Published in final edited form as:
Cancer Epidemiol Biomarkers Prev. 2014 January ; 23(1): 98–106. doi:10.1158/1055-9965.EPI-13-0437-T.

Genes-environment interactions in obesity- and diabetes-associated pancreatic cancer: A GWAS data analysis

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

[1]Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA [2]Division of Biostatistics and Human Genetics Center, School of Public Health, University of Texas Health Science Center, Houston, TX 77030 [3]Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain [4]Yale University School of Public Health, New Haven, CT, USA [5]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA [6]National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands and Department of Gastroenterology and Hepatology, University Medical Center Utrecht, Utrecht, The Netherlands [7]Samuel Lunenfeld Research Institute, Toronto General Hospital, University of Toronto, Toronto, Canada [8]Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA [9]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA [10]International Agency for Research on Cancer, Lyon, France [11]Division of Epidemiology, Public Health and Primary Care, Imperial College London, London, UK [12]Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark [13]Inserm, Centre for research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women's Health team, F-94805, Villejuif, France; Univ Paris Sud, UMRS 1018, F-94805, Villejuif, France; IGR, F-94805, Villejuif, France [14]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany [15]Department of Epidemiology, Harvard School of Public Health, 677 Huntington, Avenue, Boston, MA 02115, USA; Bureau of Epidemiologic Research, Academy of Athens, 28 Panepistimiou Street, Athens, GR-106 79, Greece; Hellenic Health Foundation, Kaisareias 13 & Alexandroupoleos Street, GR-115 27, Athens Greece [16]Dipartimento di Medicina Clinica e Chirurgia, Federico II, University, Naples Italy [17]Department of Surgery, Umeå University Hospital, Umeå, Sweden [18]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; Division of Epidemiology, Public Health and Primary Care, Imperial College London, London, UK [19]School of Clinical Medicine, University of Cambridge, UK [20]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

[†]All correspondence should be addressed to: Donghui Li, Ph.D., Department of Gastrointestinal Medical Oncology, UT MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 426, Houston, TX 77030, Phone: 713 834 6690, Fax: 713 834 6153, dli@mdanderson.org.
[*]Current address: Christopher I Amos, Department of Community and Family Medicine Geisel School of Medicine, Dartmouth College, Lebanon, NH 03766

The authors have nothing to disclose.

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.

Methods—Using 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 employing the likelihood ratio test (LRT) nested in logistic regression models and Ingenuity Pathway Analysis (IPA).

Results—After adjusting for multiple comparisons, a significant interaction of the chemokine signaling pathway with obesity ($P = 3.29 \times 10^{-6}$) and a near significant interaction of calcium signaling pathway with diabetes ($P = 1.57 \times 10^{-4}$) in modifying the risk of pancreatic cancer was 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 *GNGT2*, *RELA*, *TIAM1* and *GNAS*. None of the individual genes or SNPs except one SNP remained significant after adjusting for multiple testing. Notably, SNP rs10818684 of the *PTGS1* gene showed an interaction with diabetes ($P = 7.91 \times 10^{-7}$) at a false discovery rate of 6%.

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.

Impact—Gene-environment interaction analysis may provide new insights into the genetic susceptibility and molecular mechanisms of obesity- and diabetes-related pancreatic cancer.

Keywords

GWAS; obesity; diabetes; interaction; pancreatic cancer; genetic susceptibility

Introduction

Pancreatic cancer is the fourth leading cause of cancer death, accounting for more than 37,600 deaths each year in the United States (1). Epidemiological 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 biological knowledge of gene function and aims at combining moderate signals of SNPs 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 biological 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 biological pathways, contributing to the susceptibility of complex human diseases. While 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 (aka MODY (Mature Onset Diabetes of the Young) 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)(17), we conducted a comprehensive gene-environment (G x 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 MD Anderson Cancer Center, Mayo Clinic, Yale University, University of California at San Francisco, Memorial Sloan-Kettering Cancer Center, and University of Toronto, 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) was downloaded from the Database of Genotypes and Phenotypes (dbGaP) website (http://www.ncbi.nlm.gov/gap) with the approval of MD Anderson Institutional Review Board. We first conducted data cleaning and quality control by removing SNPs with minor allele frequency (MAF) < 5%, deviating from 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 3 release #3, NCBI build 36, dbSNP b126, 2010-05-28, MAF > 5%) for CEU, JPT/CHB, and YRI (18), we used 10,155 high-quality markers ($r^2 < 0.004$) 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 as the study subjects in the current analysis. Then, we derived the top five principal components for population substructure in the Caucasian subjects using the EIGENSTRAT (20).

Definition of pathways and genes

The pathways and genes used in the current analysis are defined as previously described (11). A total of 214 human biological pathways were downloaded from the KEGG website (21). Of these, 197 pathways with 10–500 genes each were considered in the analyses. Gene lists were downloaded from the human genome database version18 (hg18) using the UCSC Table Browser data retrieval tool (22). SNPs within 20 kb upstream or downstream of genes

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 agreements. Exposure variables included age, sex, race/ethnicity, adulthood body mass index (BMI, weight/height^2), 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 x E analysis as follows: obesity (BMI ≤30 kg/m^2 vs. >30 kg/m^2) 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 value for family history of cancer, this variable was not considered in the model.

Statistical methods

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. Prior to 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 x SNP/SNP x environment interaction analysis proposed by (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 x 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 1). Genes with a $P_{G\ x\ E}$ value < 0.05 in the interaction analysis were considered as the major contributing gene(s) 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 x E analysis at the gene level, a total of 5,127 genes were tested using LRT and logistic regression. SNPs with $P_{G\ x\ E}$ value <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 $P_{G\ x\ E}$ value ≤ 0.05 to identify over-represented canonical pathways (Ingenuity® Systems, www.ingenuity.com).

For G x 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 x E interaction analysis at the pathway level. *P* values < 1.45×10^{-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 value method with false discovery rate (FDR) at 0.10 as the significance threshold for G x 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. Case-control association did not suggest any population stratification (adjusted lambda = 0.999) (27). The prevalence of obesity (BMI > 30 kg/m^2) was 21.1% vs. 16.6%, and diabetes was 20.3% vs. 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 x E interactions at pathway level

Among the 172 pathways tested, 40 pathways showed nominal interactions ($P < 0.05$) with obesity (Supplementary Table 2) and 18 with diabetes (Supplementary Table 3). One pathway (contributing genes) remained statistically significant and one nearly so after Bonferroni correction: The chemokine signaling pathway (*GNGT2, RELA,* and *TIAM1)* interacting with obesity ($P = 3.29 \times 10^{-6}$), 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 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 (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 x 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 PI3K/AKT signaling in the pathogenesis of influenza were most over-represented in obesity-interacting genes, while molecular mechanisms of cancer pathway was most over-represented 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 4 and 5). 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).

G x E interactions at SNP level

A total of 3,859 and 3,551 SNPs exhibited nominal interactions with obesity and diabetes, respectively ($P < 0.05$), which were identified as the contributing SNPs to the genes with nominal interactions. Among these SNPs, 810 interactions with obesity and 758 interactions

with diabetes at the level of $P < 0.01$ are presented in Supplementary Tables 6 and 7, respectively. There are seven interactions with obesity and six with diabetes that had a p value of $<10^{-5}$ (Table 5). One SNP (rs10818684) of the *PTGS1* (aka *COX1*) gene actually had a q value of 0.06, which was significant at FDR < 10%. Notably, all of the top 13 SNPs displayed a differential effect on risk of pancreatic cancer between exposed (obese or diabetic) and unexposed (non-obese or non-diabetic) groups and none of them had marginal effect on risk of pancreatic cancer when the analysis was conducted in the combined dataset of exposed and unexposed individuals (Table 5).

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 likelihood ratio test (LRT) in the logistic regression model for each SNP and found 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 8).

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 9). 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.

Discussion

In this large G x 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 x 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 supported by findings from another statistical approach, i.e. IPA analysis. The major contributing genes to the chemokine signaling pathway and the top two canonical pathways identified by IPA (Table 3), e.g. *RELA, GNGT2, NFKB1, NFKB2*, and *IFNA or interleukin* genes, suggest a central role of the NFκB (Nuclear Factor kappa B) signaling mediated inflammatory and immune responses in obesity-related pancreatic cancer (2). GNGT2 (guanine nucleotide binding protein (G protein)), gamma transducing activity polypeptide 2) has been shown to mediate β-arrestin 1-induced Akt phosphorylation and NFκB activation (28). The *RELA* gene encodes the p65 protein which binds to NFκB1 forming the most abundant form of NFκB (29). NFκB is activated by many proinflammatory cytokines, and it is constitutively activated in pancreatic cancer. Increased NF-κB activity inhibits apoptosis and promotes growth, tumorigenesis, angiogenesis, invasion, and metastasis (30). Observations from this study suggest that genetic variations conferring pro-inflammatory responses may act in concerts with the chronic inflammatory state of obesity in increasing the risk of pancreatic cancer.

The current study found a nearly significant interaction of diabetes with calcium signaling pathway in modifying the risk of pancreatic cancer. In addition, nominal interactions of MAPK (mitogen-activated protein kinase) signaling pathway and pathway in cancer with diabetes were also observed. IPA analysis found that genes in molecular mechanisms of cancer were most overrepresented among diabetes-interacting genes. The physiological and biochemical roles of calcium signaling range widely, and how this pathway interacts with diabetes in modifying the risk of pancreatic cancer remains unclear. The single significant gene contributing to this pathway was *GNAS* (*GNAS* complex locus, aka adenylate cyclase-stimulating G alpha protein), which encodes the G protein α unit that couples receptors to the generation of intracellular cyclic AMP (cAMP). *GNAS* mutations have been reported in multiple types of endocrine neoplasms (31). High frequency of *GNAS* mutations were 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* (FBJ murine osteosarcoma viral oncogene homolog), the tumor promoting gene *EPAS1* (endothelial PAS domain protein 1, aka hypoxia-inducible factor 2 alpha), a tumor suppressor *DAPK3* (death-associated protein kinase 3) and *MAP2K7* (aka *MEK7, JNKK2, and SKK4*) (34). *MAP2K7* mediates the cellular responses to proinflammatory cytokines, and environmental stresses with a strong preference for activation of the JNK (c-Jun N-terminal Kinase) 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 x 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 x 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 (ESCC)(39, 40). Thus, comprehensive G x E analysis of GWAS data using multiple analytical 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 complex disease such as pancreatic cancer (42). It would be interesting to develop hybrid strategies, in line with that of (25), for pathway by environment interaction analysis in the future.

Our study has several strengths and limitations. This is the largest G x E analysis in pancreatic cancer with the most comprehensive analysis of all biological 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 x 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 x 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported from the National Institutes of Health through a supplemental grant to R01 CA98380-05 (D.Li), a MD Anderson Cancer Center Support Grant CA016672, NIH grants R01 CA169122 and R01 HL106034 (P.Wei), and the Sheikh Ahmed Center for Pancreatic Cancer Research Funds (D.Li). S. Gallinger was supported by NIH grant R01 CA09075. P.M. Bracci was supported by NIH grants R01 CA59706, R01 CA109767, R01 CA109767S1, R01 CA154823 and R03 CA89726. K.T. Khaw was supported by Medical Research Council UK and Cancer Research UK.

References

1. ACS. Cancer Facts and Figures 2012. American Cancer Society; 2012.
2. Bracci PM. Obesity and pancreatic cancer: overview of epidemiologic evidence and biologic mechanisms. Mol Carcinog. 2012; 51:53–63. [PubMed: 22162231]
3. Li D. Diabetes and pancreatic cancer. Mol Carcinog. 2012; 51:64–74. [PubMed: 22162232]
4. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet. 2009; 41:986–90. [PubMed: 19648918]
5. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010; 42:224–8. [PubMed: 20101243]
6. Yu K, Li Q, Bergen AW, Pfeiffer RM, Rosenberg PS, Caporaso N, et al. Pathway analysis by adaptive combination of P-values. Genet Epidemiol. 2009; 33:700–9. [PubMed: 19333968]
7. Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. Am J Hum Genet. 2010; 86:6–22. [PubMed: 20074509]
8. Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. Annu Rev Public Health. 2010; 31:21–36. [PubMed: 20070199]
9. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A. 2012; 109:1193–8. [PubMed: 22223662]
10. Li D, Duell EJ, Yu K, Risch HA, Olson SH, Kooperberg C, et al. Pathway analysis of genome-wide association study data highlights pancreatic development genes as susceptibility factors for pancreatic cancer. Carcinogenesis. 2012; 33:1384–90. [PubMed: 22523087]
11. Wei P, Tang H, Li D. Insights into pancreatic cancer etiology from pathway analysis of genome-wide association study data. PLoS One. 2012; 7:e46887. [PubMed: 23056513]
12. Fox CS, Heard-Costa N, Cupples LA, Dupuis J, Vasan RS, Atwood LD. Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project. BMC Med Genet. 2007; 8 (Suppl 1):S18. [PubMed: 17903300]
13. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010; 42:579–89. [PubMed: 20581827]

14. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010; 42:937–48. [PubMed: 20935630]
15. Tang H, Dong X, Hassan M, Abbruzzese JL, Li D. Body mass index and obesity- and diabetes-associated genotypes and risk for pancreatic cancer. Cancer Epidemiol Biomarkers Prev. 2011; 20:779–92. [PubMed: 21357378]
16. Pierce BL, Austin MA, Ahsan H. Association study of type 2 diabetes genetic susceptibility variants and risk of pancreatic cancer: an analysis of PanScan-I data. Cancer Causes & Control. 2011; 22:877–83. [PubMed: 21445555]
17. Petersen GM, Boffetta P. Carcinogenesis of pancreatic cancer: challenges, collaborations, progress. Mol Carcinog. 2012; 51:1–2. [PubMed: 22162226]
18. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449:851–61. [PubMed: 17943122]
19. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–59. [PubMed: 10835412]
20. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]
21. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 2010; 38:D355–60. [PubMed: 19880382]
22. Karolchik DHA, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004; 32(Database issue):D493–6. [PubMed: 14681465]
23. Chen LS, Hutter CM, Potter JD, Liu Y, Prentice RL, Peters U, et al. Insights into Colon Cancer Etiology via a Regularized Approach to Gene Set Analysis of GWAS Data. Am J Hum Genet. 2010
24. Kooperberg C, Leblanc M. Increasing the power of identifying gene x gene interactions in genome-wide association studies. Genet Epidemiol. 2008; 32:255–63. [PubMed: 18200600]
25. Murcray CE, Lewinger JP, Conti DV, Thomas DC, Gauderman WJ. Sample size requirements to detect gene-environment interactions in genome-wide association studies. Genetic Epidemiology. 2011; 35:201–10. [PubMed: 21308767]
26. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003; 100:9440–5. [PubMed: 12883005]
27. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008; 17:R122–8. [PubMed: 18852200]
28. Yang M, He RL, Benovic JL, Ye RD. beta-Arrestin1 interacts with the G-protein subunits beta1gamma2 and promotes beta1gamma2-dependent Akt signalling for NF-kappaB activation. Biochem J. 2009; 417:287–96. [PubMed: 18729826]
29. Perkins ND. Integrating cell-signalling pathways with NF-kappaB and IKK function. Nat Rev Mol Cell Biol. 2007; 8:49–62. [PubMed: 17183360]
30. Holcomb B, Yip-Schneider M, Schmidt CM. The role of nuclear factor kappaB in pancreatic cancer and the clinical applications of targeted therapy. Pancreas. 2008; 36:225–35. [PubMed: 18362834]
31. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011; 39:D945–50. [PubMed: 20952405]
32. Furukawa T, Kuboki Y, Tanji E, Yoshida S, Hatori T, Yamamoto M, et al. Whole-exome sequencing uncovers frequent GNAS mutations in intraductal papillary mucinous neoplasms of the pancreas. Sci Rep. 2011; 1:161. [PubMed: 22355676]
33. Chen M, Nemechek NM, Mema E, Wang J, Weinstein LS. Effects of deficiency of the G protein Gsalpha on energy and glucose homeostasis. Eur J Pharmacol. 2011; 660:119–24. [PubMed: 21208600]

34. Brognard J, Zhang YW, Puto LA, Hunter T. Cancer-associated loss-of-function mutations implicate DAPK3 as a tumor-suppressing kinase. Cancer Res. 2011; 71:3152–61. [PubMed: 21487036]
35. Kyriakis JM, Avruch J. Mammalian MAPK signal transduction pathways activated by stress and inflammation: a 10-year update. Physiological Reviews. 2012; 92:689–737. [PubMed: 22535895]
36. Hirosumi J, Tuncman G, Chang L, Gorgun CZ, Uysal KT, Maeda K, et al. A central role for JNK in obesity and insulin resistance. Nature. 2002; 420:333–6. [PubMed: 12447443]
37. Bogoyevitch MA, Ngoei KRW, Zhao TT, Yeap YYC, Ng DCH. c-Jun N-terminal kinase (JNK) signaling: recent advances and challenges. Biochimica et Biophysica Acta. 2010; 1804:463–75. [PubMed: 19900593]
38. Asim M, Chaturvedi R, Hoge S, Lewis ND, Singh K, Barry DP, et al. Helicobacter pylori induces ERK-dependent formation of a phospho-c-Fos c-Jun activator protein-1 complex that causes apoptosis in macrophages. J Biol Chem. 2010; 285:20343–57. [PubMed: 20410304]
39. Wu C, Chang J, Ma B, Miao X, Zhou Y, Liu Y, et al. The Case-Only Test for Gene–Environment Interaction is Not Uniformly Powerful: An Empirical Example. Genetic Epidemiology. 2013
40. Wu C, Kraft P, Zhai K, Chang J, Wang Z, Li Y, et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. Nat Genet. 2012; 44:1090–7. [PubMed: 22960999]
41. Hsu L, Jiao S, Dai JY, Hutter C, Peters U, Kooperberg C. Powerful cocktail methods for detecting genome-wide gene-environment interaction. Genet Epidemiol. 2012; 36:183–94. [PubMed: 22714933]
42. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–53. [PubMed: 19812666]

Table 1

Distribution of demographics and risk factors among cases and controls

Variable	Case (n = 2,028) n (%)	Control (n =2,109) n (%)	$P(\chi^2)$	AOR(95%CI)[a]
Age group				
≤50	199 (9.81)	236 (11.19)		
51–60	563 (27.76)	575 (27.26)		
61–70	710 (35.01)	713 (33.81)		
>70	556 (27.42)	585 (27.74)	0.49	
Race[b]				
Non-Hispanic Whites	2,008 (99.26)	2,092 (99.19)		
Hispanics	8 (0.40)	13 (0.62)		
Blacks	0 (0)	2 (0.09)		
Others	7 (0.35)	2 (0.09)	0.12	
Sex				
Female	920 (45.36)	968 (45.90)		
Male	1,108 (54.64)	1,141 (54.10)	0.73	
Smoking[c]				
Never	801 (39.63)	1,008 (47.91)		1.00
Ever	1,220 (60.37)	1,096 (52.09)	<0.001	1.43 (1.26–1.63)
Pack-years[c]				
0	801 (39.63)	1008 (47.91)		1.00
<20	463 (22.91)	485 (23.05)		1.23 (1.04–1.45)
≥20	757 (37.46)	611 (29.04)	<0.001	1.60 (1.38–1.86)
History of diabetes[d]				
No	1,583 (79.71)	1,877 (90.50)		1.00
Yes	403 (20.29)	197 (9.50)	<0.001	2.35 (1.94–2.84)
BMI (kg/m^2)[e]				
≤25	764 (37.95)	885 (42.45)		1.00
25–29.9	824 (40.93)	854 (40.96)		1.07 (0.93–1.24)
≥30	425 (21.11)	346 (16.59)	<0.001	1.22 (1.02–1.47)

Abbreviations: AOR: adjusted odds ratio; 95% CI: 95% confidence interval; BMI: body mass index.

[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.

Table 2

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

KEGG code	Pathway description	Risk factor	No. of genes/genes with marginal effect[a]	No. of SNPs/eigenSNPs in the interaction analysis[b]	$P_{G \times E}$[c]	Major contributing genes[d]
hsa04062[e]	Chemokine Signaling[e]	Obesity	175/27	695/181	3.29×10^{-6}	*GNGT2 RELA TIAM1*
hsa05200	Pathways in cancer	Obesity	315/37	806/212	5.35×10^{-4}	*CBLC RELA*
hsa04060	Cytokine-cytokine receptor interaction	Obesity	247/36	422/149	6.97×10^{-4}	*IFNA13 IL22RA1 IL2RA*
hsa04020	Calcium signaling pathway	Diabetes	171/24	759/190	1.57×10^{-4}	*GNAS*
hsa04010	MAPK signaling pathway	Diabetes	260/32	523/154	3.56×10^{-4}	*FOS MAP2K7*
hsa05200	Pathways in cancer	Diabetes	315/37	806/212	4.46×10^{-4}	*DAPK3 EPAS1 FOS*

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

[b] Number of SNPs in the "reconstructed" pathways/number of principal components for LRT.

[c] *P* value was estimated by LRT in 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.

[d] Genes with $P_{G \times E} \leq 0.05$ in logistic regression and $P \leq 0.10$ in PCA-LRT.

[e] Pathways remained significant after Bonferroni correction ($P < 1.45 \times 10^{-4}$)

Table 3

Top overrepresented canonical pathways in genes interacting with risk factors ($P < 10^{-8}$)

Biological process	Risk factor	*P* Value[a]	Ratio[b]	Contributing genes
Role of RIG1-like Receptors in Antiviral Innate Immunity	Obesity	6.71×10^{-11}	12/49 (0.25)	*TRAF6 RELA IFNA7 IFNA4 NFKB2 IFNA10 IFNA16 NFKB1 IFNA1/IFNA13 IFNA5 IFNA14 IFNA6*
Role of PI3K/AKT Signaling in the Pathogenesis of Influenza	Obesity	8.64×10^{-9}	12/74 (0.12)	*RELA IFNA7 IFNA4 NFKB2 GSK3B IFNA10 IFNA16 NFKB1 IFNA1/IFNA13 IFNA5 IFNA14 IFNA6*
Molecular Mechanisms of Cancer	Diabetes	1.03×10^{-9}	24/378 (0.063)	*TP53 FYN ARHGEF4 GNAS CYCS AXIN1 ADCY4 PRKAR2A ARHGEF1 CDC42 RAC3 SIN3A RB1 FOS CDH1 NFKBIA GNAT1 PAK3 RHOA RASGRP1 PIK3CD BMP6 CHEK2 E2F2*

[a] Calculated using Fisher's exact test (right-tailed).

[b] Number 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.

Table 4

Top genes interacting with obesity or diabetes on risk of pancreatic cancer

Chromosome location	Gene	Gene Full name	Risk factor	No. of SNPs/eigenSNP[a]	$P_{G \times E}$[b]
2q33	*AOX1*	*aldehyde oxidase 1*	Obesity	28/6	2.70×10^{-4}
3p24-p22	*RAB5A*	*RAB5A, member RAS oncogene family*	Obesity	15/4	3.31×10^{-4}
4q21	*HERC3*	*hect domain and RLD 3*	Obesity	13/3	5.64×10^{-4}
22q13.1	*GALR3*	*galanin receptor 3*	Obesity	4/2	6.46×10^{-4}
22q13.1	*GCAT*	*glycine C-acetyltransferase*	Obesity	5/2	6.70×10^{-4}
9q13-q21	TJP2	*tight junction protein 2 (zona occludens 2)*	Obesity	24/7	6.71×10^{-4}
8p11.21	CHRNA6	*cholinergic receptor, nicotinic, alpha 6*	Obesity	2/2	7.72×10^{-4}
11p15.3-p15.1	*WEE1*	*WEE1 homolog (S. pombe)*	Diabetes	8/3	3.02×10^{-5}
9q33.2	*OR1J1*	*olfactory receptor, family 1, subfamily J, member 1*	Diabetes	13/6	1.69×10^{-4}
2q11.2	*RPL31*	*ribosomal protein L31*	Diabetes	14/3	2.01×10^{-4}
20q13.2-q13.3	*MC3R*	*melanocortin 3 receptor*	Diabetes	20/7	2.22×10^{-4}
9q13	*GCNT1*	*glucosaminyl (N-acetyl) transferase 1, core 2*	Diabetes	26/6	9.27×10^{-4}

[a] Number of SNPs in the gene/number of principal components used for analysis.

[b] *P* value was estimated by the interaction term in logistic regression model with adjustment of age, sex, study site, pack years (quantitative), obesity or diabetes as appropriate, and five principal components for population structure.

Table 5

Top SNPs interacting with obesity or diabetes

SNPs	Gene	$P_{G \times E}$ [a]	q value	[b] MA# in case/ control	OR (95% CI)[c]	*P* value[c]	MA# in case/control	OR (95% CI)	*P* Value	OR (95% CI)	*P* value
				Non-obese			Obese			All	
rs366193	*ATP6V0A4*	4.3×10^{-6}	0.34	1240/1416	0.94 (0.86–1.04)	2.4×10^{-1}	373/221	1.68 (1.36–2.07)	1.1×10^{-6}	1.05 (0.96–1.14)	3.2×10^{-1}
rs13070718	*RAB5A*	1.2×10^{-5}	0.35	459/424	1.23 (1.06–1.41)	4.8×10^{-3}	94/118	0.61 (0.45–0.81)	7.2×10^{-4}	1.07 (0.95–1.22)	2.7×10^{-1}
rs7043149	*TJP2*	1.3×10^{-5}	0.35	1358/1613	0.88 (0.79–0.96)	6.7×10^{-3}	399/268	1.41 (1.15–1.73)	8.8×10^{-4}	0.95 (0.88–1.04)	2.9×10^{-1}
rs9314865	*TJP2*	3.2×10^{-5}	0.47	1476/1510	1.14 (1.04–1.26)	6.6×10^{-3}	351/341	0.73 (0.60–0.89)	2.4×10^{-3}	1.05 (0.97–1.15)	2.5×10^{-1}
rs2309424	*TJP2*	3.3×10^{-5}	0.47	1476/1511	1.14 (1.04–1.26)	6.6×10^{-3}	351/341	0.73 (0.60–0.89)	2.4×10^{-3}	1.05 (0.97–1.15)	2.5×10^{-1}
rs4329331	*TJP2*	3.5×10^{-5}	0.47	1143/1335	0.91 (0.83–1.01)	7.4×10^{-2}	335/215	1.46 (1.18–1.8)	5.1×10^{-4}	0.99 (0.91–1.09)	8.6×10^{-1}
rs4338173	*TJP2*	4.5×10^{-5}	0.51	1142/1331	0.92 (0.83–1.01)	8.3×10^{-2}	335/215	1.46 (1.18–1.8)	5.1×10^{-4}	0.99 (0.91–1.09)	9.0×10^{-1}
				Non-diabetic			Diabetic			All	
rs10818684	*PTGS1*	7.9×10^{-7}	0.06	869/962	1.09 (0.98–1.21)	1.2×10^{-1}	174/136	0.53 (0.40–0.69)	2.3×10^{-6}	0.99 (0.89–1.09)	8.0×10^{-1}
rs10179599	*RPL31*	6.6×10^{-6}	0.26	667/722	1.11 (0.99–1.25)	7.5×10^{-2}	133/107	0.54 (0.40–0.72)	2.0×10^{-5}	1.01 (0.90–1.12)	8.8×10^{-1}
rs2872220	*WEE1*	1.0×10^{-5}	0.26	407/549	0.86 (0.75–0.98)	2.7×10^{-2}	130/32	2.19 (1.46–3.30)	1.2×10^{-5}	0.96 (0.84–1.09)	5.1×10^{-1}
rs2278725 [d]	*RPL31, NPAS2*	1.9×10^{-5}	0.33	821/874	1.14 (1.03–1.28)	1.6×10^{-2}	168/121	0.60 (0.46–0.79)	2.3×10^{-4}	1.05 (0.95–1.16)	3.6×10^{-1}
rs1771792	*OR1J1*	2.2×10^{-5}	0.33	397/396	1.21 (1.04–1.40)	1.3×10^{-2}	71/63	0.51 (0.36–0.74)	2.5×10^{-4}	1.07 (0.93–1.23)	3.2×10^{-1}
rs13017465	*RPL31, NPAS2*	2.5×10^{-5}	0.33	821/874	1.14 (1.03–1.28)	1.6×10^{-2}	168/120	0.61 (0.46–0.8)	3.0×10^{-4}	1.05 (0.95–1.16)	3.5×10^{-1}

[a] Obtained 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.

[b] Minor allele counts in cases and controls.

[c] Odds ratio (95% confidence interval) not adjusted for covatiates and *P* values from Chi-square test.

[d] One SNP may be assigned to two genes because SNPs located 20Kb up- or down-stream of the gene region were included for each gene.