 SNPs), and TIAM1 (107 SNPs), the three major contributing genes in the chemokine signaling pathway. We first conducted LRT in the logistic regression model for each SNP and found that 17 SNPs were significant at the 0.05 level. We further analyzed the interaction pattern of the 17 SNPs using standard interaction analysis method and identified 8 synergisms, 4 antagonisms, and 5 undefined (Supplementary Table S8).

In light of the strong linkage between obesity and diabetes, we investigated the overlap between genes/SNPs interacting with obesity and diabetes on the risk of pancreatic cancer, as well as the overlap between genes/SNPs marginally associated with these two risk factors (Supplementary Table S9). At the significance level of 0.001, there were no overlapping genes/SNPs between obesity and diabetes; at a less stringent significance level of 0.01, there was a moderate 1% to 3% overlapping genes/SNPs. As a result, our analyses here did not support strong overlap between genetic factors interacting with obesity and diabetes on the risk of pancreatic cancer.

There are a total of 120 SNPs genotyped for GNGT2 (8 SNPs), RELA (5 SNPs), and TIAM1 (107 SNPs), the three major contributing genes in the chemokine signaling pathway. We first conducted LRT in the logistic regression model for each SNP and found that 17 SNPs were significant at the 0.05 level. We further analyzed the interaction pattern of the 17 SNPs using standard interaction analysis method and identified 8 synergisms, 4 antagonisms, and 5 undefined (Supplementary Table S8).

In light of the strong linkage between obesity and diabetes, we investigated the overlap between genes/SNPs interacting with obesity and diabetes on the risk of pancreatic cancer, as well as the overlap between genes/SNPs marginally associated with these two risk factors (Supplementary Table S9). At the significance level of 0.001, there were no overlapping genes/SNPs between obesity and diabetes; at a less stringent significance level of 0.01, there was a moderate 1% to 3% overlapping genes/SNPs. As a result, our analyses here did not support strong overlap between genetic factors interacting with obesity and diabetes on the risk of pancreatic cancer.

Table 2. Top six pathways interacting with obesity or diabetes in modifying the risk of pancreatic cancer

<table>
<thead>
<tr>
<th>KEGG code</th>
<th>Pathway description</th>
<th>Risk factor</th>
<th>Number of genes/genes with marginal effect</th>
<th>Number of SNPs/eigenSNPs in the interaction analysis</th>
<th>Major contributing genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04062a</td>
<td>Chemokine signalinga</td>
<td>Obesity</td>
<td>175/27</td>
<td>695/181</td>
<td>GNGT2 RELA TIAM1</td>
</tr>
<tr>
<td>hsa05200a</td>
<td>Pathways in cancer</td>
<td>Obesity</td>
<td>315/37</td>
<td>806/212</td>
<td>CBLC RELA</td>
</tr>
<tr>
<td>hsa04060</td>
<td>Cytokine–cytokine receptor interaction</td>
<td>Obesity</td>
<td>247/36</td>
<td>422/149</td>
<td>IFNA13 IL22RA1 IL2RA</td>
</tr>
<tr>
<td>hsa04020</td>
<td>Calcium signaling pathway</td>
<td>Diabetes</td>
<td>171/24</td>
<td>759/190</td>
<td>GNAS</td>
</tr>
<tr>
<td>hsa04010</td>
<td>MAPK signaling pathway</td>
<td>Diabetes</td>
<td>260/32</td>
<td>523/154</td>
<td>FOS MAP2K7</td>
</tr>
<tr>
<td>hsa05200</td>
<td>Pathways in cancer</td>
<td>Diabetes</td>
<td>315/37</td>
<td>806/212</td>
<td>DAPK3 EPAS1 FOS</td>
</tr>
</tbody>
</table>

aNumber of genes making up the pathway/number of genes survived the PCA-LRT ($P \leq 0.10$).

bNumber of SNPs in the “reconstructed” pathways/number of principal components for LRT.

P value was estimated by LRT in the logistic regression model with adjustment of age, sex, study site, pack-years (continuous), obesity or diabetes as appropriate, and five principal components for population structure.

cGenes with $P_g \times e \leq 0.05$ in logistic regression and $P \leq 0.10$ in PCA-LRT.

dPathways remained significant after the Bonferroni correction ($P < 1.45 \times 10^{-4}$).

Table 3. Top overrepresented canonical pathways in genes interacting with risk factors ($P < 10^{-6}$)

<table>
<thead>
<tr>
<th>Biologic process</th>
<th>Risk factor</th>
<th>P valuea</th>
<th>Ratiob</th>
<th>Contributing genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role of RIG1-like receptors in antiviral innate immunity</td>
<td>Obesity</td>
<td>$6.71 \times 10^{-11}$</td>
<td>12/49 (0.25)</td>
<td>TRAF6 RELA IFNA7 IFNA4 NFKB2 IFNA10 IFNA16 NFKB1 IFNA1/IFNA2 IFNAS IFNA14 IFNA6</td>
</tr>
<tr>
<td>Role of PI3K/AKT signaling in the pathogenesis of influenza</td>
<td>Obesity</td>
<td>$8.64 \times 10^{-9}$</td>
<td>12/74 (0.12)</td>
<td>RELA IFNA7 IFNA4 NFKB2 GS3KB IFNA10 IFNA16 NFKB1 IFNA1/IFNA2 IFNAS IFNA5 IFNA6</td>
</tr>
<tr>
<td>Molecular mechanisms of cancer</td>
<td>Diabetes</td>
<td>$1.03 \times 10^{-9}$</td>
<td>24/378 (0.063)</td>
<td>TP53 FYN ARHI3EF4 GNAS CYCS AXIN1 ADCY4 PRKAR2A ARHI3EF1 CDC42 RAC3 SINGA RB1 FOS CDH1 NFKB1 GNA1 PAK3 RHOA RASGRP1 PIK3CD BMP6 CHEK2 E2F2</td>
</tr>
</tbody>
</table>

aCalculated using the Fisher exact test (right-tailed).

bNumber of genes interacting with a risk factor of interest ($P \leq 0.05$) in a given pathway divided by total number of genes making up that pathway.
Discussion

In this large G × E analysis in pancreatic cancer, two approaches, from pathway to gene and from gene to pathway, suggest consistent findings that highlight the interactions of inflammatory response pathways with obesity and insulin resistance or cancer-related pathways with diabetes in modifying the risk of pancreatic cancer. We also observed that SNPs without marginal effects had strong differential effects on cancer risk between exposed and unexposed individuals. These preliminary findings underscore the potential value and the challenges of comprehensive G × E analysis in revealing molecular mechanisms that may underlie complex disease.

Using LRT-logistic regression analysis, the current study identified a statistically significant interaction of the chemokine signaling pathway with obesity in modifying the risk of pancreatic cancer. This association was consistently identified by both PCA-LRT and IPA analysis. The major contributing genes to these interactions included the oncogenic FBJ murine osteosarcoma viral oncogene homolog (aka EPAS1), the tumor-promoting angiogenic neoplasms (31). High frequency of GNAS mutations have been reported in multiple types of endocrine neoplasms (31). High frequency of GNAS mutations was also found in intraductal papillary mucinous neoplasm of the pancreas, but not in pancreatic ductal cancer (32). However, studies in mice indicate that mutations of this gene lead to obesity, glucose intolerance, and insulin resistance (33). It is possible that GNAS variants contributed to diabetes-associated pancreatic cancer via the mechanism of altered cAMP signaling transduction or enhanced insulin resistance. In addition to the calcium-signaling pathway, the most notable interaction of diabetes was with genes or pathways involved in cancer, which was consistently identified by both PCA-LRT and IPA approaches. The major contributing genes to these interactions included the oncogenic FOS (FB) murine osteosarcoma viral oncogene homolog, the tumor-promoting gene EPAS1 (endothelial PAS domain protein 1, aka
### Table 5. Top SNPs interacting with obesity or diabetes

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Gene</th>
<th>$P_g$</th>
<th>$q$ value</th>
<th>$\text{OR (95% CI)}$</th>
<th>$P$ value</th>
<th>$\text{MA# in case/control}$</th>
<th>$\text{OR (95% CI)}$</th>
<th>$P$ value</th>
<th>$\text{MA# in case/control}$</th>
<th>$\text{OR (95% CI)}$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs366193</td>
<td>ATP6V0A4</td>
<td>$4.3 \times 10^{-6}$</td>
<td>0.34</td>
<td>1.240/1,416</td>
<td>0.94 (0.86-1.04)</td>
<td>$2.4 \times 10^{-1}$</td>
<td>373/221</td>
<td>1.68 (1.36-2.07)</td>
<td>$1.1 \times 10^{-6}$</td>
<td>1.05 (0.96-1.14)</td>
<td>$3.2 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs13070718</td>
<td>RAB5A</td>
<td>1.2</td>
<td>0.35</td>
<td>459/424</td>
<td>1.23 (1.06-1.41)</td>
<td>$4.8 \times 10^{-3}$</td>
<td>94/118</td>
<td>0.61 (0.45-0.81)</td>
<td>$7.2 \times 10^{-4}$</td>
<td>1.07 (0.95-1.22)</td>
<td>$2.7 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs7043149</td>
<td>TJP2</td>
<td>1.3</td>
<td>0.35</td>
<td>1,358/1,613</td>
<td>0.88 (0.79-0.98)</td>
<td>$6.7 \times 10^{-3}$</td>
<td>399/268</td>
<td>1.41 (1.15-1.73)</td>
<td>$8.8 \times 10^{-4}$</td>
<td>0.95 (0.88-1.04)</td>
<td>$2.9 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs9314865</td>
<td>TJP2</td>
<td>3.2</td>
<td>0.47</td>
<td>1,476/1,510</td>
<td>1.14 (1.04-1.26)</td>
<td>$6.6 \times 10^{-3}$</td>
<td>351/341</td>
<td>0.73 (0.60-0.89)</td>
<td>$2.4 \times 10^{-3}$</td>
<td>1.05 (0.97-1.15)</td>
<td>$2.5 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs2399424</td>
<td>TJP2</td>
<td>3.3</td>
<td>0.47</td>
<td>1,476/1,511</td>
<td>1.14 (1.04-1.26)</td>
<td>$6.6 \times 10^{-3}$</td>
<td>351/341</td>
<td>0.73 (0.60-0.89)</td>
<td>$2.4 \times 10^{-3}$</td>
<td>1.05 (0.97-1.15)</td>
<td>$2.5 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs4323331</td>
<td>TJP2</td>
<td>3.5</td>
<td>0.47</td>
<td>1,143/1,335</td>
<td>0.91 (0.83-1.01)</td>
<td>$7.4 \times 10^{-2}$</td>
<td>335/215</td>
<td>1.46 (1.18-1.8)</td>
<td>$5.1 \times 10^{-4}$</td>
<td>0.99 (0.91-1.09)</td>
<td>$8.6 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs4338173</td>
<td>TJP2</td>
<td>4.5</td>
<td>0.51</td>
<td>1,142/1,331</td>
<td>0.92 (0.83-1.01)</td>
<td>$8.3 \times 10^{-2}$</td>
<td>335/215</td>
<td>1.46 (1.18-1.8)</td>
<td>$5.1 \times 10^{-4}$</td>
<td>0.99 (0.91-1.09)</td>
<td>$9.0 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs10818684</td>
<td>PTGS1</td>
<td>7.9</td>
<td>0.06</td>
<td>869/962</td>
<td>1.09 (0.98-1.21)</td>
<td>$1.2 \times 10^{-1}$</td>
<td>174/136</td>
<td>0.53 (0.40-0.69)</td>
<td>$2.3 \times 10^{-6}$</td>
<td>0.99 (0.89-1.09)</td>
<td>$8.0 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs10179599</td>
<td>RPL31</td>
<td>6.6</td>
<td>0.26</td>
<td>667/722</td>
<td>1.11 (0.99-1.25)</td>
<td>$7.5 \times 10^{-2}$</td>
<td>133/107</td>
<td>0.54 (0.40-0.72)</td>
<td>$2.0 \times 10^{-6}$</td>
<td>1.01 (0.90-1.12)</td>
<td>$8.8 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs2872220</td>
<td>WEE1</td>
<td>1.0</td>
<td>0.26</td>
<td>407/549</td>
<td>0.86 (0.75-0.98)</td>
<td>$2.7 \times 10^{-2}$</td>
<td>130/32</td>
<td>2.19 (1.46-3.30)</td>
<td>$1.2 \times 10^{-1}$</td>
<td>0.96 (0.84-1.09)</td>
<td>$5.1 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs2278725</td>
<td>RPL31, NAPAS2</td>
<td>1.9</td>
<td>0.33</td>
<td>821/874</td>
<td>1.14 (1.03-1.28)</td>
<td>$1.6 \times 10^{-2}$</td>
<td>168/121</td>
<td>0.60 (0.46-0.79)</td>
<td>$2.3 \times 10^{-4}$</td>
<td>1.05 (0.95-1.16)</td>
<td>$3.6 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs1771769</td>
<td>OR1J1</td>
<td>2.2</td>
<td>0.33</td>
<td>397/396</td>
<td>1.21 (1.04-1.40)</td>
<td>$1.3 \times 10^{-2}$</td>
<td>71/63</td>
<td>0.51 (0.36-0.74)</td>
<td>$2.5 \times 10^{-4}$</td>
<td>1.07 (0.93-1.23)</td>
<td>$3.2 \times 10^{-1}$</td>
</tr>
<tr>
<td>rs1301747</td>
<td>RPL31, NAPAS2</td>
<td>2.5</td>
<td>0.33</td>
<td>821/874</td>
<td>1.14 (1.03-1.28)</td>
<td>$1.6 \times 10^{-2}$</td>
<td>168/120</td>
<td>0.61 (0.46-0.8)</td>
<td>$3.0 \times 10^{-4}$</td>
<td>1.05 (0.95-1.16)</td>
<td>$3.5 \times 10^{-1}$</td>
</tr>
</tbody>
</table>

*aObtained from LRT-nested in logistic regression models with adjustment for age, sex, study site, diabetes or obesity, pack-years (quantitative), and five principal components for population structure.

bMinor allele counts in cases and controls.

cOdds ratio (95% confidence interval) not adjusted for covariates and $P$ values from a $\chi^2$ test.

dOne SNP may be assigned to two genes because SNPs located 20 kb up- or downstream of the gene region were included for each gene.
hypoxia-inducible factor 2 α, a tumor suppressor DAPK3 (death-associated protein kinase 3), and MAP2K7 (aka MEK7, JNK2, and SKK4; ref. 34). MAP2K7 mediates the cellular responses to proinflammatory cytokines and environmental stresses with a strong preference for activation of the c-jun-NH2-kinase (JNK) pathway (35). JNK signaling plays a central role in obesity and insulin resistance (36) as well as in regulating apoptosis (37). FOS proteins can dimerize with c-jun, thereby forming the transcription factor complex AP-1 that regulates cell proliferation, differentiation, and transformation as well as apoptosis (38). Overall, the results of our study highlight pathways and genes that have been implicated in cancer development, especially those associated with insulin resistance and apoptosis, in diabetes-related pancreatic cancer.

Several studies have previously suggested the possibility to increase statistical power of G × E analyses by focusing on genes with marginal effects only (24). Our findings that SNPs with the smallest P value for interaction were those without any marginal effect suggest that G × E analysis limited to such genes/SNPs may miss genetic variants that have a true impact on disease risk among exposed individuals only, consistent with a recently reported SNP by alcohol intake interaction influencing the risk of esophageal squamous-cell carcinoma (39, 40). Thus, a comprehensive G × E analysis of GWAS data using multiple analytic methods with complementary strengths as undertaken here and suggested by previous research (41) may be a necessary and useful approach to unveiling missing heritability of a complex disease such as pancreatic cancer (42). It would be interesting to develop hybrid strategies, in line with that of ref. (25), for pathway by environment interaction analysis in the future.

Our study has several strengths and limitations. This is the largest G × E analysis in pancreatic cancer with the most comprehensive analysis of all biologic pathways identified from KEGG using an agnostic approach. We used a PCA approach to reduce the dimensionality of the GWAS data and increased the probability of finding useful information. The analysis was based on high-quality genotype and exposure data with extensive quality control measures. We also applied stringent criteria to control false-positive reporting. However, our sample size is still relatively small for a full G × E GWAS analysis. Our findings cannot be replicated due to the lack of available datasets. Thus, the possibility that some associations are spurious findings cannot be excluded, which limits the generalization of the results. Nevertheless, the pathways and genes found interacting with obesity and diabetes are highly relevant to pancreatic cancer and are supported by other experimental evidence. Our results underscore the interactions of inflammation-related genes with obesity and insulin resistance or cancer-related genes with diabetes in modifying the risk of pancreatic cancer. G × E analysis offers an opportunity to identify genetic factors linking obesity and diabetes to pancreatic cancer. Such information would provide scientific rationales for the development of novel strategies in personalized prevention of pancreatic cancer.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors' Contributions
Conception and design: H. Tang, P. Wei, H.A. Risch, H.B. Bueno-de-Mesquita, P. Mueters, K.-T. Khaw, D. Li
Development of methodology: H. Tang, P. Wei, E.A. Holly, H.P.M. Peeters
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.) : J.C. Duell, H.A. Risch, D.H. Olson, H.B. Bueno-de-Mesquita, S. Gallinger, E.A. Holly, C.M. Petersen, R.R. McWilliams, A. Tjonneland, M.-C. Boutron-Ruault, R. Kaaks, D. Trichopoulos, M. Sund, P.H.M. Peeters, K.-T. Khaw, D. Li
Analysis and interpretation of data (e.g., statistical analysis, bioinformatics, computational analysis) : H. Tang, P. Wei, H.A. Risch, P. Mueters, J. C. Amos, P. H. M. Peeters, D. Li
Writing, review, and/or revision of the manuscript: H. Tang, P. Wei, E.J. Duell, H.A. Risch, D.H. Olson, H.B. Bueno-de-Mesquita, E.A. Holly, G.M. Petersen, P.M. Bracci, M. Jenab, E. Riboli, A. Tjonneland, M.-C. Boutron-Ruault, R. Kaaks, D. Trichopoulos, M. Sund, P.H.M. Peeters, K.-T. Khaw, C.J. Amos, D. Li
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): H. Tang, S. Gallinger, P. H. M. Peeters, K.-T. Khaw, D. Li
Study supervision: H. Tang, P. Wei, H.A. Risch, E.A. Holly, D. Li

Grant Support
This study was supported from the NIH through a supplemental grant R01 CA98380-05 (to D. Li), an MD Anderson Cancer Center Support Grant CA016672, NIH grants R01 CA169122 and R01 HL106034 (to P. Wei), and the Sheikh Ahmed Center for Pancreatic Cancer Research Funds (to D. Li). S. Gallinger was supported by NIH grant R01 CA096755. P.M. Braccia was supported by NIH grants R01 CA59706, R01 CA110976, R01 CA110976S1, R01 CA154823, and R03 CA89726. K.-T. Khaw was supported by Medical Research Council UK and Cancer Research UK.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 26, 2013; revised October 3, 2013; accepted October 8, 2013; published OnlineFirst October 17, 2013.

References


Genes–Environment Interactions in Obesity- and Diabetes-Associated Pancreatic Cancer: A GWAS Data Analysis

Hongwei Tang, Peng Wei, Eric J. Duell, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-13-0437-T

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2013/10/17/1055-9965.EPI-13-0437-T.DC1

Cited articles
This article cites 41 articles, 8 of which you can access for free at:
http://cebp.aacrjournals.org/content/23/1/98.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.