JNCI J Natl Cancer Inst (2020) 112(10): djz247

doi: 10.1093/jnci/djz247
First published online April 5, 2020
Editorial

EDITORIAL

Identifying Novel Genetic Markers Through a Transcription-Wide Association Study: Can This Be a Path to Reducing the Burden of Pancreatic Cancer?

Jeanine M. Genkinger PhD,[1,2,*] Gloria H. Su PhD,[2,3,4] Regina M. Santella PhD[2,5]

[1]Department of Epidemiology, Mailman School of Public Health, 722 w 168th St, Rm 712, Columbia University, New York, NY 10032, USA; [2]Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, 1130 St Nicholas Ave, New York, NY 10032, USA; [3]Department of Pathology and Cell Biology, Columbia University Medical Center, 1130 St. Nicolas Avenue, ICRC 10-04, New York, NY 10032, USA; [4]Department of Otolaryngology and Head and Neck Surgery, Columbia University Medical Center, 180 Fort Washington Avenue, 7th floor, New York, NY 10032, USA and [5]Department of Environmental Health Sciences, Mailman School of Public Health, 630 West 168th St Room 16-410, Columbia University, New York, NY 10032, USA

*Correspondence to: Jeanine M. Genkinger, PhD, MHS, Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 w 168th St, Rm 712, New York, NY 10032, USA (e-mail: jg3081@columbia.edu).

Pancreatic cancer is one of the most lethal malignancies worldwide, and rates of occurrence and mortality are similar; in 2018, an estimated 458 918 cases of pancreatic cancer occurred globally and 432 242 died from the disease (1). Of those diagnosed with pancreatic cancer, half will die within 6 months, and the 5-year survival for pancreatic cancer is a dismal 8% (2). Another alarming fact is that incidence rates have increased by 1.5% per year since 2004 in the United States (3). Several factors contribute to this high fatality rate, including few known modifiable risk factors, no effective screening tools, and lack of early diagnostic symptoms unique to pancreatic cancer. The exact factors driving this increasing trend in incidence have yet to be elucidated.

Familial and genetic risk are some of the strongest identified risk factors for pancreatic cancer incidence. Family history of pancreatic cancer results in pancreatic cancer risk of 1.68 (eg, because of any relative diagnosed with pancreatic cancer) to up to fivefold (eg, because of having an affected sibling), depending on the relationship of the familial risk (4,5). Increased risks are also observed when examining the number of first-degree relatives affected with pancreatic cancer (summary relative ratio [RR] = 4.6, 95% confidence interval [CI] = 0.5 to 16.4; summary RR = 6.4, 95% CI = 1.8 to 16.4; summary RR = 32.0, 95% CI = 10.2 to 74.7, for one, two, or three relatives affected, respectively) (6). In addition, genetic syndromes are suggested to confer 4–40% increased pancreatic cancer risk (eg, familial atypical multiple mole melanoma). However, only 10–15% of all pancreatic cancers are considered familial and/or because of a genetic predisposition (7).

To date, a number of genes have been implicated in pancreatic cancer risk (eg, *BRCA1*, *BRCA2*, *PALB2*, *ATM*, *CDKN2A*, *APC*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *PRSS1*, and *STK11*), as well as the ABO blood type (7). A recent meta-analysis conducted with the largest pancreatic cancer genome-wide association studies (GWAS) that included up to 11 537 cases and 17 107 controls observed several new genome-wide significant loci (eg, single-nucleotide polymorphisms [SNPs] located on the *NOC2L*) (8). Most genetic susceptibility loci have been identified through GWAS approaches; although GWAS have provided critical information regarding low penetrance genetic risk, there are some inherent limitations of a GWAS-only approach. For example, prior research has demonstrated that GWAS only identifies a modest proportion of the variance in disease risk (9). These findings suggest that undiscovered loci, in addition to important lifestyle and environmental factors and variations in host microbiota, relevant for pancreatic carcinogenesis still exist. Further, the functionality of these loci identified through GWAS with regards to disease traits is unclear; GWAS often identify multiple SNPs within an associated loci that are in linkage disequilibrium through use of tagging SNPs and does not necessarily isolate the single (or groups of) loci that mechanistically are relevant to pancreatic carcinogenesis. Use of Transcription-Wide Association Study (TWAS) approaches provides a clear advantage over GWAS-only approaches. With recent advances in high-throughput techniques, TWAS has the ability to extend the knowledge that we have gained from GWAS to identify novel and support putative risk variants associated with pancreatic cancer that alter gene expression and subsequently modify protein abundance or function (10). Prior research has demonstrated that TWAS has been able to identify numerous susceptibility genes across complex traits that were not identified in GWAS (11).

In their work, Zhong et al. (12) conducted a comprehensive and robust TWAS of pancreatic cancer using data from two gene expression quantitative trait loci datasets on gene expression in normal pancreatic tissues of Europeans. The Laboratory of Translational Genomics (LTG) data are from "normal" tumor adjacent tissues and the Genotype Tissue Expression (GTEx)

Received: December 16, 2019; **Accepted:** January 3, 2020

data are from tissues from cancer-free autopsies. GWAS data were from two large-scale consortia: the Pancreatic Cancer Cohort Consortium and the Pancreatic Cancer Case-Control Consortium. Two algorithms (FUSION and MetaXan) were used to construct genetically regulated expression models using the genotyping data and RNA-seq transcriptome data. The authors identified 25 genes with genetically predicted expression that was associated with pancreatic adenocarcinoma risk. Of these 25 genes, 14 were located at 11 novel loci. Twelve of these genes (three at novel loci) remained statistically significant after Bonferroni correction. However, after conditioning on lead GWAS variants, two of the gene signals appear to mostly be explained by lead GWAS variants at chr1p36.12/CELA3B and chr9q31.1/SMC2. Variability in loci was noted across datasets and algorithms. Yet the authors confirmed a number of putative loci and observed high correlation in their findings across datasets and algorithms. They tested the robustness of their findings by employing gene expression data from three different sources (LTG, GTEx, and the combination LTG + GTEx).

These results varied greatly from a prior TWAS, conducted by Gong et al. (13), using data from a GWAS (14) and gene expression data from ONCOMINE (15) using the FUSION algorithm. Gong et al. (13) identified 19 genes; yet none of the genes identified in their study overlapped with Zhong et al. (12), nor did any of their genes identified in their study (13) reach statistical significance when tested against a Bonferroni corrected *P* value. A number of factors could have contributed to this heterogeneity in results; key differences existed between the populations and approaches employed between these two publications. Gong et al. (13) performed their TWAS analysis using a smaller sample size that may have limited power, used a different gene expression dataset, used only the FUSION algorithm, and included multiple ancestries. The contrast in results (statistically significant loci) between these two studies suggests that employment of multiple gene expression datasets, large sample sizes, and population stratification are critical components of TWAS. As the sample sizes for both the GWAS and TWAS increase, statistical power to detect smaller effect sizes improve.

Overall, this comprehensive and innovative study by Zhong et al. (12) identified candidate risk genes for pancreatic carcinogenesis at novel and previously reported loci through multiple algorithms and gene expression datasets. Although some of the 12 genes identified (eg, *TERT*, *PDX1*) have known functional impacts on pancreatic tumorigenesis, other genes require further interrogations to delineate their roles in pancreatic development and progression. If confirmed in other studies, each of these novel loci identified may better elucidate pancreatic cancer etiology and provide new targets for prevention, early detection, and treatment—something that is greatly needed for this highly fatal disease with no currently effective screening tools or curative treatment protocols.

Notes

The authors have no conflicts of interest to disclose.

References

1. Ferlay J, Soerjomataram I, Ervik M, et al. Globocan 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11; 2013 http://globocan.iarc.fr. Accessed December 2019.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin*. 2017;67(1):7–30.
3. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin*. 2012;62(1):10–29.
4. Molina-Montes E, Gomez-Rubio P, Márquez M, et al. Risk of pancreatic cancer associated with family history of cancer and other medical conditions by accounting for smoking among relatives. *Int J Epidemiol*. 2018;47(2):473–483.
5. Fehringer G, Gallinger S, Borgida A, et al. The association of family history of cancer and medical history with pancreatic cancer risk. *Pancreas*. 2014;43(5):812–814.
6. Permuth-Wey J, Egan KM. Family history is a significant risk factor for pancreatic cancer: results from a systematic review and meta-analysis. *Fam Cancer*. 2009;8(2):109–117.
7. Klein AP. Genetic susceptibility to pancreatic cancer. *Mol Carcinog*. 2012;51(1):14–24.
8. Klein AP, Wolpin BM, Risch HA, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat Commun*. 2018;9(1):556.
9. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747–753.
10. Wu C, Pan W. Integrating eQTL data with GWAS summary statistics in pathway-based analysis with application to schizophrenia. *Genet Epidemiol*. 2018;42(3):303–316.
11. Pavlides JM, Zhu Z, Gratten J, McRae AF, Wray NR, Yang J. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med*. 2016;8(1):84.
12. Zhong J, Jermusyk A, Wu L, et al. A transcriptome-wide association study (TWAS) identifies novel candidate susceptibility genes for pancreatic cancer. *J Natl Cancer Inst*. 2020;112(10):djz246.
13. Gong L, Zhang D, Lei Y, Qian Y, Tan X, Han S. Transcriptome-wide association study identifies multiple genes and pathways associated with pancreatic cancer. *Cancer Med*. 2018;7(11):5727–5732.
14. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet*. 2009;41(9):986–990.
15. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. 2007;9(2):166–180.