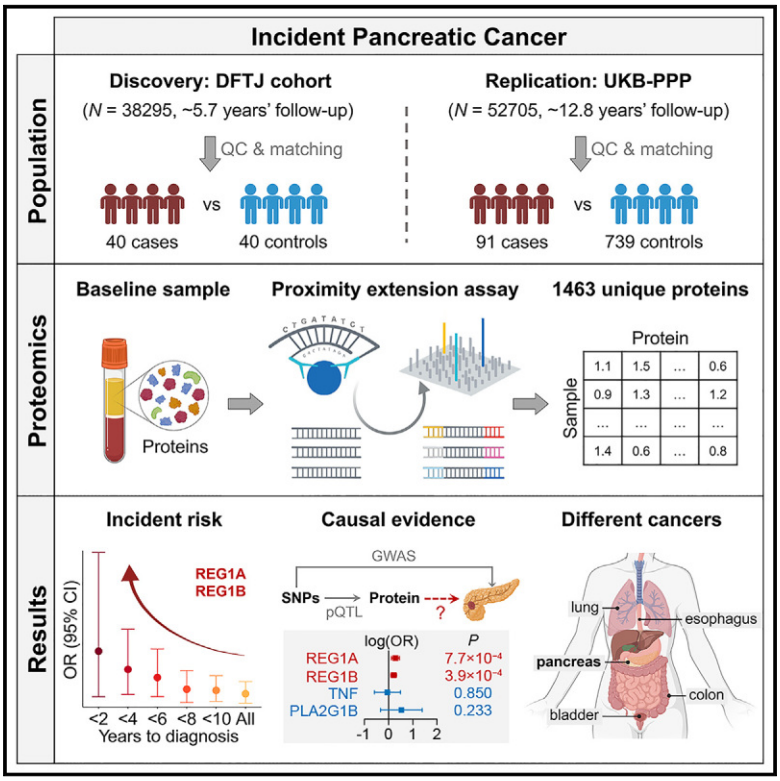
Cell Genomics

Short Article

Identification of biomarkers and potential therapeutic targets for pancreatic cancer by proteomic analysis in two prospective cohorts

Graphical abstract



Highlights

- Prospective analysis of 1,463 circulating proteins on pancreatic cancer (PC) risk
- REG1A and REG1B are associated with the risk of incident PC in two cohorts
- Mendelian randomization supports potential causality of REG1A and REG1B on PC

Authors

Jingjing Lyu, Minghui Jiang, Ziwei Zhu, ..., Tangchun Wu, Jiang Chang, Chaolong Wang

Correspondence

changjiang815@hust.edu.cn (J.C.), chaolong@hust.edu.cn (C.W.), wut@tjmu.edu.cn (T.W.)

In brief

Lyu et al. identified circulating REG1A and REG1B proteins to associate with the incident risk of pancreatic cancer in two prospective cohorts from China and the UK. With additional causal evidence from Mendelian randomization, this study highlights REG1A and REG1B as promising biomarkers and potential therapeutic targets for pancreatic cancer.

Lyu et al., 2024, Cell Genomics 4, 100561
June 12, 2024
https://doi.org/10.1016/j.xgen.2024.100561

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Short Article

Identification of biomarkers and potential therapeutic targets for pancreatic cancer by proteomic analysis in two prospective cohorts

Jingjing Lyu,[1,2] Minghui Jiang,[1,2] Ziwei Zhu,[1,2] Hongji Wu,[1,2] Haonan Kang,[1,2] Xingjie Hao,[1,2] Shanshan Cheng,[1,2] Huan Guo,[1,3] Xia Shen,[4] Tangchun Wu,[1,3,*] Jiang Chang,[1,5,*] and Chaolong Wang[1,2,6,*]

[1]Ministry of Education Key Laboratory of Environment and Health, State Key Laboratory of Environmental Health (Incubating), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Occupational and Environmental Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[4]Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China

[5]Department of Health Toxicology, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[6]Lead contact

*Correspondence: changjiang815@hust.edu.cn (J.C.), chaolong@hust.edu.cn (C.W.), wut@tjmu.edu.cn (T.W.)

https://doi.org/10.1016/j.xgen.2024.100561

SUMMARY

Pancreatic cancer (PC) is the deadliest malignancy due to late diagnosis. Aberrant alterations in the blood proteome might serve as biomarkers to facilitate early detection of PC. We designed a nested case-control study of incident PC based on a prospective cohort of 38,295 elderly Chinese participants with ~5.7 years' follow-up. Forty matched case-control pairs passed the quality controls for the proximity extension assay of 1,463 serum proteins. With a lenient threshold of $p < 0.005$, we discovered regenerating family member 1A (REG1A), REG1B, tumor necrosis factor (TNF), and phospholipase A2 group IB (PLA2G1B) in association with incident PC, among which the two REG1 proteins were replicated using the UK Biobank Pharma Proteomics Project, with effect sizes increasing steadily as diagnosis time approaches the baseline. Mendelian randomization analysis further supported the potential causal effects of REG1 proteins on PC. Taken together, circulating REG1A and REG1B are promising biomarkers and potential therapeutic targets for the early detection and prevention of PC.

INTRODUCTION

Pancreatic cancer (PC) ranks as the third leading cause of cancer-related mortality and is one of the deadliest malignancies with 5-year relative survival rates as low as 10%, partly due to its late diagnosis.[1,2] Early detection can significantly enhance survival rates, providing a 24%–37% increase compared to patients diagnosed at advanced stages.[1] Furthermore, the progression of PC is gradual, taking an average of 11.7 years from its initiation to invasive stages and thus providing a sufficient time window for early detection.[3] Therefore, early screening and intervention of high-risk individuals is important to improve the survival of PC.

The blood proteome is a composite of circulating proteins, including those secreted for normal physiological function and those leaked from damaged cells or tissues.[4] Abnormal changes in the blood proteome often signify the occurrence of lesions within the body, including the development of tumors. Traditional cancer biomarkers, such as carbohydrate antigen 19-9 (CA19-9),[5] CA125,[6] CA242,[7] and carcinoembryonic antigen,[8] have demonstrated varying levels of upregulation in patients with PC compared to controls. However, their specificity to PC remains relatively low,[5,9] and their effectiveness was only demonstrated in progressed cancers.[10] To fill the gap, several studies have endeavored to identify alternative protein biomarkers for the early diagnosis of PC.

Chronic inflammation is considered a key mediator in the development of PC.[11,12] Inflammatory proteins, such as C-reactive protein (CRP), interleukin-6 (IL-6), and tumor necrosis factor (TNF), were found to associate with PC in case-control studies,[13–17] but limited evidence has been provided by prospective studies.[18–21] Only one prospective cohort study found a significant association between CRP and incident PC after excluding patients with pancreatitis.[22] Recently, a Japanese cohort comprising 111 incident PC cases and 773 controls reported no significant association between 67 inflammatory proteins and the incidence of PC.[18] Similarly, in a Chinese cohort of 610 incident PC cases and 623 controls, only two out of 92

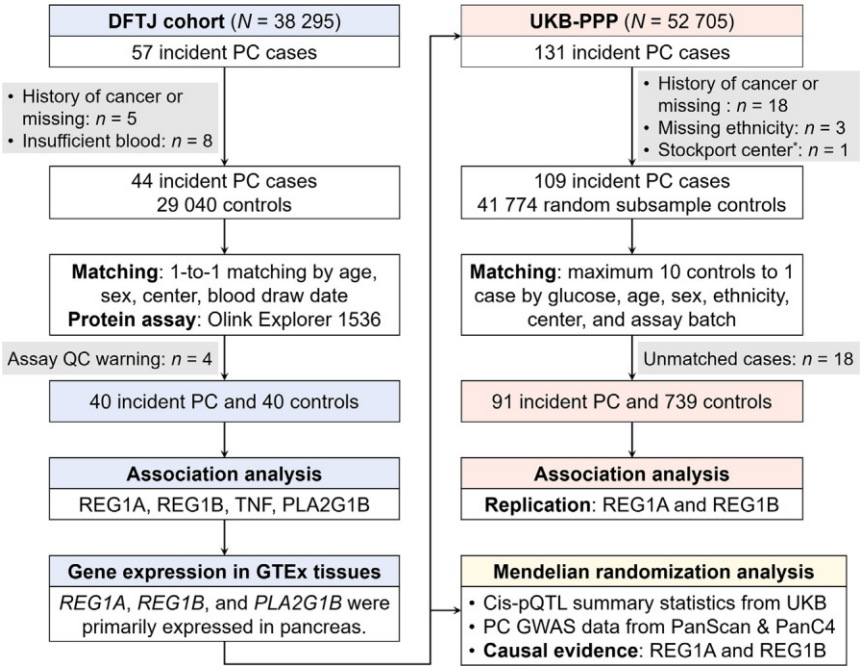


Figure 1. Flowchart of the study design

n, number of incident PC cases.

*Stockport center was excluded from our analysis because no glucose measurement was provided by this center.

inflammatory protein biomarkers (monocyte chemotactic protein 3 and angiopoietin-2) showed a significant association at a false discovery rate of 0.05.[23] Besides inflammatory proteins, several prospective studies focused on proteins related to the metabolic function of the pancreas, such as lipid metabolism and insulin secretion.[24–28] Among the limited number of pancreas-function-related proteins examined by these studies, adiponectin,[24] leptin,[25] and phospholipid fatty acids[26] were associated with the risk of PC, while the relationship of insulin-like growth factors (IGFs) and IGF-binding proteins to PC risk remained unclear.[27,28] Further validations were required to confirm these findings.

In this study, we aimed to identify serum protein biomarkers associated with incident PC in a prospective Chinese cohort. In the discovery phase, we used the Olink Explore 1536 panel to measure 1,463 unique serum proteins in 44 pairs of incident PC cases and controls nested in the Dongfeng-Tongji (DFTJ) cohort.[29] Replication analysis was conducted for the top protein biomarkers using the plasma proteomics dataset from the UK Biobank Pharma Proteomics Project (UKB-PPP).[30] Finally, for further validation, we evaluated the potential causal effects of the proteins on PC via two-sample Mendelian randomization (MR) analysis by leveraging summary statistics from a large-scale protein quantitative trait locus (pQTL) study[30] and genome-wide association studies (GWASs) of PC in European populations.[31–34]

RESULTS

Associations between serum proteins and incident PC in the DFTJ cohort

We included 44 incident cases of PC and designed a nested case-control study by matching one cancer-free control to each case based on sex, age, hospital, and blood draw date in the DFTJ cohort[29] (Figure 1). After quality controls (STAR Methods), 40 case-control pairs were included in the downstream analyses. The mean (SD) age was 68.48 (8.18) years and 45% were males among the PC cases (Table S1). Compared to controls, PC cases had higher levels of fasting blood glucose ($p = 0.022$, paired t test) and hemoglobin A1c (HbA1c) levels ($p = 0.023$, paired t test) and a higher prevalence of type 2 diabetes (T2D; $p = 0.014$, Cochran-Mantel-Haenszel test). No significant differences were detected in the education level, BMI, smoking status, and drinking status between PC cases and their matched controls.

Among 1,463 serum proteins included in the Olink Explore 1536 panel, several associations with incident PC risk were unveiled at a lenient threshold of $p < 0.005$ (conditional logistic regression, Wald test), including two regenerating family member 1 (REG1) proteins (REG1A, odds ratio [OR] = 2.86 [95% confidence interval (CI): 1.40, 5.85]; REG1B, OR = 2.58 [1.33, 5.03]), TNF (OR = 4.88 [1.80, 13.23]), and phospholipase A2 group IB (PLA2G1B; OR = 0.38 [0.19, 0.74]) (Figure 2; Table S2). These associations were robust when additionally adjusting for T2D (Figure S1), and there was no discernible heterogeneity between males and females (Table S3). *REG1A* and *REG1B* genes are clustered on chromosome 2p12, and their protein expression levels were highly correlated (Spearman's correlation $r_s = 0.93$). By examining gene expression profiles across 54 tissues in the GTEx Project,[35] we found that *REG1A*, *REG1B*, and *PLA2G1B* were primarily expressed in the pancreas (Figure S2). Furthermore, when we stratified cases into two groups by every 3 years of their follow-up time to PC diagnosis, we found that REG1A and REG1B were significantly overexpressed in cases diagnosed 3–6 years from the baseline ($p < 0.05$, t test, Figure S3). In contrast, the expression levels of TNF and PLA2G1B were significantly different between cases and controls only when cases were diagnosed within 3 years from the baseline.

Replication of candidate proteins in the UKB-PPP dataset

The replication samples were selected from 52,705 baseline samples measured on the same Olink Explore 1536 panel in the UKB-PPP.[30] After preprocessing and quality controls, we performed matching between 109 incident PCs and 41,774 random baseline controls based on blood glucose, age, sex, ethnicity, center, and batch of protein assay. After removing 18 cases without matched controls, the replication analyses were

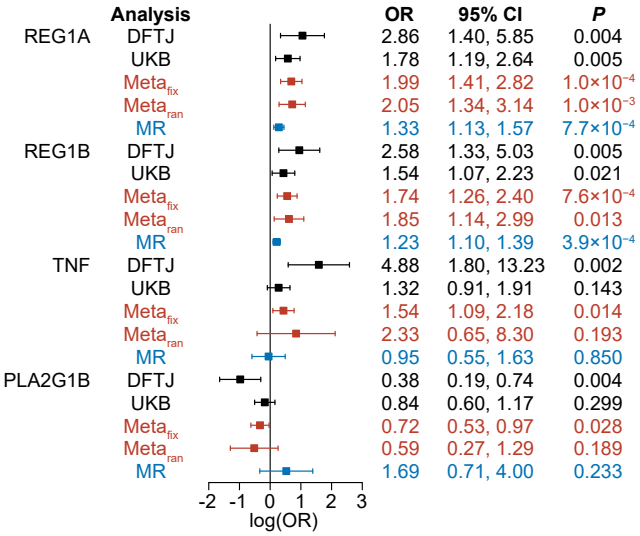


Figure 2. The estimated effects of REG1A, REG1B, TNF, and PLA2G1B on PC risks

OR indicates the odds ratio per 1-SD increase in the protein expression. Log (OR) indicates the natural logarithm of OR with a 95% confidence interval. Analysis of UKB data was restricted to cases diagnosed within 6 years of follow-up. Results of the meta-analysis ($Meta_{fix}$ and $Meta_{ran}$) and the Mendelian randomization (MR) are shown in red and blue text, respectively. $Meta_{fix}$, fixed-effect meta-analysis; $Meta_{ran}$, random-effect meta-analysis. *p* values were derived from conditional logistic regression for DFTJ and UKB, from inverse variance weighted method for $Meta_{fix}$, $Meta_{ran}$, and MR.

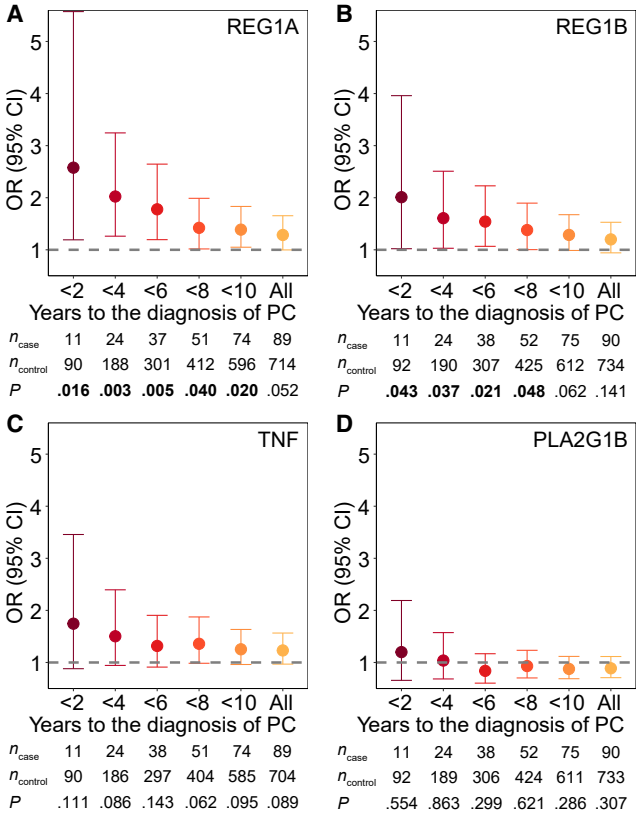


Figure 3. Association between plasma proteins and incident PC risks in the UKB-PPP

(A) REG1A.
(B) REG1B.
(C) TNF.
(D) PLA2G1B.

The mean times between baseline and diagnosis of PC in each group were 1.0, 2.1, 3.2, 4.2, 5.7, and 6.5 years, respectively. Note that the samples in each group were cumulative, with the last group including all cases and matched controls. n_{case}, number of PC cases. $n_{control}$, number of matched controls. *p* values were derived from conditional logistic regression. OR indicates the odds ratio per 1-SD increase in the protein expression.

based on 91 incident PCs and 739 matched controls, whose baseline characteristics were presented in Table S4.

When including all 91 matched strata in the association analysis, we found no significant protein, potentially due to the wide time interval between the baseline and diagnosis (Figure 3; Table S5). Thus, we performed subgroup analyses by the time interval to the diagnosis of PC. The subsets of strata included cases diagnosed within 2, 4, 6, 8, and 10 years from the baseline, respectively. Remarkably, ORs of REG1A, REG1B, and TNF showed a clear increasing trend when the time to diagnosis was shorter, although no significant association was found for TNF (Figure 3). When we restricted the analysis to 38 incident cases within 6 years of follow-up (like that of the DFTJ cohort), REG1A and REG1B showed significant associations with PC risk (OR = 1.78 [1.19, 2.64] for REG1A; OR = 1.54 [1.07, 2.23] for REG1B) (Figure 2). The results were similar when additionally adjusting for blood glucose (Figure S1).

Compared to the estimates in DFTJ, the effect sizes of the two REG1 proteins in the UKB-PPP were smaller but not significantly different ($p_{het} \approx 0.2$, Q test, Table S6). Combining evidence from DFTJ and the UKB-PPP with fixed-effect meta-analysis, one SD increase in the expression level of REG1A (OR = 1.99 [1.41, 2.82]) or REG1B (OR = 1.74 [1.26, 2.40]) could almost double the risk of incident PC (Figure 2).

MR analysis between candidate proteins and PC

We performed MR analysis to evaluate the causal effects of candidate proteins on PC risk, based on published PC GWAS datasets (Table S7)[31–34] and pQTL summary statistics from the UKB-PPP.[30] For the PC GWAS, we performed quality controls, imputation, association, and meta-analysis on the individual-level data to generate summary statistics (STAR Methods; Figure S4). We selected 12, 11, 4, and 1 *cis*-pQTLs as the instrumental variables (IVs) for REG1A, REG1B, TNF, and PLA2G1B, respectively (STAR Methods). All IVs have *F* statistics greater than 10, indicating little weak instrumental bias (Table S8). No heterogeneity or pleiotropy was detected among IVs for REG1A and REG1B (Figure S5). Intriguingly, we found significant causal effects on PC for REG1A (OR = 1.33 [1.13, 1.57]) and REG1B (OR = 1.23 [1.10, 1.39]) but no evidence for TNF (OR = 0.95 [0.55, 1.63]) or PLA2G1B (OR = 1.69 [0.71, 4.00]) (Figure 2). Furthermore, causal effect estimates for REG1A and REG1B were largely unchanged when we additionally adjusted for T2D using multivariable MR (Figure S6).

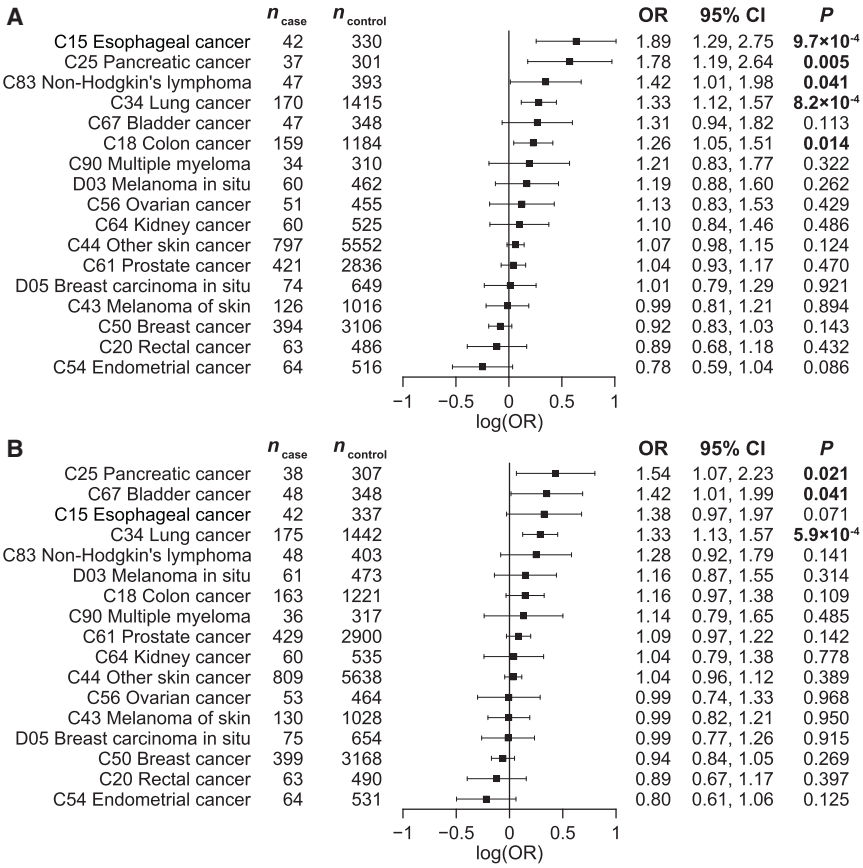


Figure 4. Association of REG1 proteins with different types of incident cancer within 6 years of follow-up

(A) REG1A.

(B) REG1B.

OR indicates the odds ratio per 1-SD increase in the protein expression. Log (OR) indicates the natural logarithm of OR with a 95% confidence interval. n_{case}, number of incident cases; $n_{control}$, number of matched controls. *p* values were derived from conditional logistic regression.

For two REG1 proteins with causal evidence, we conducted colocalization analysis to assess IV assumptions. Bayesian colocalization analysis based on *coloc*[36] yielded moderate evidence to support that REG1 proteins and PC shared the same causal variant (PP_{H4} = 0.483). More importantly, there was no evidence suggesting that MR analyses of REG1 proteins and PC were confounded by different causal variants in LD (PP_{H3} = 0.033) (Figure S7). In addition, we performed the conditional test of proportional colocalization,[37] which did not reject colocalization for either REG1A (*p* = 0.34) or REG1B (*p* = 0.78) (Figure S8). The proportionality constant was significantly different from zero for REG1B ($p = 3.0 \times 10^{-4}$, Lagrange multiplier test) but not for REG1A (*p* = 0.17, Lagrange multiplier test). Overall, there was evidence supporting colocalization for both REG1 proteins, with stronger evidence favoring REG1B.

Association between REG1 proteins and other cancers

We compared the expression of REG1A and REG1B between incident cases and matched controls for 17 cancers with at least 100 incident cases during the follow-up, based on the UKB-PPP (Figure 4). Using the same analysis strategy as applied to PC, we found that in addition to PC, both REG1A (OR = 1.33 [1.12, 1.57]) and REG1B (OR = 1.33 [1.13, 1.57]) were significantly elevated in incident lung cancer, for which ORs were smaller than those of PC but the *p* values were more significant due to the larger sample size. REG1A was also elevated in incident esophageal cancer (OR = 1.89 [1.29, 2.75]), with an effect size comparable to that of PC (Figure 4).

DISCUSSION

In this study, we investigated prospective associations of 1,463 serum protein biomarkers with PC risk based on incident case-control samples nested in an elderly Chinese cohort. We identified REG1A, REG1B, TNF, and PLA2G1B to associate with incident PC in the DFTJ cohort and successfully replicated REG1A and REG1B in the UKB-PPP. Moreover, MR analyses based on independent large-scale pQTL and GWAS datasets further supported the causal effects of REG1A and REG1B on PC.

REG1A and *REG1B* are members of the REG1 family tandemly clustered on chromosome 2p12. Because REG1 proteins are synthesized in the islet β cells[38,39] and are involved in islet cell regeneration and diabetogenesis,[40,41] we hypothesized that pancreatic lesions causing islet cell damage would stimulate the cytothesis and proliferation of islet β cells and thus activate the secretion of REG1 proteins. Furthermore, REG1A and REG1B contain the c-type lectin-like domain,[42] which can promote tumor growth and malignancy by binding to the carbohydrates on the surface of tumor cells.[43] In fact, both mRNA and proteins of *REG1* genes have been shown to overexpress in cases of many cancers, especially in the human digestive system,[38,44,45] highlighting the diagnostic and prognostic value of *REG1* genes. However, no study has investigated the predictive value of REG1 proteins for incident PC. Our study, for the first time, showed that REG1A and REG1B proteins were elevated in the blood several years before the incidence of PC, even after adjustment of the status of T2D, suggesting that REG1 proteins can signal the onset of PC before it progresses to an invasive stage. This early signaling allows for the timely use of computed tomography (CT) scans and biopsies to confirm the diagnosis and subsequently facilitates early intervention to enhance the survival of PC. Moreover, causal evidence from MR suggested that REG1A and REG1B were involved in the development of PC and thus might serve as potential therapeutic targets for the prevention of PC. Nevertheless, their nearby genomic locations and high correlation in the protein expression levels make it

difficult to distinguish which of the two, REG1A or REG1B, is causal for PC, while the conditional test of proportional colocalization slightly favors REG1B over REG1A. In addition, it is worth noting that circulating REG1 proteins were also elevated in the incident cases of esophageal and lung cancers in the UKB-PPP, despite their low RNA expression in normal tissues of esophageal and lung (Figure S2). The mechanism of how circulating REG1A and REG1B contribute to these cancers remains unclear, while it has been suggested that REG1 proteins might induce the expression of IL-6 to exert effects on esophageal cancer cell biology.[46]

Although the associations of TNF and PLA2G1B with incident PC were not replicated in the UKB-PPP, their biological functions were relevant to PC. TNF is an important pro-inflammatory cytokine involved in cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation.[47] In progressed PC, TNF could stimulate tumor cell growth by upregulating the expression of the epidermal growth factor receptor[48] and increase pro-inflammatory signaling by upregulating the expression of *CXCL1* in tumor cells.[49] Nonetheless, there is limited evidence for TNF on incident PC from prospective studies.[23] PLA2G1B is a secreted member of the phospholipase A2 enzymes produced by pancreatic acinar cells[50] and has been shown to be downregulated in pancreatic tumors.[51] PLA2G1B displays enzymatic activity after meals and is highly correlated with body weight gain.[52] When *PLA2G1B* was inhibited, mice were protected from obesity induced by a high-fat or high-carbohydrate diet.[52,53] Consistent with the downregulation of *PLA2G1B*, weight loss is an important distinction between diabetes mellitus caused by PC and T2D.[54] Considering that the UKB-PPP only has 38 incident PC cases within 6 years of follow-up, future replication efforts for TNF and PLA2G1B are warranted.

This study has several strengths. Firstly, we utilized two prospective cohorts, which allowed for the establishment and subsequent replication of the association between circulating proteins and the incidence of PC. The prospective design, which collects data prior to the development of PC, is crucial to strengthen the validity and reliability of our findings. Secondly, with the ability to predict PC incidence several years prior to diagnosis, our identified protein biomarkers showed significant clinical importance in improving PC screening efforts. Thirdly, moving beyond mere associations, the inclusion of MR analysis strengthens our study by delving into the causal relationship between the identified proteins and PC risk, thereby laying a robust foundation for considering these proteins as potential intervention and therapeutic targets.

In summary, this study contributes important findings to the field of PC research. By utilizing prospective cohorts, we have identified circulating protein biomarkers that can predict PC incidence years before diagnosis. The replication of our findings and the use of MR analysis further support the association and suggest potential causality between REG1 proteins and PC risk. These findings have important implications for PC screening and highlight the potential of these proteins as early diagnostic biomarkers and therapeutic targets. Future investigations with larger sample sizes and comprehensive protein profiling are warranted to validate and expand upon our findings.

Limitations of the study

First, because of the low incidence of PC and the prospective design of our study, we have limited numbers of incident cases in both discovery and replication datasets. Therefore, we made a conscious decision to focus on the identification of protein biomarkers that can aid in early detection of PC rather than on risk prediction models. Because of the small sample size, we selected the top-ranked association signals with a lenient threshold of $p < 0.005$ in the discovery stage without strictly controlling for multiple testing of 1,463 proteins. Despite this limitation, our findings of REG1A and REG1B were supported by independent replication analysis and MR analysis. Second, we did not match cases and controls by T2D, despite T2D being a known risk factor of PC and its prevalence being different between cases and controls, because the relationship between T2D, circulating proteins, and PC might be complicated.[55] Matching on T2D might obscure the true effect of a protein on PC when either T2D or the protein is a mediator on the causal pathway. Instead, we performed sensitivity analysis adjusting for T2D, and the results remained consistent. Third, we have observed some heterogeneity between the results from DFTJ and the UKB-PPP, which might be attributed to the difference in sample type (serum versus plasma), sample status (fasting versus non-fasting), or ethnicity (Chinese versus European). Fourth, our choice of serum sample in the discovery stage might limit our ability to identify the potential role of clotting factors in PC. But compared to plasma, serum samples also have several advantages, including less cell contamination, better stability for some measurands, and the absence of anticoagulants.[56] Finally, the overexpression of REG1 proteins in incident esophageal and lung cancers highlights an important issue of specificity when utilizing REG1 proteins as PC biomarkers. In general, the screening specificity can be improved by a combination of multiple biomarkers, including those from gastroscopy or CT scans.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Study design
  - Measurement of proteins
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis of the DFTJ data
  - Replication using the UKB-PPP dataset
  - Mendelian randomization analysis
  - Colocalization analysis
  - Association of protein biomarkers with different cancers

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2024.100561.

ACKNOWLEDGMENTS

We thank Dr. Jian Yang from Westlake University, China, for constructive comments. Part of the graphical abstract was created with BioRender (https://biorender.com/). This work was funded by the Natural Science Foundation of China (82325044, 82021005, and 82192903), the National Key R&D Program of China (2022YFF1203300), the Natural Science Fund for Distinguished Young Scholars of Hubei Province (2022CFA046 and 2020CFA067), and the Fundamental Research Funds for the Central Universities (HUST: 2019kfyXJJS036 and 2023BR030).

AUTHOR CONTRIBUTIONS

C.W., J.C., and T.W. conceived and supervised the study. J.L. performed experiments. J.L., M.J., Z.Z., H.W., H.K., X.H., and S.C. cleaned and analyzed data. J.L. drafted the manuscript with input from C.W., J.C., X.S., and H.G. All authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 4, 2023
Revised: December 12, 2023
Accepted: April 21, 2024
Published: May 15, 2024

REFERENCES

1. Siegel, R.L., Giaquinto, A.N., and Jemal, A. (2024). Cancer statistics, 2024. CA. Cancer J. Clin. *74*, 12–49.
2. Mizrahi, J.D., Surana, R., Valle, J.W., and Shroff, R.T. (2020). Pancreatic cancer. Lancet *395*, 2008–2020.
3. Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R.H., Eshleman, J.R., Nowak, M.A., et al. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature *467*, 1114–1117.
4. Suhre, K., McCarthy, M.I., and Schwenk, J.M. (2021). Genetics meets proteomics: perspectives for large population-based studies. Nat. Rev. Genet. *22*, 19–37.
5. Winter, J.M., Yeo, C.J., and Brody, J.R. (2013). Diagnostic, prognostic, and predictive biomarkers in pancreatic cancer. J. Surg. Oncol. *107*, 15–22.
6. Meng, Q., Shi, S., Liang, C., Xiang, J., Liang, D., Zhang, B., Qin, Y., Ji, S., Xu, W., Xu, J., et al. (2017). Diagnostic accuracy of a CA125-based biomarker panel in patients with pancreatic cancer: a systematic review and meta-analysis. J. Cancer *8*, 3615–3622.
7. Zhang, Y., Yang, J., Li, H., Wu, Y., Zhang, H., and Chen, W. (2015). Tumor markers CA19-9, CA242 and CEA in the diagnosis of pancreatic cancer: a meta-analysis. Int. J. Clin. Exp. Med *8*, 11683–11691.
8. Meng, Q., Shi, S., Liang, C., Liang, D., Xu, W., Ji, S., Zhang, B., Ni, Q., Xu, J., and Yu, X. (2017). Diagnostic and prognostic value of carcinoembryonic antigen in pancreatic cancer: a systematic review and meta-analysis. Onco. Targets Ther *10*, 4591–4598.
9. Yu, J., Ploner, A., Kordes, M., Löhr, M., Nilsson, M., de Maturana, M.E.L., Estudillo, L., Renz, H., Carrato, A., Molero, X., et al. (2021). Plasma protein biomarkers for early detection of pancreatic ductal adenocarcinoma. Int. J. Cancer *148*, 2048–2058.
10. Kim, J.E., Lee, K.T., Lee, J.K., Paik, S.W., Rhee, J.C., and Choi, K.W. (2004). Clinical usefulness of carbohydrate antigen 19-9 as a screening test for pancreatic cancer in an asymptomatic population. J. Gastroenterol. Hepatol. *19*, 182–186.
11. Shadhu, K., and Xi, C. (2019). Inflammation and pancreatic cancer: an updated review. Saudi J. Gastroenterol. *25*, 3–13.
12. Zambirinis, C.P., Pushalkar, S., Saxena, D., and Miller, G. (2014). Pancreatic cancer, inflammation, and microbiome. Cancer J. *20*, 195–202.
13. Lindgaard, S.C., Sztupinszki, Z., Maag, E., Chen, I.M., Johansen, A.Z., Jensen, B.V., Bojesen, S.E., Nielsen, D.L., Hansen, C.P., Hasselby, J.P., et al. (2021). Circulating protein biomarkers for use in pancreatic ductal adenocarcinoma identification. Clin. Cancer Res. *27*, 2592–2603.
14. Mroczko, B., Groblewska, M., Gryko, M., Kędra, B., and Szmitkowski, M. (2010). Diagnostic usefulness of serum interleukin 6 (IL-6) and C-reactive protein (CRP) in the differentiation between pancreatic cancer and chronic pancreatitis. J. Clin. Lab. Anal. *24*, 256–261.
15. Błogowski, W., Deskur, A., Budkowska, M., Sałata, D., Madej-Michniewicz, A., Dąbkowski, K., Dołęgowska, B., and Starzyńska, T. (2014). Selected cytokines in patients with pancreatic cancer: a preliminary report. PLoS One *9*, e97613.
16. Shaw, V.E., Lane, B., Jenkinson, C., Cox, T., Greenhalf, W., Halloran, C.M., Tang, J., Sutton, R., Neoptolemos, J.P., and Costello, E. (2014). Serum cytokine biomarker panels for discriminating pancreatic cancer from benign pancreatic disease. Mol. Cancer *13*, 114.
17. Zhang, P., Zou, M., Wen, X., Gu, F., Li, J., Liu, G., Dong, J., Deng, X., Gao, J., Li, X., et al. (2014). Development of serum parameters panels for the early detection of pancreatic cancer. Int. J. Cancer *134*, 2646–2655.
18. Ma, E., Shimazu, T., Song, M., Charvat, H., Sawada, N., Yamaji, T., Inoue, M., Camargo, M.C., Kemp, T.J., Pfeiffer, R.M., et al. (2022). Circulating inflammation markers and pancreatic cancer risk: a prospective case-cohort study in Japan. Cancer Epidemiol. Biomarkers Prev. *31*, 236–241.
19. Grote, V.A., Kaaks, R., Nieters, A., Tjønneland, A., Halkjær, J., Overvad, K., Skjelbo Nielsen, M.R., Boutron-Ruault, M.C., Clavel-Chapelon, F., Racine, A., et al. (2012). Inflammation marker and risk of pancreatic cancer: a nested case–control study within the EPIC cohort. Br. J. Cancer *106*, 1866–1874.
20. Bao, Y., Giovannucci, E.L., Kraft, P., Qian, Z.R., Wu, C., Ogino, S., Gaziano, J.M., Stampfer, M.J., Ma, J., Buring, J.E., et al. (2013). Inflammatory plasma markers and pancreatic cancer risk: a prospective study of five US cohorts. Cancer Epidemiol. Biomarkers Prev. *22*, 855–861.
21. Douglas, J.B., Silverman, D.T., Weinstein, S.J., Graubard, B.I., Pollak, M.N., Tao, Y., Virtamo, J., Albanes, D., and Stolzenberg-Solomon, R.Z. (2011). Serum C-reactive protein and risk of pancreatic cancer in two nested, case-control studies. Cancer Epidemiol. Biomarkers Prev. *20*, 359–369.
22. Sollie, S., Michaud, D.S., Sarker, D., Karagiannis, S.N., Josephs, D.H., Hammar, N., Santaolalla, A., Walldius, G., Garmo, H., Holmberg, L., et al. (2019). Chronic inflammation markers are associated with risk of pancreatic cancer in the Swedish AMORIS cohort study. BMC Cancer *19*, 858.
23. Kartsonaki, C., Pang, Y., Millwood, I., Yang, L., Guo, Y., Walters, R., Lv, J., Hill, M., Yu, C., Chen, Y., et al. (2022). Circulating proteins and risk of pancreatic cancer: a case-subcohort study among Chinese adults. Int. J. Epidemiol. *51*, 817–829.
24. Bao, Y., Giovannucci, E.L., Kraft, P., Stampfer, M.J., Ogino, S., Ma, J., Buring, J.E., Sesso, H.D., Lee, I.M., Gaziano, J.M., et al. (2013). A prospective study of plasma adiponectin and pancreatic cancer risk in five US cohorts. J. Natl. Cancer Inst. *105*, 95–103.
25. Stolzenberg-Solomon, R.Z., Newton, C.C., Silverman, D.T., Pollak, M., Nogueira, L.M., Weinstein, S.J., Albanes, D., Männistö, S., and Jacobs, E.J. (2015). Circulating leptin and risk of pancreatic cancer: a pooled analysis from 3 cohorts. Am. J. Epidemiol. *182*, 187–197.
26. Matejcic, M., Lesueur, F., Biessy, C., Renault, A.L., Mebirouk, N., Yammine, S., Keski-Rahkonen, P., Li, K., Hémon, B., Weiderpass, E., et al. (2018). Circulating plasma phospholipid fatty acids and risk of pancreatic cancer in a large European cohort. Int. J. Cancer *143*, 2437–2448.
27. Wolpin, B.M., Michaud, D.S., Giovannucci, E.L., Schernhammer, E.S., Stampfer, M.J., Manson, J.E., Cochrane, B.B., Rohan, T.E., Ma, J., Pollak,

M.N., and Fuchs, C.S. (2007). Circulating insulin-like growth factor binding protein-1 and the risk of pancreatic cancer. Cancer Res. *67*, 7923–7928.

28. Wolpin, B.M., Michaud, D.S., Giovannucci, E.L., Schernhammer, E.S., Stampfer, M.J., Manson, J.E., Cochrane, B.B., Rohan, T.E., Ma, J., Pollak, M.N., and Fuchs, C.S. (2007). Circulating insulin-like growth factor axis and the risk of pancreatic cancer in four prospective cohorts. Br. J. Cancer *97*, 98–104.

29. Wang, F., Zhu, J., Yao, P., Li, X., He, M., Liu, Y., Yuan, J., Chen, W., Zhou, L., Min, X., et al. (2013). Cohort profile: the Dongfeng-Tongji cohort study of retired workers. Int. J. Epidemiol. *42*, 731–740.

30. Sun, B.B., Chiou, J., Traylor, M., Benner, C., Hsu, Y.-H., Richardson, T.G., Surendran, P., Mahajan, A., Robins, C., Vasquez-Grinnell, S.G., et al. (2023). Plasma proteomic associations with genetics and health in the UK Biobank. Nature *622*, 329–338.

31. Amundadottir, L., Kraft, P., Stolzenberg-Solomon, R.Z., Fuchs, C.S., Petersen, G.M., Arslan, A.A., Bueno-de-Mesquita, H.B., Gross, M., Helzlsouer, K., Jacobs, E.J., et al. (2009). Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat. Genet. *41*, 986–990.

32. Petersen, G.M., Amundadottir, L., Fuchs, C.S., Kraft, P., Stolzenberg-Solomon, R.Z., Jacobs, K.B., Arslan, A.A., Bueno-de-Mesquita, H.B., Gallinger, S., Gross, M., et al. (2010). A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat. Genet. *42*, 224–228.

33. Wolpin, B.M., Rizzato, C., Kraft, P., Kooperberg, C., Petersen, G.M., Wang, Z., Arslan, A.A., Beane-Freeman, L., Bracci, P.M., Buring, J., et al. (2014). Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat. Genet. *46*, 994–1000.

34. Childs, E.J., Mocci, E., Campa, D., Bracci, P.M., Gallinger, S., Goggins, M., Li, D., Neale, R.E., Olson, S.H., Scelo, G., et al. (2015). Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat. Genet. *47*, 911–916.

35. Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G.A., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., et al. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

36. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.

37. Patel, A., Whittaker, J.C., and Burgess, S. (2024). A frequentist test of proportional colocalization after selecting relevant genetic variants. Preprint at arXiv. https://doi.org/10.48550/arXiv.2402.12171.

38. Yuan, R.-H., Jeng, Y.-M., Chen, H.-L., Hsieh, F.-J., Yang, C.-Y., Lee, P.-H., and Hsu, H.-C. (2005). Opposite roles of human pancreatitis-associated protein and REG1A expression in hepatocellular carcinoma: association of pancreatitis-associated protein expression with low-stage hepatocellular carcinoma, β-catenin mutation, and favorable prognosis. Clin. Cancer Res. *11*, 2568–2575.

39. Terazono, K., Uchiyama, Y., Ide, M., Watanabe, T., Yonekura, H., Yamamoto, H., and Okamoto, H. (1990). Expression of reg protein in rat regenerating islets and its co-localization with insulin in the Beta cell secretory granules. Diabetologia *33*, 250–252.

40. Astorri, E., Guglielmi, C., Bombardieri, M., Alessandri, C., Buzzetti, R., Maggi, D., Valesini, G., Pitzalis, C., and Pozzilli, P. (2010). Circulating Reg1α proteins and autoantibodies to Reg1α proteins as biomarkers of β-cell regeneration and damage in type 1 diabetes. Horm. Metab. Res. *42*, 955–960.

41. Zhou, L., Zhang, R., Wang, L., Shen, S., Okamoto, H., Sugawara, A., Xia, L., Wang, X., Noguchi, N., Yoshikawa, T., et al. (2010). Upregulation of REG Iα accelerates tumor progression in pancreatic cancer with diabetes. Int. J. Cancer *127*, 1795–1803.

42. Li, Q., Wang, H., Zogopoulos, G., Shao, Q., Dong, K., Lv, F., Nwilati, K., Gui, X.Y., Cuggia, A., Liu, J.L., and Gao, Z.H. (2016). Reg proteins promote acinar-to-ductal metaplasia and act as novel diagnostic and prognostic markers in pancreatic ductal adenocarcinoma. Oncotarget *7*, 77838–77853.

43. Geijtenbeek, T.B.H., and Gringhuis, S.I. (2009). Signalling through C-type lectin receptors: shaping immune responses. Nat. Rev. Immunol. *9*, 465–479.

44. Astrosini, C., Roeefzaad, C., Dai, Y.-Y., Dieckgraefe, B.K., Jöns, T., and Kemmner, W. (2008). REG1A expression is a prognostic marker in colorectal cancer and associated with peritoneal carcinomatosis. Int. J. Cancer *123*, 409–413.

45. Harada, K., Zen, Y., Kanemori, Y., Chen, T.C., Chen, M.F., Yeh, T.S., Jan, Y.Y., Masuda, S., Nimura, Y., Takasawa, S., et al. (2001). Human REG I gene is up-regulated in intrahepatic cholangiocarcinoma and its precursor lesions. Hepatology *33*, 1036–1042.

46. Usami, S., Motoyama, S., Koyota, S., Wang, J., Hayashi-Shibuya, K., Maruyama, K., Takahashi, N., Saito, H., Minamiya, Y., Takasawa, S., et al. (2010). Regenerating gene I regulates interleukin-6 production in squamous esophageal cancer cells. Biochem. Biophys. Res. Commun. *392*, 4–8.

47. Sethi, J.K., and Hotamisligil, G.S. (2021). Metabolic Messengers: tumour necrosis factor. Nat. Metab. *3*, 1302–1312.

48. Schmiegel, W., Roeder, C., Schmielau, J., Rodeck, U., and Kalthoff, H. (1993). Tumor necrosis factor alpha induces the expression of transforming growth factor alpha and the epidermal growth factor receptor in human pancreatic cancer cells. Proc. Natl. Acad. Sci. USA *90*, 863–867.

49. Bianchi, A., De Castro Silva, I., Deshpande, N.U., Singh, S., Mehra, S., Garrido, V.T., Guo, X., Nivelo, L.A., Kolonias, D.S., Saigh, S.J., et al. (2023). Cell-Autonomous Cxcl1 Sustains Tolerogenic Circuitries and Stromal Inflammation via TNF in Pancreatic Cancer. Cancer Discov. *13*, 1428–1453.

50. Hui, D.Y. (2019). Group 1B phospholipase A2 in metabolic and inflammatory disease modulation. Biochim. Biophys. Acta. Mol. Cell Biol. Lipids *1864*, 784–788.

51. Goonesekere, N.C.W., Andersen, W., Smith, A., and Wang, X. (2018). Identification of genes highly downregulated in pancreatic cancer through a meta-analysis of microarray datasets: implications for discovery of novel tumor-suppressor genes and therapeutic targets. J. Cancer Res. Clin. Oncol. *144*, 309–320.

52. Kuefner, M.S. (2021). Secretory phospholipase A2s in insulin resistance and metabolism. Front. Endocrinol. *12*, 732726.

53. Hui, D.Y., Cope, M.J., Labonté, E.D., Chang, H.T., Shao, J., Goka, E., Abousalham, A., Charmot, D., and Buysse, J. (2009). The phospholipase A2 inhibitor methyl indoxam suppresses diet-induced obesity and glucose intolerance in mice. Br. J. Pharmacol. *157*, 1263–1269.

54. Sharma, A., Kandlakunta, H., Nagpal, S.J.S., Feng, Z., Hoos, W., Petersen, G.M., and Chari, S.T. (2018). Model to determine risk of pancreatic cancer in patients with new-onset diabetes. Gastroenterology *155*, 730–739.e3.

55. Li, Y., Bian, X., Wei, S., He, M., and Yang, Y. (2019). The relationship between pancreatic cancer and type 2 diabetes: cause and consequence. Cancer Manag. Res. *11*, 8257–8268.

56. Plebani, M., Banfi, G., Bernardini, S., Bondanini, F., Conti, L., Dorizzi, R., Ferrara, F.E., Mancini, R., and Trenti, T. (2020). Serum or plasma? An old question looking for new answers. Clin. Chem. Lab. Med. *58*, 178–187.

57. Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N., et al. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat. Genet. *50*, 1505–1513.

58. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.

59. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

60. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. *48*, 1443–1448.

61. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.

62. Wang, C., Zhan, X., Bragg-Gresham, J., Kang, H.M., Stambolian, D., Chew, E.Y., Branham, K.E., Heckenlively, J., FUSION Study; and Fulton, R., et al. (2014). Ancestry estimation and control of population stratification for sequence-based association studies. Nat. Genet. *46*, 409–415.

63. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.

64. Staley, J.R., Blackshaw, J., Kamat, M.A., Ellis, S., Surendran, P., Sun, B.B., Paul, D.S., Freitag, D., Burgess, S., Danesh, J., et al. (2016). PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics *32*, 3207–3209.

65. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. Genet. Epidemiol. *37*, 658–665.

66. Burgess, S., and Thompson, S.G. (2015). Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am. J. Epidemiol. *181*, 251–260.

67. Li, P., Stuart, E.A., and Allison, D.B. (2015). Multiple imputation: a flexible tool for handling missing data. JAMA *314*, 1966–1967.

68. Borenstein, M., Hedges, L.V., Higgins, J.P.T., and Rothstein, H.R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. Res. Synth. Methods *1*, 97–111.

69. Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

70. Hemani, G., Tilling, K., and Davey Smith, G. (2017). Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. *13*, e1007081.

71. Sanderson, E., Davey Smith, G., Windmeijer, F., and Bowden, J. (2019). An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. Int. J. Epidemiol. *48*, 713–727.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Olink Explore 1536	Olink	https://olink.com/products-services/explore/
Deposited data		
Proteomic data from the DFTJ	This paper	National Genomics Data Center: https://ngdc.cncb.ac.cn/omix: accession no. OMIX006240
Gene expression data of from GTEx v8	GTEx[35]	https://www.gtexportal.org/home/
Genetic, proteomic, and phenotype data from UKB-PPP	UK Biobank	https://biobank.ndph.ox.ac.uk/showcase/
GWAS data of pancreatic cancer	dbGaP	dbGaP: phs000206.v5.p3.c1; phs000206.v5.p3.c2; phs000648.v1.p1
GWAS summary statistics of T2D	Mahajan et al.[57]	https://www.diagram-consortium.org/index.html
1000 Genomes Project (Phase 3, v5a)	The 1000 Genomes Project	https://www.internationalgenome.org/
Human Genome Diversity Project	Li et al.[58]	http://www.hagsc.org/hgdp/
pQTL summary statistics from UKB-PPP	Sun et al.[30]	https://doi.org/10.1101/2022.06.17.496443
Software and algorithms		
R	R Core Team	https://www.r-project.org/
R package meta v6.0.0		https://www.r-project.org/
PLINK	Purcell et al.[59]	https://www.cog-genomics.org/plink/2.0
Eagle v2.4.1	Loh et al.[60]	https://github.com/poruloh/Eagle
Minimac4	Das et al.[61]	https://github.com/statgen/Minimac4
LASER v2.04	Wang et al.[62]	http://csg.sph.umich.edu/chaolong/LASER/
EPACTS v3.3.2		https://github.com/statgen/EPACTS
METAL	Willer et al.[63]	https://github.com/statgen/METAL
PhenoScanner v2	Staley et al.[64]	http://www.phenoscanner.medschl.cam.ac.uk/
R package TwoSampleMR v0.5.6	Burgess et al.[65] Burgess et al.[66]	https://github.com/MRCIEU/TwoSampleMR
R package coloc v5.1.1	Giambartolomei et al.[36]	https://github.com/chr1swallace/coloc
R package prop.coloc v1.1.0	https://arxiv.org/abs/2402.12171	https://github.com/ash-res/prop-coloc

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Chaolong Wang (chaolong@hust.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The proteomics and phenotype data from the Dongfeng-Tongji study have been deposited at the China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (OMIX: OMIX006240, https://ngdc.cncb.ac.cn/omix/release/OMIX006240).
- The genetics, proteomics and phenotype data of UK Biobank participants was obtained under the application number 88159.
- The individual-level GWAS datasets of pancreatic cancer were downloaded from the dbGaP: phs000206.v5.p3.c1; phs000206.v5.p3.c2; phs000648.v1.p1.
- The pQTL summary statistics from the UKB-PPP study are available at http://ukb-ppp.gwas.eu/.

- This study did not generate any unique code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The DFTJ cohort is a Chinese prospective cohort of retired workers. A total of 38,295 retired workers of the Dongfeng Motor Corporation in Shiyan, Hubei, China, were enrolled in 2013 and followed up until December 31, 2018. The UKB-PPP data included 52,705 baseline samples randomly selected from the UK Biobank. We accessed the UK Biobank data under the application number 88159.

METHOD DETAILS

Study design

We designed a nested case-control study of incident PC based on the DFTJ cohort[29] (Figure 1). In a median follow-up of 5.7 years, 57 incident PC cases were confirmed by trained medical personnel through medical insurance records, medical records, and death records in the health-care system. PC was defined by the International Classification of Diseases, 10th Revision (ICD-10, C25). After excluding individuals with a cancer history (n = 5) or insufficient blood samples (n = 8), we included 44 incident cases of PC. We matched one cancer-free control to each case by sex, age, hospital, and blood draw date (Figure S9).

Measurement of proteins

Baseline fasting serum samples of the DFTJ cohort were stored at −80°C refrigerators. After thawing to −20°C, 4°C, and room temperature, we aliquoted 20μL serum sample for each of the 88 individuals. All samples were randomized on a 96-well plate before shipping on dry ice to the Olink laboratory in Shanghai, China. A total of 1,463 unique protein biomarkers were measured using the Olink Explore 1536 panel, which was based on the proximity extension assay (PEA) technique. Each protein was quantified by a Normalized Protein eXpression (NPX) value. While some NPX values were below the limit of detection (LOD), we kept the reported values for the analysis. Among 88 samples, 4 samples, including 3 cases and 1 control, failed to pass the assay quality controls (i.e., exceeding 0.3 NPX deviation from the median value of the internal controls). We thus removed these 4 samples and their matched samples, leading to a final sample size of 40 case-control pairs for downstream analyses.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of the DFTJ data

Baseline characteristics, including education level, smoking status, drinking status, BMI, and glycemic indices, between PC cases and their matched controls were compared by paired *t*-test for continuous variables and Cochran-Mantel-Haenszel test for categorical variables. T2D was defined as self-reported physician diagnosis of T2D, use of hypoglycemic drugs or insulin, fasting blood glucose $\geq$7.0 mmol/L, or HbA1c $\geq$ 6.5%.

We applied the rank-based inverse normal transformation to standardize protein expression values. Conditional logistic regression (CLR) was used to estimate the odds ratio (OR) and the corresponding 95% confidence interval (CI) for each protein. Given the limited sample size of the DFTJ data, we used a lenient threshold of $p < 0.005$ to select proteins for further replication and MR analyses. We examined the gene expression profile of each protein across 54 tissues based on the RNA sequencing data from the GTEx Project, version 8.[35] In addition, we performed sensitivity analyses by adjusting for T2D and stratifying by sex. We stratified cases into two groups by every 3 years of their follow-up time to PC diagnosis and compared the protein expression, after regressing on age and sex, of each group to the controls by *t*-test.

All statistical analyses and visualizations were performed in R (version 4.1.2 and 4.2.0).

Replication using the UKB-PPP dataset

The replication samples were selected from the UKB-PPP, including 52,705 baseline samples measured on the same Olink Explore 1536 panel.[30] Among these 52,705 samples, 131 incident PC (ICD-10, C25) cases were documented in a median follow-up of 12.8 years, and 46,380 were from a random subsample of the full cohort. We excluded participants with a self-reported cancer history or missing disease history, with unknown ethnicity, or from the Stockport center, resulting in 109 incident PCs and 41,774 controls from the random subsample (Figure 1). Considering blood samples in UKB were non-fasting, we used blood glucose level as a matching variable between controls and cases. The Stockport center was excluded because no glucose measurement was provided by this center. We imputed the missing glucose level (missing rate = 12%) using the predictive mean matching method based on age, sex, ethnicity, center, and fasting time.[67] We performed multiple imputations 5 times, each with 10 iterations, and took the mean of the imputed values.

Instead of using all available samples, we performed matching between 109 incident PCs and 41,774 random baseline controls based on blood glucose (within ±1 mmol/L), age (within ±1 year), sex, ethnicity, center, and batch of protein assay, because matching allows for better control of potential nonlinear confounding effects of the matching variables. Each case was set to match a maximum of 10 controls to avoid extremely unbalanced case-control ratios. All matched controls were free of cancer and death events within

the same follow-up time of their cases. We further excluded 18 cases because no controls could be matched. Finally, we included 91 incident PCs and 739 matched controls in the analysis. We omitted missing data and standardized the NPX values of each protein by the rank-based inverse normal transformation. CLR-based association analysis was applied to all matched strata, and to subsets of strata in which cases were diagnosed in 2, 4, 6, 8, and 10 years from the baseline, respectively.

We combined results of the DFTJ study and the UKB-PPP subset restricting to those diagnosed with PC in 6 years by inverse variance weighted (IVW) fixed-effect and random-effect meta-analysis.[68] Both Q and I^2 statistics were reported for the heterogeneity between the DFTJ and the UKB analyses.

Mendelian randomization analysis

We performed two-sample MR to evaluate causal effects of the identified protein biomarkers on PC. Individual-level GWAS data of PC were obtained from the Pancreatic Cancer Cohort and Case-Control Consortium (PanScan) and the Pancreatic Cancer Case-Control Consortium (PanC4),[31–34] including four datasets totaling 7,843 PC cases and 7,719 controls (Table S7). We performed standard quality controls on each dataset using PLINK.[59] Briefly, we excluded SNPs with call rate < 0.95, minor allele count (MAC) < 3, or Hardy-Weinberg equilibrium (HWE) test $p < 10^{-6}$, and samples with call rate < 0.9, inbreeding coefficient < −0.1, mismatched sex, or duplicates. Phasing and imputation were performed on each dataset using Eagle (v2.4.1)[60] and Minimac4,[61] with 2,504 samples from the 1000 Genome Project as the reference panel.[69] SNPs with imputation $R^2 \leq 0.7$ or minor allele frequency (MAF) ≤ 0.005 were removed. Finally, we excluded samples with non-European ancestry based on principal components analysis (PCA) with worldwide samples from the Human Genome Diversity Project,[62] samples with missing data on age or sex, 2nd degree or closer relatedness, and PCA outliers (>6 SD from the mean in any of the top 10 principal components). The final sample sizes of the 4 datasets were shown in Table S7, totaling 7,176 PC cases and 6,983 controls. Based on genotyping data, we confirmed there was no sample overlap between studies.[59] We performed GWAS across 10,625,681 imputed variants in each dataset separately using logistic regression Wald test, adjusting for sex, age, and top 10 principal components, followed by IVW fixed-effect meta-analysis.[63]

We obtained pQTL summary statistics of 1,463 plasma proteins based on 34,557 UKB-PPP participants.[30] For each protein, we selected independent *cis*-pQTLs, which were within ±500kb from the transcription start site and had association $p < 5 \times 10^{-8}$,[70] as the instrumental variables (IVs). Candidate IVs with sample sizes <10,000 in the PC GWAS were excluded. Independent SNPs with linkage disequilibrium (LD) $r^2 < 0.01$ in 1Mb windows were extracted by the clumping method in PLINK.[59] We further excluded IVs associated with BMI, diabetes, glucose, or insulin-related traits ($p < 1 \times 10^{-5}$) using PhenoScanner.[64] In this step, we only excluded one IV for REG1B, rs4853451, which was reported to weakly associate with BMI at $p = 1.86 \times 10^{-6}$. We computed the proportion of variance explained (*PVE*) and the standard F statistic for each IV.[71] *PVE* was defined as $2f(1-f)\beta^2$, where f and β denote MAF and the pQTL effect size, respectively. We applied the IVW method to estimate the causal effect of each protein on PC.[65]

We conducted multivariable MR (MVMR) analysis with adjustment for T2D. GWAS summary statistics of T2D were obtained from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) study,[57] including 74,124 cases and 824,006 controls. We first extracted SNPs that are present and had sample sizes > 10,000 in all three datasets (i.e., *cis*-pQTL, T2D GWAS, and PC GWAS). We then selected IVs for MVMR as those significantly associated ($p < 5 \times 10^{-8}$) with the protein or T2D, followed by LD clumping ($r^2 < 0.01$ in 10Mb windows) using PLINK.[59] We estimated the causal effect of each protein on PC with adjustment of T2D using the IVW MVMR method.[66]

Colocalization analysis

We applied the Bayesian colocalization method *coloc*,[36] with prior probabilities set as $p_1 = p_2 = 1 \times 10^{-4}$ and $p_{12} = 5 \times 10^{-5}$. We also applied the conditional test of proportional colocalization, which does not assume a single causal variant, with default settings in the *prop.coloc* R package.[37] The LD matrix for *prop.coloc* was calculated based on 34,422 baseline samples of European ancestry in the UKB-PPP (batches 0 to 6).[30]

Association of protein biomarkers with different cancers

Based on ICD-10, we extracted a total of 128 cancer types from the UKB. We focused on 17 cancers with at least 100 incident cases during the follow-up among 52,705 baseline samples in UKB-PPP. For each cancer, we followed the same analysis procedure of PC to match each case to a maximum of 10 controls based on blood glucose (within ±1 mmol/L), age (within ±1 year), sex, ethnicity, center, and batch of protein assay. CLR-based association analyses were restricted to the strata with cases diagnosed within 6 years from the baseline.