# Associations of diabetes, circulating protein biomarkers, and risk of pancreatic cancer

Short title: Diabetes, proteomics, and pancreatic cancer

Yuanjie Pang, DPhil[1], Jun Lv, PhD[1,2,3], Ting Wu, BM[1], Canqing Yu, PhD[1,2], Yu Guo, MSc[4,5], Yiping Chen, DPhil[6,7], Ling Yang, PhD[6,7] Iona Y. Millwood, DPhil[6,7], Robin G. Walters, PhD[6,7], Xiaoming Yang, PhD[6], Rebecca Stevens, MS[6], Robert Clarke, MD[6], Junshi Chen, MD[8], Liming Li*, MD[1,2], Zhengming Chen*, DPhil[6,7], Christiana Kartsonaki*, DPhil[6,7]

*Joint senior authors

1. Department of Epidemiology & Biostatistics, School of Public Health, Peking University, 38 Xueyuan Road, Beijing 100191, China
2. Peking University Center for Public Health and Epidemic Preparedness & Response, Beijing, China
3. Key Laboratory of Molecular Cardiovascular Sciences, Ministry of Education, Peking University, Beijing, China
4. National Clinical Research Center of Cardiovascular Diseases, National Center for Cardiovascular Diseases
5. Fuwai Hospital, Chinese Academy of Medical Sciences, 167 Beilishi Road, Xicheng District, Beijing 100037, China
6. Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, Big Data Institute Building, Roosevelt Drive, University of Oxford, UK
7. Medical Research Council Population Health Research Unit (MRC PHRU) at the University of Oxford, Nuffield Department of Population Health, University of Oxford, UK
8. National Center for Food Safety Risk Assessment, 37 Guangqu Road, Beijing 100021, China

**Address for correspondence:**

Dr. Yuanjie Pang
Department of Epidemiology & Biostatistics
School of Public Health
Peking University
38 Xueyuan Road
Beijing, 100191, China
Tel: 86-010-82801528
Fax: 86-010-82801530
yuanjie_p@163.com

Prof. Liming Li
Dept. of Epidemiology & Biostatistics
School of Public Health
Peking University
38 Xueyuan Road
Beijing, 100191, China
Tel: 86-010-82801528
Fax: 86-010-82801530
lmleeph@vip.163.com

**Word count:** abstract 250, main text 2977
(3 tables, 3 figures and 11 supplementary tables/figures)

Revised in June 2023

## Abstract

**Background:** Type 2 diabetes (T2D) is associated with strong risk of pancreatic cancer (PC), but the underlying mechanisms are not fully understood.

**Methods:** We conducted a case-subcohort study involving 610 PC cases and 623 subcohort participants with 92 protein biomarkers measured in baseline plasma samples. Genetically-instrumented T2D was derived using 86 single-nucleotide polymorphisms (SNPs), including insulin resistance (IR) SNPs.

**Results:** In observational analyses of 623 subcohort participants (mean age, 52 years; 61% women), T2D was positively associated with 13 proteins (SD difference: IL6: 0.52 [0.23-0.81]; IL10: 0.41 [0.12-0.70]), of which 8 were nominally associated with incident PC. The 8 proteins potentially mediated 36.9% (18.7%-75.0%) of the association between T2D and PC. In MR, no associations were observed for genetically-determined T2D with proteins, but there were positive associations of genetically-determined IR with IL6 and IL10 (SD difference: 1.23 [0.05-2.41] and 1.28 [0.31-2.24]). In two-sample MR, fasting insulin was associated with both IL6 and PC, but no association was observed between IL6 and PC.

**Conclusions:** Proteomics were likely to explain the association between T2D and PC, but were not causal mediators. Elevated fasting insulin driven by insulin resistance might explain the associations of T2D, proteomics, and PC.



***Keywords****: diabetes; proteomics; pancreatic cancer; Chinese*

## Introduction

Pancreatic cancer (PC) ranks the 10th commonest cancer globally[1]. The prognosis of PC is abysmal with a 5-year survival of only 5-10%[2]. Currently, there are no effective treatments for PC, though several randomised controlled trials are ongoing[3-5]. Previous studies in Western countries and East Asia have shown that lifestyle factors (e.g. smoking, alcohol, physical inactivity, adiposity) are possible risk factors for PC[6,7].

Type 2 diabetes (T2D) is the most important risk factor for PC. A recent meta-analysis has shown that T2D was associated with 52% higher risk of PC both in Europeans and East Asians[8]. A recent report from the China Kadoorie Biobank (CKB) demonstrated that serum protein biomarkers were associated with higher risk of PC, reflecting angiogenesis, IL6 signaling, and autophagy pathways[9]. Taken together, these studies provide opportunities to understand the biological mechanisms between T2D and PC.

Mendelian randomisation studies have shown no evidence of a causal association between T2D and PC, but suggested a causal role of fasting insulin in the aetiology of PC[10]. To provide insights on the underlying mechanisms linking T2D and PC, evidence is needed to compare and contrast the genetic associations of diabetes and fasting insulin with proteomics.

Therefore, the objectives of this study were to examine the conventional observational and genetically estimated associations of T2D with inflammation and immune-associated proteins in a subcohort of the CKB. We also evaluated the extent to which these protein biomarkers mediated the association between T2D and PC.

## Results

Of the 1233 participants included in the case-subcohort study, the mean age at study baseline was higher in pancreatic cancer cases than that of subcohort participants (60.3 [SD 9.0] vs 52.1 [10.5 years]). Cases were more likely to be male, regular smokers, regular alcohol drinkers, and have higher SBP and prevalent diabetes at baseline (**Table 1**). Among pancreatic cancer cases, the median time from study entry to diagnosis was 5.3 years (interquartile range [IQR] 4.3) and mean age at diagnosis was 66.0 (SD 8.9).

*Observational associations of proteomics, diabetes and PC*

After adjusting for multiple comparisons, there were positive associations between T2D and 13 of the 90 protein biomarkers (at 5% FDR; **Figure 2A** and **Supplementary Table 4**). The associations of proteomics with long-standing diabetes were similar to those with new-onset diabetes (*p*-value for heterogeneity >0.05, **Supplementary Figure 1**). Of the 13 diabetes-associated protein biomarkers, there were nominal associations of 8 proteins with risk of incident PC (**Figure 2C**), including CAIX, IL18, IL6, ANGPT2, MCP3, CD8a, TNFSF14, and TIE2. Mediation analysis showed that the proportion mediated by individual protein biomarker ranged from 4.4% (0.6%-13.0%) (TNFSF14) to 17.1% (6.2%-60.0%) (CD8a), while the proportion mediated by all 8 proteins was 36.9% (18.7%-75.0%, **Table 2**). Compared with long-term risk of PC (occurring after 1 year of follow-up), protein biomarkers mediated a smaller proportion of the association between T2D and short-term risk of PC (within 1 year, 26.0% [15.1%-67.0%], **Supplementary Table 5**).

*Genetic associations of proteomics with diabetes*

In genetic analyses, there were no associations of GRS-T2D with the 13 T2D-associated

protein biomarkers (at 5% FDR, **Figure 2B** and **Supplementary Table 4**). Similarly, no associations were observed for GSR-BC and these protein biomarkers (**Supplementary Table 4**). In contrast, there were nominal associations of GRS-IR with 2 of the 13 T2D-associated protein biomarkers, with SD difference of 1.23 (0.05-2.41) for IL6 and 1.28 (0.31-2.24) for IL10 associated with genetically determined IR (**Figure 2B**). The Pearson correlation coefficient between the observational and genetic estimates was 0.045 for GRS-T2D and 0.28 for GRS-IR. The Cochran Q test showed no evidence for differences between the genetic and observational estimates, with the exception of 5 proteins (NOS3, CAIX, IL18, ANGPT2, and MCP3).

*Genetic associations of IL6, fasting insulin and PC*

**Figure 3** and **Supplementary Table 6** shows the results of two-sample Mendelian randomisation. There were no genetic associations of T2D with IL6 and PC. In contrast, there were genetic associations of fasting insulin with IL6 (SD difference 0.23 [0.07-0.40]) and with PC (OR 1.90 [1.28-2.83]). However, there was no association between genetically-determined IL6 activity and PC (PR 1.65 [0.86-3.16]). Two-step MR showed that IL6 mediated 8.8% of the genetic association between fasting insulin and PC.

*Subgroup and sensitivity analyses*

For IL6R and IL10, protein biomarkers on the IL6/IL6R pathway, there was a negative genetic association of fasting insulin with IL6R (SD difference -0.23 [-0.40 to -0.06]) and no association with IL10 (-0.15 [-0.39 to 0.09], **Supplementary Table 6**). Bi-directional MR showed that genetically-determined IL6 activity was not associated with T2D or fasting insulin (SD difference 0.003 [-0.028 to 0.035] and 0.004 [-0.028 to 0.035]). In MR-Egger

and weighted median analyses, the results were generally consistent with those estimated using individual participant–level data (**Supplementary Table 7-8**).

## Discussion

In this Chinese population, diabetes was associated with a range of protein biomarkers, which mediated 37% of the association between T2D and PC. Despite the observational associations, there was no evidence of genetic associations of T2D with these proteins. In contrast, there was evidence of genetic associations of insulin resistance with IL6 and IL10. Two-sample MR showed that there was possible causal association between fasting insulin and IL6. Nonetheless, there was no evidence of genetic association between altered IL6 activity and PC. Findings of this study suggested that proteomics were likely to explain the association between T2D and PC, but were not causal mediators. Elevated fasting insulin driven by insulin resistance might explain the associations of T2D, proteomics, and PC.

Previous studies have focused on the protein signatures associated with incidence of T2D[21-23], rather than whether T2D causes disturbances in circulating proteins. A recent analysis in the Malmö Diet and Cancer-Cardiovascular Cohort study showed that 18 protein biomarkers associated with waist-to-hip ratio were associated with incident T2D[21], which included IL18, CCL20, and LAP-TGF-beta1 overlapping with the current study. Despite the small number of incident events in the subcohort, we showed consistent results for the associations between proteomics and incident T2D with the Swedish study (**Supplementary Table 9**). Another European study measured proteomics using the SOMAScan platform and conducted bi-directional Mendelian randomisation on T2D and

protein biomarkers[23]. Despite the different coverage from Olink Immuno-Oncology panel (used in the present study), this study showed no evidence of genetic associations between T2D and protein biomarkers involved in amino acid metabolism, growth hormone receptors, tumour necrosis factor superfamily, and renin-angiotensin and kallikrein-kinin systems (**Supplementary Table 10**). Consistent with this study, our study showed no evidence of genetic associations between T2D and protein biomarkers involved in the same metabolism pathways and/or belonging to the same classes.

Despite the null association between genetically-determined T2D and protein biomarkers, our study reported a positive genetic association between insulin resistance and IL6, which was confirmed by two-sample Mendelian randomisation on fasting insulin and IL6. The discrepancy between genetically-determined T2D and fasting insulin on disease phenotypes has been reported for PC. Pooled analysis of the PanScan and PanC4 reported no evidence of a causal relationship between T2D and PC but showed that genetically increased fasting insulin was causally associated with a higher risk of PC[10]. As hyperinsulinemia occurs in the early stage of T2D, insulin may be a confounder for any observed association between diabetes and disease phenotype (PC or elevated IL6). Therefore, insulin per se rather than diabetes as a consequence of hyperinsulinemia may be the causal factor.

Nonetheless, two-sample Mendelian randomisation found no evidence of genetic association between altered IL6 activity and PC. On one hand, the lack of genetic association between altered IL6 activity and PC may be due to weak instrument bias. Indeed, F-statistics for the 3 SNPs for IL6 ranged from 24 to 138. Large GWAS with

increased power is needed to develop better instruments for IL6 signaling. Alternatively, IL6 elevation may be a manifestation in the subclinical stage of PC, rather than a cause of PC. IL6 has been shown to be associated with an adverse prognosis and progression of PC[24,25], and RCTs have been undertaken to investigate IL6R as a potential drug target for PC[26]. However, previous prospective cohort studies in Europeans showed no evidence of association between IL6 and PC risk[27,28]. Indeed, a previous report in CKB showed that the positive association between IL6 and PC was stronger for short-term PC (occurring in the first year) than long-term PC[9], suggesting higher levels of IL6 in the time preceding diagnosis. It is likely that in the early stage of PC, elevated IL6 is driven by fasting insulin associated with T2D. Despite the strong observational associations, proteomics are not likely to be causal mediators between T2D/fasting insulin and PC. More studies are warranted to explore possible causal mediators between T2D/fasting insulin and PC, with a special focus on the IL6/IL6R signaling pathway.

Our study on proteomics along with two-sample Mendelian randomisation using GWAS summary statistics is one of the first attempts to search for causal mediators for PC. Previous studies on causal mediators of non-communicable diseases have focused on coronary heart disease and chronic kidney disease and identified CSF1 and CXCL12 as promising drug targets for coronary heart disease and HER2 and uromodulin for chronic kidney disease [29,30]. As IL6/IL6R signaling pathway is the most promising target for PC, future studies are warranted to quantify protein candidates from IL6 and IL6R genes identified from trans-ethnic GWAS and to develop customise targeted panels to measure proteomics.

The strengths of the CKB included use of a prospective design, coverage of a wide range of blood-based protein biomarkers involved in multiple biological pathways, and use of three complementary types of analyses to assess genetically estimated associations of T2D, proteins, and PC in the same study population. This study also had several limitations. First, IR phenotypes such as fasting insulin and homeostasis model of assessment insulin resistance were unavailable in CKB. Although we constructed GRS-IR using 6 SNPs identified in previous GWAS, we were not able to examine the associations of GRS-insulin resistance with fasting insulin or measures of insulin resistance. As a result, the genetic association of GRS-IR with IL6 could be attributable to either circulating levels of insulin or insulin sensitivity. However, we showed consistent results when using fasting insulin SNPs in two-sample MR (**Supplementary Table 6**). Second, it is possible that a subset of SNPs included in the diabetes genetic score may affect protein biomarkers independently of T2D, potentially violating the assumptions of Mendelian randomisation. However, we showed that weighted median and MR-Egger estimates were largely consistent with the inverse-variance weighted estimates in CKB (**Supplementary Figure 8**). Third, the three SNPs for IL6 activity included 2 *cis*-SNPs on IL6R gene and 1 *trans*-SNP on TDRD10 gene. Therefore, non-specific (i.e. horizontally pleiotropic) effects of the IL6 *trans*-acting instruments cannot be excluded. However, we showed no horizontal ploeiotropy in MR-Egger analysis (**Supplementary Table 7**). Fourth, plasma samples in CKB were stored at -80°C for ~10 years before conducting the proteomics assay. However, an external study showed that the storage time explained between 5% to 35% of the variation for single proteins. Five proteins (CD40L, FASLG, IL13, LAP-TGF-beta1, and MMP7) were included in the current study and the variance explained ranged from 4.9% to

34.9%. Fifth, we only examined the observational associations of diabetes, proteins, and risk of pancreatic cancer, without investigating the genetic associations between diabetes and pancreatic cancer in CKB. This is because GWAS data were only available for a subset of the case-subcohort study and there was lack of power in genetic analyses. However, previous studies conducted in Europeans reported a positive genetic association between IR and pancreatic cancer[10]. The genetic analyses of T2D and pancreatic cancer were conducted using summary-level GWAS data published in previous studies (**Figure 3**). Sixth, diabetes was assessed at study baseline and incident diabetes occurring during the follow-up was not accounted for. However, inclusion of incident diabetes might mask the temporal associations between diabetes and proteins. By utilising a GRS for diabetes we were able to identify alterations in proteins associated with diabetes. Seventh, although the majority of pancreatic cancer cases were likely to be pancreatic ductal adenocarcinoma, we do not have detailed information on histological subtypes for all cases. However, previous study in CKB showed that the association between diabetes and PC was in agreement with previous studies ascertaining pancreatic ductal adenocarcinoma by pathological examinations[8], suggesting that misclassification of PC diagnosis should minimally bias our results. Last, although our study measured proteomics in a relatively large number of participants, the sample size might not be large enough to identify some associations of smaller magnitude between T2D and protein biomarkers and between protein biomarkers and PC risk.

In conclusion, our study showed that T2D was associated with several inflammatory and immune-associated protein biomarkers. For these protein biomarkers, there was

suggestive evidence of causal associations of insulin resistance with IL6 and IL10. Using summary statistics of large GWAS in Europeans and East Asians, there was a causal association of fasting insulin with IL6 but no causal association of IL6 with PC, suggesting that IL6 is not a causal mediator between fasting insulin and PC. More studies are warranted to provide potential insights into the biological mechanisms linking T2D, IL6 signaling pathway, and PC.

## Methods

### *Study population*

The CKB is a prospective cohort study of 512,891 Chinese adults aged 30-79 years who were recruited from 10 regions (five urban and five rural) in China during 2004-08. Details of the CKB design, survey methods, and long-term follow-up have been previously described[11]. Ethics approval was obtained from the Oxford University Tropical Research Ethics Committee, the Chinese Centre for Disease Control and Prevention (CDC) Ethical Review Committee and the local CDC of each study area. All participants provided written informed consent.

### *Case-subcohort study of pancreatic cancer*

A case-subcohort study was designed to examine the associations of proteomics with risk of incident pancreatic cancer, involving 700 cases of pancreatic cancer (International Statistical Classification of Diseases and Related Health Problems, Tenth Revision [ICD-10] code C25) that accumulated

until January 1, 2016, and a subcohort of 700 participants selected from the baseline cohort using simple random sampling with genome-wide genotyping data. 167 participants were excluded for not passing quality control (i.e. with either a quality control warning or precipitation), leaving 1233 participants for the present study.

*Proteomics assay*

The Olink Immuno-Oncology assay measured 92 protein biomarkers selected to include proteins known or suspected to be involved in promotion and inhibition of tumour immunity, chemotaxis, vascular and tissue remodelling, apoptosis and cell killing, and metabolism and autophagy. The Olink method is based on proximity extension assay technology, to obtain normalized protein expression values, which is an arbitrary unit on a log2 scale[12-14]. All biomarkers were standardized to have a SD of 1. Assessment of baseline diabetes, lifestyle factors, and other covariates are described in the **eMethods**.

*Ascertainment of T2D*

A 10-ml nonfasting (with the time since the participant last ate recorded) blood sample was collected from participants into an ethylene diamine tetraacetic acid vacutainer (EDTA) vacutainer (BD Hemogard, NJ, US). Immediate on-site testing of RPG level was undertaken using the SureStep Plus System (Johnson & Johnson, California, US), regularly calibrated with manufacturer quality control solution. Participants with glucose levels ≥7.8 mmol/L and <11.1 mmol/L were invited to return for a fasting plasma glucose (FPG) test the next day. RPG data were unavailable for 8341 participants (because of a delay in making the on-site test available in certain regions).

Previously diagnosed diabetes was defined by the question “Has a doctor ever told you that you had diabetes?”. Among positive respondents, additional information about age at diagnosis and current use of certain medications for the treatment of diabetes (insulin and chlorpropamide or metformin) and CVDs (e.g. aspirin, lipid and blood pressure lowering agents) was collected. Among those without previously diagnosed diabetes, screen-detected diabetes was defined as (1) RPG ≥7.0 mmol/L if the time since last eating was ≥8 h, (2) RPG ≥11.1 mmol/L if the time since last eating was <8 h, or (3) FPG ≥7.0 mmol/L on subsequent testing.

*Genetic risk score for T2D*

We constructed a genetic risk score (GRS) for T2D using 86 single nucleotide polymorphisms (SNPs) developed in Asian and European populations and validated in Chinese (GRS-T2D) (**Supplementary Table 1**)[15]. Detailed selection criteria of these SNPs have been reported elsewhere[15]. Briefly, these 86 SNPs involved T2D SNPs originally reported among South Asians, East Asians, and Europeans which did not demonstrate heterogeneity in associations with T2D between European and East Asian populations if first reported in Europeans. This GRS has been shown to predict T2D in CKB with good performance (C statistic 0.593 [0.586, 0.600]). We also constructed GRS using T2D-associated variants with specific pathophysiological mechanisms: beta cell dysfunction (GRS-BC) (24 SNPs) and insulin resistance (GRS-IR) (6 SNPs).

Beta cell dysfunction related SNPs were identified by: (1) association with decreased homeostasis model assessment of beta cell function in individuals without diabetes[16]; (2) association with one of the beta cell function indices during an

oral glucose tolerance tes[17,18]; (3) presence in a locus influencing beta cell function according to cluster analysis[17]; and/or (4) the existence of rare variants responsible for forms of monogenic diabetes characterised by insulin secretory failure. IR-related SNPs were identified by: (1) association with increased HOMA-IR ($p$<0.05, β> 0 for risk allele) in individuals without diabetes[18] or with decreased insulin sensitivity index[16]; (2) association with fasting insulin[19]; (3) presence in a locus influencing insulin sensitivity according to cluster analysis[17]; (4) association with increased triacylglycerol or other IR-related traits[19]; and (5) not acting primarily through obesity[16].

*Statistical methods*

The observational associations between T2D and proteomics were conducted in the subcohort (**Figure 1**). Linear regression was used to assess the associations of T2D with protein markers, adjusted for age, age squared, sex, area, education, household income, smoking, alcohol, self-rated health, and fasting time (i.e. time since last having eaten). For each biomarker, adjusted SD differences and 95% confidence intervals (CI) associated with total T2D were estimated. To assess whether the associations differed between new-onset diabetes and long-standing diabetes, we classified diabetes status by time since diagnosis of T2D (≤2 years and >2 years since diagnosis to distinguish between new-onset and long-standing diabetes, respectively). The associations between proteins and risk of pancreatic cancer were assessed using Cox proportional hazards models (**Figure 1**), using the Prentice pseudo-partial likelihood, adjusted for age, age squared (to account for the non-linear association between age and the outcome), sex, area, education, household income, smoking, alcohol, self-rated health, and fasting time. There was one pancreatic cancer case occurring in the subcohort, and this case was treated as

contributing to risk sets from the time of entering the study to its event time. To assess the role of protein biomarkers in the association between T2D and PC, proportion mediated by individual protein biomarker and a protein score was calculated using the “mediation” package in R[20]. Protein biomarkers were selected which were associated with both T2D and risk of PC with a false-discovery rate (FDR)-adjusted *p*-value <0.05. A protein score was constructed by summing the concentrations of these proteins, weighted by the coefficient of each protein on T2D. This process was repeated with time censored one year after study entry to examine the mediation effect of protein biomarkers between T2D and short-term risk of pancreatic cancer.

The genetic associations between T2D and proteomics were estimated in the subcohort (**Figure 1**). In Mendelian randomisation analysis, we calculated the genetically estimated associations of T2D with proteomics by the 2-stage least squares estimator method using individual participant-level data. In the first stage, the associations between GRS-T2D and diabetes were examined in 75,736 participants in the GWAS population subset using linear regression, adjusting for age, age squared, sex, area, the first 12 principal components, education, smoking, and alcohol. In the second stage, the associations of the resulting estimated T2D with proteomics were examined in the subcohort of 623 individuals using linear regression with the same adjustments. Significance was assessed at a 5% FDR in the observational analysis of T2D with protein biomarkers. Unadjusted p-values are reported for the genetic associations of T2D with protein biomarkers and observational associations of protein biomarkers with PC to avoid overcorrection.

For proteomics, IL6 was selected for exploration of causal effects because in CKB there was (1) observational association of T2D with IL6, (2) genetic association of IR with IL6,

and (3) observational association of IL6 with PC risk. Previous studies have suggested that genetically elevated fasting insulin is associated with PC[10]. To understand the underlying mechanisms, we examined the genetic associations of fasting insulin with IL6 and PC using two-sample Mendelian randomisation. We reported in the **eMethods** in the **Supplementary Materials** details of SNP selection (**Supplementary Table 2-3**) as well as methods for two-sample Mendelian randomisation and two-step Mendelian randomisation.

**Acknowledgements**

The chief acknowledgment is to the participants, the project staff, and the China National Centre for Disease Control and Prevention (CDC) and its regional offices for access to death and disease registries. The Chinese National Health Insurance scheme provides electronic linkage to all hospital admission data.

**Data availability**

CKB investigators are committed to sharing this important resource with the wider scientific community, so that the potential value of the CKB resource can be maximised. Open access to the CKB resource has begun in a phased approach. To facilitate the process a Data Access Committee (see http://www.ckbiobank.org/site/Data+Access) has been established, comprising not only senior CKB scientists but also external experts in related fields. For any external data access requests, an outline proposal defining the purpose of the investigation, the data/samples required and the time-scale for the analysis needs to be completed and submitted for review by the study executive committee. The access request review will assess the scientific merit of the proposal to ensure that research questions are legitimate and that there is no duplication of effort. Only proposals complying with the activities listed in the participant's original consent and with the study's ethical approval will be considered.

To facilitate future collaboration and streamline data sharing and access, a detailed policy document on data access and a related IT platform has been developed and made available on the study web site (www.ckbiobank.org). The policy reflects the principles of

the data access policies promoted by the study funders, as well as certain specific conditions already agreed with the original funder (the Kadoorie Foundation) and Chinese government. Information on access to the CKB resource is actively disseminated through workshops, seminars and conference presentations, in published articles, and through the study website. To date over 250 researchers have registered through our data sharing system and over 100 datasets have been securely delivered to open access users and collaborators using this facility.

**Funding**

This work was supported by National Natural Science Foundation of China (91846303, 81941018). The CKB baseline survey and the first re-survey were supported by a grant from the Kadoorie Charitable Foundation in Hong Kong. The long-term follow-up is supported by grants (2016YFC0900500, 2016YFC0900501, 2016YFC0900504) from the National Key R&D Program of China, and Chinese Ministry of Science and Technology (2011BAI09B01). Dr Pang acknowledges support from the Peking University Medicine Fund of Fostering Young Scholars' Scientific & Technological Innovation (BMU2022RCZX022), the Fundamental Research Funds for the Central Universities, and the Peking University Start-up Grant (BMU2022PY014). The proteomics analysis was funded by NDPH Pump Priming Award, Pancreatic Cancer UK (A102016RIFProfZChen), and Cancer Research UK Oxford Centre (C552/A17720). The funders had no role in the study design, data collection, data analysis and interpretation, writing of the report, or the decision to submit the article for publication.

**Conflict of interest**

The authors declare that they have no competing interests.

## Contribution statement

YP, LL, ZC, and CK had full access to the data. YP and CK conducted data analysis and are responsible for accuracy of the results and the decision to submit for publication. All authors were involved in study design, conduct, long-term follow-up, review and coding of disease events, interpretation of the results, or writing the report. All authors approved the final version of the manuscript. CK is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## References


1. Kocarnik JM, Compton K, Dean FE, Fu W, Gaw BL, Harvey JD, et al. Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *JAMA Oncol.* 2022; 8(3): 420-444.
2. The global, regional, and national burden of pancreatic cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol.* 2019; 4(12): 934-947.
3. Versteijne E, van Dam JL, Suker M, Janssen QP, Groothuis K, Akkermans-Vogelaar JM, et al. Neoadjuvant chemoradiotherapy versus upfront surgery for resectable and borderline resectable pancreatic cancer: long-term results of the Dutch randomized PREOPANC trial. *J Clin Oncol.* 2022; 40(11): 1220-1230.
4. Zhu X, Cao Y, Liu W, Ju X, Zhao X, Jiang L, et al. Stereotactic body radiotherapy plus pembrolizumab and trametinib versus stereotactic body radiotherapy plus gemcitabine for locally recurrent pancreatic cancer after surgical resection: an open-label, randomised, controlled, phase 2 trial. *Lancet Oncol.* 2022; 23(3): 105-115.
5. Padrón LJ, Maurer DM, O'Hara MH, O'Reilly EM, Wolff RA, Wainberg ZA, et al. Sotigalimab and/or nivolumab with chemotherapy in first-line metastatic pancreatic cancer: clinical and immunologic analyses from the randomized phase 2 PRINCE trial. *Nat Med.* 2022; 28(6): 1167-1177.
6. Pang Y, Holmes MV, Chen Z, Kartsonaki C. A review of lifestyle, metabolic risk factors, and blood-based biomarkers for early diagnosis of pancreatic ductal adenocarcinoma. *J Gastroenterol Hepatol.* 2019; 34(2): 330-345.
7. Pang Y, Holmes MV, Kartsonaki C, Guo Y, Yang L, Bian Z, et al. Young adulthood and adulthood adiposity in relation to incidence of pancreatic cancer: a prospective study of 0.5 million Chinese adults and a meta-analysis. *J Epidemiol Commun Health* 2017; 71(11): 1059-1067.
8. Pang Y, Kartsonaki C, Guo Y, Bragg F, Yang L, Bian Z, et al. Diabetes, plasma glucose and incidence of pancreatic cancer: A prospective study of 0.5 million Chinese adults and a meta-analysis of 22 cohort studies. *Int J Cancer.* 2017; 140(8): 1781-1788.
9. Kartsonaki C, Pang Y, Millwood I, Yang L, Guo Y, Walters R, et al. Circulating proteins and risk of pancreatic cancer: a case-subcohort study among Chinese adults. *Int J Epidemiol.* 2022; 51(3): 817-829.
10. Carreras-Torres R, Johansson M, Gaborieau V, Haycock PC, Wade KH, Relton CL, et al. The role of obesity, type 2 diabetes, and metabolic factors in pancreatic cancer: A Mendelian Randomization study. *J Natl Cancer Inst.* 2017; 109(9). doi: 10.1093/jnci/djx012.
11. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011; 40(6): 1652-1666.
12. Lundberg M, Thorsen SB, Assarsson E, Villablanca A, Tran B, Gee N, et al.

Multiplexed homogeneous proximity ligation assays for high-throughput protein biomarker research in serological material. *Mol Cell Proteomics.* 2011; 10(4): M110.004978.
13. Assarsson E, Lundberg M, Holmquist G, Björkesten J, Thorsen SB, Ekman D, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One.* 2014; 9(4): e95192. doi: 10.1371/journal.pone.0095192.
14. Olink. Olink validation document for the Immuno-Oncology panel (article number 95310). Accessed December 2018. https://www.olink.com/resources-support/document-download-center/
15. Gan W, Bragg F, Walters RG, Millwood IY, Lin K, Chen Y, et al. Genetic predisposition to type 2 diabetes and risk of subclinical atherosclerosis and cardiovascular diseases among 160,000 Chinese adults. *Diabetes.* 2019; 68(11): 2155-2164.
16. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* 2012; 44(9): 981-990.
17. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert MF, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes.* 2014; 63(6): 2158-2171.
18. Prokopenko I, Poon W, Mägi R, Prasad BR, Salehi SA, Almgren P, et al. A central role for GRB10 in regulation of islet function in man. *PLoS Genet.* 2014; 10(4): e1004235.
19. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Na Genet.* 2012; 44(6): 659-669.
20. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software* 2014; 59(5): 1-38.
21. Bao X, Xu B, Yin S, Pan J, Nilsson PM, Nilsson J, et al. Proteomic profiles of body mass index and waist-to-hip ratio and their role in incidence of diabetes. *J Clin Endocrinol Metab.* 2022; 107(7): 2982-2990.
22. Molvin J, Pareek M, Jujic A, Melander O, Råstam L, Lindblad U, et al. Using a Targeted Proteomics Chip to Explore Pathophysiological Pathways for Incident Diabetes- The Malmö Preventive Project. *Sci Rep.* 2019; 9(1): 272. doi: 10.1038/s41598-018-36512-y.
23. Elhadad MA, Jonasson C, Huth C, Wilson R, Gieger C, Matias P, et al. Deciphering the plasma proteome of type 2 diabetes. *Diabetes.* 2020; 69(12): 2766-2778.
24. Palmquist C, Dehlendorff C, Calatayud D, Hansen CP, Hasselby JP, Johansen JS. Prediction of unresectability and prognosis in patients undergoing surgery on suspicion of pancreatic cancer using Carbohydrate Antigen 19-9, Interleukin 6, and YKL-40. *Pancreas.* 2020; 49(1): 53-61.
25. Ramsey ML, Talbert E, Ahn D, Bekaii-Saab T, Badi N, Bloomston PM, et al. Circulating interleukin-6 is associated with disease progression, but not cachexia in pancreatic cancer. *Pancreatology.* 2019; 19(1): 80-87.
26. Long KB, Tooker G, Tooker E, Luque SL, Lee JW, Pan X, et al. IL6 receptor blockade enhances chemotherapy efficacy in pancreatic ductal adenocarcinoma. *Mol Cancer*

*Ther.* 2017; 16(9): 1898-908.
27. Grote VA, Kaaks R, Nieters A, Tjønneland A, Halkjær J, Overvad K, et al. Inflammation marker and risk of pancreatic cancer: a nested case-control study within the EPIC cohort. *Br J Cancer.* 2012; 106(11): 1866-1874.
28. Bao Y, Giovannucci EL, Kraft P, Qian ZR, Wu C, Ogino S, et al. Inflammatory plasma markers and pancreatic cancer risk: a prospective study of five U.S. cohorts. *Cancer Epidemiol, Biomarkers Prev.* 2013; 22(5): 855-861.
29. Sjaarda J, Gerstein H, Chong M, Yusuf S, Meyre D, Anand SS, et al. Blood CSF1 and CXCL12 as Causal Mediators of Coronary Artery Disease. *J Am Coll Cardiol.* 2018; 72(3): 300-310.
30. Sjaarda J, Gerstein HC, Yusuf S, Treleaven D, Walsh M, Mann JFE, et al. Blood HER2 and Uromodulin as Causal Mediators of CKD. *J Am Soc Nephrol.* 2018; 29(4): 1326-1335.

**Table 1. Baseline characteristics of pancreatic cancer cases and subcohort participants**

| | Cases (n = 610) | Subcohort (n = 623) |
|---|---|---|
| Age (SD), years | 60.3 (9.0) | 52.1 (10.5) |
| Female, % | 50.6 | 60.9 |
| **Socioeconomic and lifestyle factors** | | |
| Urban region, % | 48.7 | 50.1 |
| ≥9 years of education, % | 20.5 | 16.4 |
| Household income ≥35,000 yuan/year, % | 20.2 | 18.0 |
| Ever-regular smoker, % | | |
| Male | 80.8 | 73.7 |
| Female | 7.1 | 4.2 |
| Ever-regular alcohol drinking, % | | |
| Male | 46.7 | 34.2 |
| Female | 3.9 | 2.6 |
| Total physical activity (SD), MET-h/day | 19.1 (14.0) | 20.4 (14.5) |
| **Blood pressure and anthropometry** | | |
| SBP (SD), mmHg | 137.2 (21.2) | 131.3 (21.8) |
| BMI (SD), kg/m$^2$ | 23.8 (3.5) | 23.8 (3.5) |
| Body fat percentage (SD) | 26.9 (9.1) | 28.5 (8.5) |
| **Prior disease history, %** | | |
| T2D* | 13.6 | 6.3 |
| Coronary heart disease | 4.3 | 3.7 |
| Stroke or TIA | 2.7 | 2.4 |
| Family history of diabetes$^\dagger$ | 4.2 | 4.3 |
| Family history of cancer$^\dagger$ | 13.7 | 12.4 |

Abbreviations: BMI, body mass index; SBP, systolic blood pressure; MET, metabolic equivalent of task; T2D, type 2 diabetes; TIA, transient ischemic attack.

* Self-reported or screen-detected diabetes

$^\dagger$Family history of diabetes or cancer included family history of any of father, mother, or siblings.

**Table 2. Associations of T2D and risk of PC and the mediating effect of protein biomarkers**

| | HR (95% CI) | % mediated |
|---|---|---|
| Overall coefficient | 1.62 (1.29, 2.03) | |
| *Adding protein biomarkers* | | |
| CAIX | 1.51 (1.19, 1.91) | 5.8 (-4.0, 15.0) |
| IL18 | 1.45 (1.15, 1.83) | 16.5 (4.9, 40.0) |
| IL6 | 1.46 (1.16, 1.84) | 9.4 (3.6, 27.0) |
| ANGPT2 | 1.49 (1.18, 1.87) | 11.1 (3.5, 23.0) |
| MCP3 | 1.42 (1.13, 1.80) | 12.7 (5.0, 24.0) |
| CD8a | 1.49 (1.19, 1.88) | 17.1 (6.2, 60.0) |
| TNFSF14 | 1.54 (1.22, 1.94) | 4.4 (0.6, 13.0) |
| TIE2 | 1.53 (1.22, 1.93) | 9.0 (-2.9, 32.0) |
| **Protein score** | **1.17 (0.92, 1.49)** | **36.9 (18.7, 75.0)** |

The observational estimates were adjusted for age, age squared, sex, area, education, household income, smoking, alcohol, self-rated health, and fasting time. A protein score was constructed by summing the concentrations of 8 proteins, weighted by the coefficient of each protein on T2D.

## Figure legends

### Figure 1. Flow diagram

A flow diagram to show participants whose data were used to estimate observational and genetic associations of T2D, proteomics, and PC in the China Kadoorie Biobank (CKB). Abbreviations: T2D, type 2 diabetes; PC, pancreatic cancer; FI, fasting insulin; IR, insulin resistance.

### Figure 2. Associations of T2D, proteomics, and PC

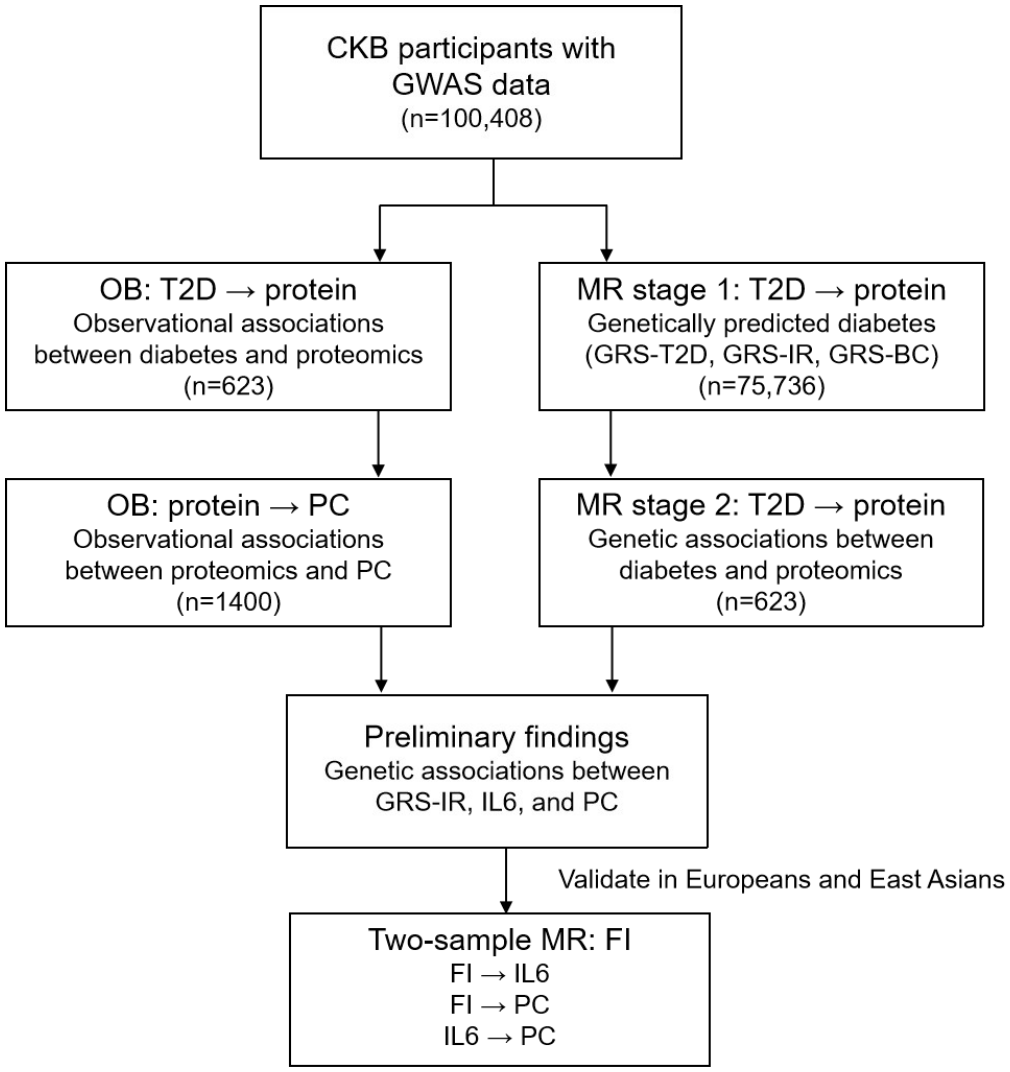
(a), Adjusted SD differences (95% CI) of protein biomarkers associated with T2D for 13 protein biomarkers with FDR-corrected $p$-value < 0.05. (b), Corresponding estimates associated with genetically-determined T2D. The observational estimates were adjusted for age, age squared, sex, area, education, household income, smoking, alcohol, self-rated health, and fasting time. The mendelian randomization estimates were adjusted forage, age squared, sex, area, the first 12 principal components, education, smoking, and alcohol. Open boxes denote GRS-T2D. Gray boxes denote GRS-IR. (c), Adjusted hazard ratios (HRs) with 95% CIs of T2D per 1-SD higher protein biomarkers. Models were adjusted for age, age squared, sex, area, education, household income, smoking, alcohol, self-rated health, and fasting time. Within each column, the size of the box was inversely proportional to the variance of the SD difference or logHR. Abbreviation: OB, observational; MR, Mendelian randomisation; T2D, type 2 diabetes; IR, insulin resistance; PC, pancreatic cancer.

### Figure 3. Observational and genetic associations of T2D, IL6, and PC

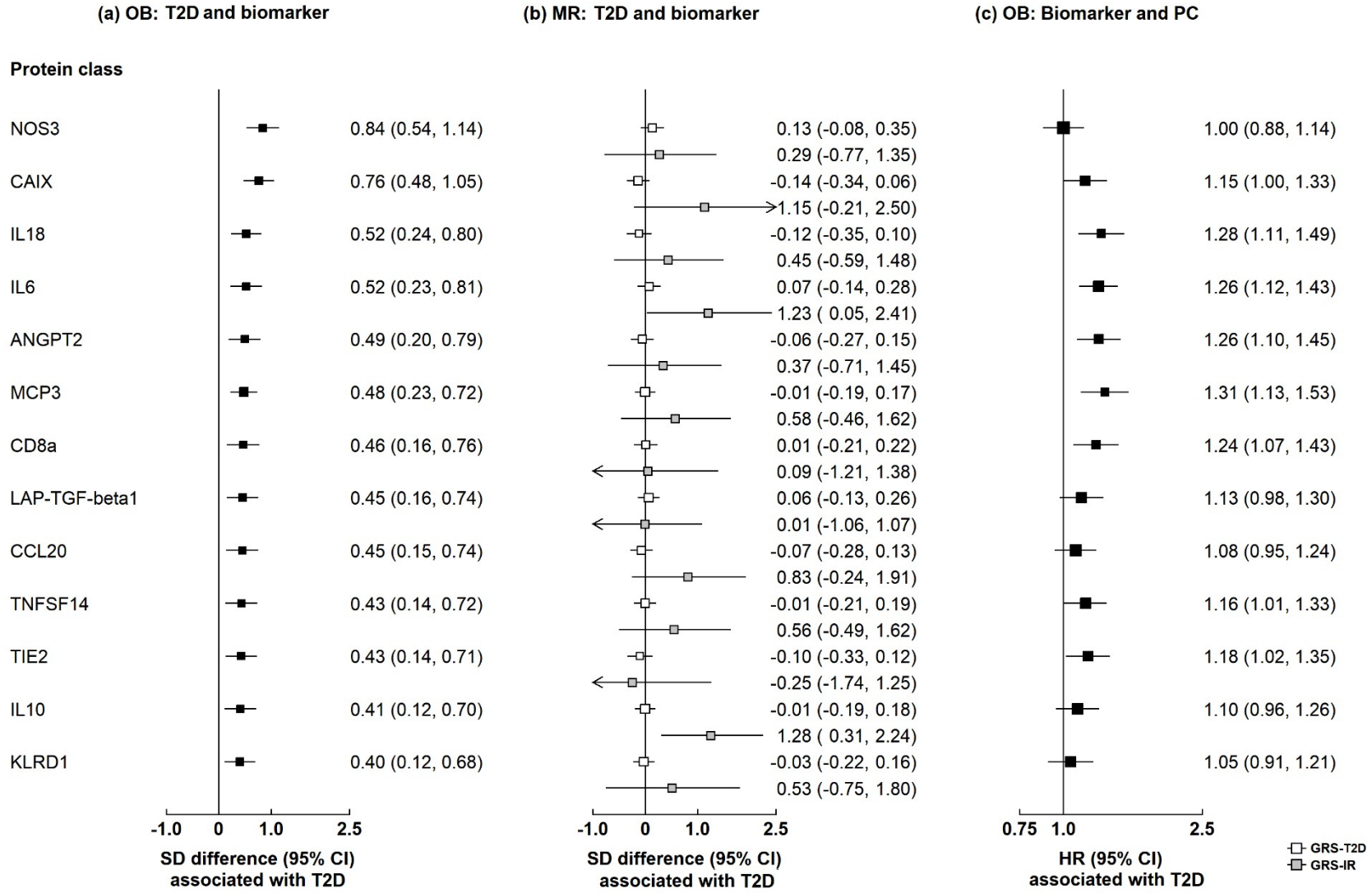
(a), Observational analysis of diabetes, IL6, and PC in CKB. (b), Genetic analysis of T2D,

IL6, and PC using summary-level GWAS statistics in East Asians and Europeans. (c), Genetic analysis of FI, IL6, and PC using summary-level GWAS statistics in East Asians and Europeans. Genetic instruments were obtained from (1) Europeans; (2) BBJ and PanScan; (3) BBJ and PanScan. Abbreviation: T2D, type 2 diabetes; FI, fasting insulin; PC, pancreatic cancer.

**Figure 1. Flow diagram**

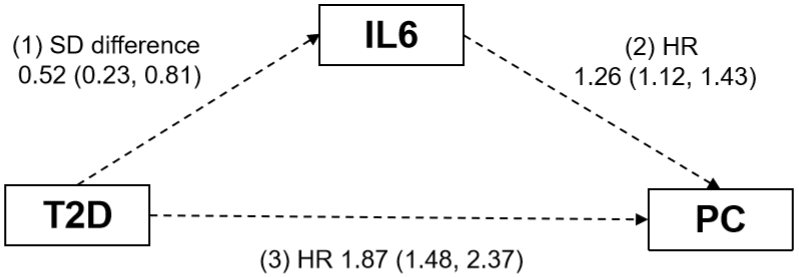


555 **Figure 2. Associations of T2D, proteomics, and PC**
556

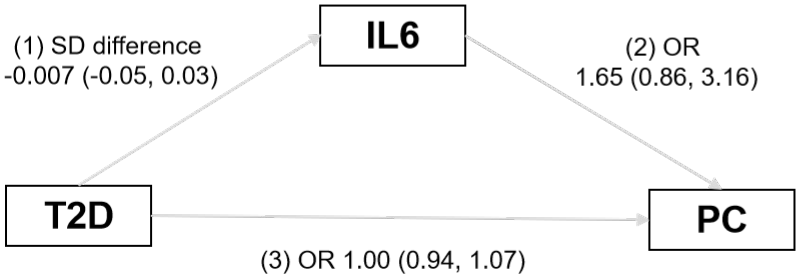


557

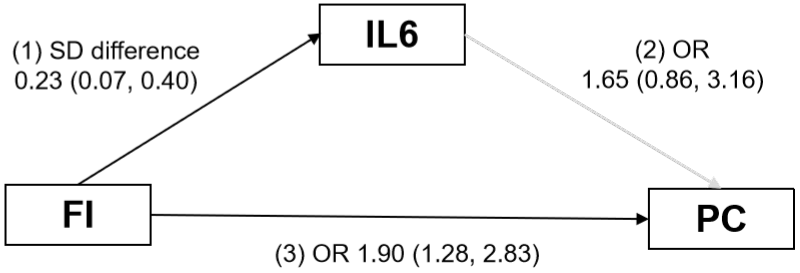
**Figure 3. Observational and genetic associations of T2D, IL6, and PC**




(a) Observational analysis




(b) Genetic analysis: T2D




(c) Genetic analysis: FI