

Association between cancer antigen 19-9 and diabetes risk: A prospective and Mendelian randomization study

Zhaoyang Li¹, Jing Wang¹, Xu Han¹, Fei Wang¹, Hua Hu¹, Jing Yuan¹, Ping Yao¹, Sheng Wei², Huan Guo¹, Dan Zheng¹, Yuhan Tang¹, Handong Yang³, Meian He^{1*}

¹Department of Occupational and Environmental Health, State Key Laboratory of Environmental Health for Incubating, ²Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, and ³Dongfeng Central Hospital, Dongfeng Motor Corporation and Hubei University of Medicine, Shiyan, Hubei, China

Keywords

Cancer antigen 19-9, Diabetes mellitus, Mendelian randomization

*Correspondence

Meian He
Tel: +86-27-8365-7914
Fax: +86-27-8369-2560
E-mail address:
hemeian@gmail.com

J Diabetes Investig 2020; 11: 585–593

doi: 10.1111/jdi.13166

ABSTRACT

Aims/Introduction: Elevated serum cancer antigen 19-9 (CA19-9) levels have been found in diabetes patients in most observational studies; however, whether there is a causal association between CA19-9 and diabetes mellitus is unclear.

Materials and Methods: Our study was carried out based on the Dongfeng-Tongji cohort comprising 27,009 individuals. We first investigated the associations between serum CA19-9 levels and incident diabetes mellitus risk in a prospective cohort study (12,700 individuals). Then, we explored the potential causal relationship between CA19-9 and diabetes mellitus risk in a cross-sectional study (3,349 diabetes mellitus patients and 8,341 controls) using Mendelian randomization analysis. A weighted genetic risk score was calculated by adding the CA19-9 increasing alleles in five single-nucleotide polymorphisms (rs17271883, rs3760776 and rs3760775 in *FUT6*, rs11880333 in *CA11*, rs265548 in *B3GNT3*, and rs1047781 in *FUT2*), which were identified in a previous genome-wide association study on serum CA19-9 levels.

Results: In the prospective study, a total of 1,004 incident diabetes mellitus patients were diagnosed during a mean 4.54-year follow-up period. Elevated serum CA19-9 level was associated with a higher incident diabetes risk after adjustment for confounders, with a hazard ratio of 1.20 (95% confidence interval 1.11–1.30) per standard deviation (12.17 U/mL) CA19-9 increase. Using the genetic score to estimate the unconfounded effect, we did not find a causal association of CA19-9 with diabetes risk (odds ratio per weighted CA19-9-increasing allele: 0.99, 95% confidence interval 0.94–1.04; $P = 0.61$).

Conclusions: The present study did not support a causal association of serum CA19-9 with diabetes risk. CA19-9 might be a potential biomarker of incident diabetes mellitus risk.

INTRODUCTION

Diabetes mellitus is a major chronic disease worldwide, and the prevalence of diabetes in adults is expected to reach 10.4% by 2040¹. In China, the estimated prevalence of total diagnosed and undiagnosed diabetes is 10.9% among adults². Accumulating evidence suggests that islet inflammation might be related to the early pathogenesis of diabetes^{3,4}. However, the relationship between pancreatic cancer and diabetes is complicated and

has not been clear^{5,6}. Even though studies showed that long-term diabetes mellitus is a risk factor for pancreatic cancer, an increasing amount of evidence reminds us that pancreatic cancer might be a diabetogenic factor^{5,7,8}.

Cancer antigen 19-9 (CA19-9), a member of the Lewis antigen family, is expressed in some tissues, such as pancreatic and biliary ductal cells, in very small amounts in the normal human body. A higher level of CA19-9 has often been used as a clinical tumor marker of pancreatic cancer⁹. Elevated serum CA19-9 levels were observed not only in malignant tumors, but also

Received 28 August 2019; revised 23 October 2019; accepted 25 October 2019

in patients with inflammatory conditions, including pancreatitis⁹ and diabetes mellitus^{10–12}. A positive association between high serum CA19-9 levels with the degree of impaired glucose regulation has also been reported¹³. In addition, a decrease of serum CA19-9 levels was observed in patients undergoing diabetes treatment¹⁴. Recently, a prospective study with 2,391 participants found that higher levels of serum CA19-9 were significantly correlated with incident diabetes among the middle-aged and elderly Chinese¹⁵. However, some studies did not support a correlation between CA19-9 and glycated hemoglobin (HbA1c) levels in diabetes patients¹⁶ or find an elevated level of CA19-9 in diabetes patients¹⁷. In addition, an early study showed that insulin secreted by islets might promote the secretion of CA19-9 through a paracrine pathway¹⁸. However, results from an animal study showed that insulin might partly be responsible for the activities of galactosyltransferase, which can transfer galactose to N-acetylglucosamine during the process of CA19-9 biosynthesis¹⁹. This contradiction reminds us that whether CA19-9 levels are associated with diabetes risk and if this association is causal or not still remain to be elucidated. The Mendelian randomization analysis, which makes full use of the characteristics of natural randomization of genetic variants independent of confounding factors²⁰, might allow us to clarify the causal associations of serum CA19-9 levels and diabetes mellitus risk.

In the current study, we investigated the association between serum CA19-9 levels and incident diabetes mellitus risk using the data of the Dongfeng-Tongji cohort (DFTJ cohort) study including 27,009 individuals. Furthermore, we used the principle of Mendelian randomization and five CA19-9 related single-nucleotide polymorphisms (SNPs), which were identified in our previous genome-wide association study (GWAS)²¹, as a proxy of serum CA19-9 to examine whether serum CA19-9 levels are causally related to diabetes mellitus risk.

METHODS

Study population

The DFTJ cohort is a prospective cohort study launched in Shiyang City in the Hubei province of China, and detailed information of this cohort study has been described previously²². Briefly, a total of 27,009 retired workers of the Dongfeng Motor Corporation were recruited between September 2008 and June 2010.

Participants filled out questionnaire information, underwent physical measurements and clinical examinations, and provided fasting blood samples at baseline. Finally, a total of 25,987 individuals completed the first follow-up period from June 2013 to October 2013.

In the observational analyses, 1,387 individuals died during the follow-up period. We excluded participants who were lost at the first follow-up visit ($n = 1,031$); and also those with diabetes, coronary heart disease, stroke in consideration of the commonality in risk factors, and the close relationship among coronary heart disease, stroke and diabetes; patients with tumor

($n = 8,358$), hepatobiliary diseases and hepatitis ($n = 4,671$); and patients with abnormal serum levels (higher than twice the upper limit of the normal level) of aspartate transaminase (AST; >80 U/L), alanine transaminase (ALT; >80 U/L) and total bilirubin (TB; >35 $\mu\text{mol/L}$; $n = 183$) at baseline. Patients with missing or abnormal data (>300 U/mL) on baseline CA19-9 levels ($n = 66$) were also excluded. A total of 12,700 participants (5,485 men and 7,215 women) were included in the final analyses (Figure 1).

Among the whole DFTJ study population, genotype data were available for 13,626 individuals. The Mendelian randomization analyses were based on a case-control study design. After exclusion of participants with missing data on genetic risk score (GRS; $n = 1,100$) and participants with self-reported tumor at baseline or during the follow-up period ($n = 836$), 11,690 participants remained for the Mendelian randomization analyses. The cases included both prevalent cases and incident cases ($n = 3,349$), whereas the controls were the rest of the population ($n = 8,341$).

Ethics approval and consent to participate

The study protocol and informed consent procedure were approved by the Ethics and Human Subject committee of the

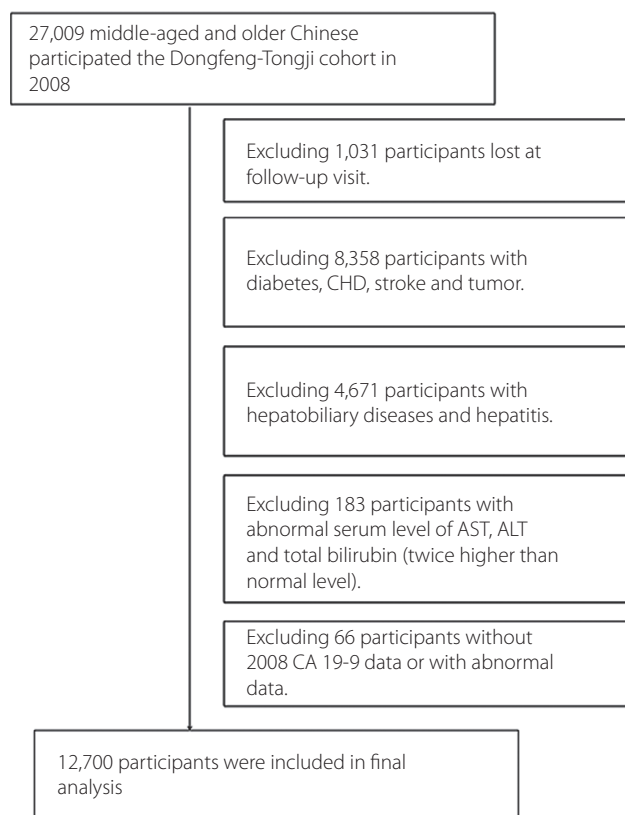


Figure 1 | Flowchart of the participants excluded from the present cohort study.

School of Public Health, Tongji Medical College, Huazhong University of Science and Technology and Dongfeng General Hospital, Dongfeng Motor Corporation. The study was carried out in compliance with the Declaration of Helsinki, and all participants signed the informed consent form.

Assessment of covariates

Baseline information, including sociodemographic characteristics and lifestyle habits, were gathered by trained investigators who administered the interview questionnaire. The physical examinations were also carried out at the same time.

Blood samples were collected after an overnight fast. Levels of serum CA19-9, fasting blood glucose (FBG) and the HbA1c levels were measured at the hospital's laboratory following standard laboratory procedures.

Ascertainment of baseline and incident diabetes

The definition of diabetes mellitus was based on two criteria from the American Diabetes Association²³ including FBG level ≥ 7.0 mmol/L or HbA1c level $\geq 6.5\%$. Furthermore, those with self-report of physician's diagnosis of diabetes or use of diabetes medication were also defined as having diabetes. Because we did not measure the HbA1c levels at baseline, the diabetes mellitus diagnosis at baseline was based on the FBG level ≥ 7.0 mmol/L, self-report of physician's diagnosis of diabetes or use of diabetes medication. Eventually, a total of 1,004 people were diagnosed with incident diabetes.

Genotyping and construction of the genetic score

The selected SNPs were previously reported to be associated with serum CA19-9 levels in our GWAS²⁰. Six SNPs (rs1047781, rs17271883, rs3760776, rs3760775, rs11880333 and rs265548) mapping to four genes (*FUT2*, *FUT6*, *CA11* and *B3GNT3*) were genotyped. Among the participants with genotype data, 1,452 individuals were genotyped with Affymetrix Genome-Wide Human SNP Array 6.0 chips (Santa Clara, CA, USA); 6,156 individuals were genotyped with Illumina Human OmniZhongHua-8 chips (San Diego, CA, USA). Another 6,018 individuals were genotyped with the iPLEX system (Sequenom, San Diego, CA, USA) and/or the TaqMan assay (Applied Biosystems, Foster, CA, USA) in the 384-well format. No SNPs were in linkage disequilibrium with each other (all $r^2 < 0.2$ and $D' < 0.6$; Table S1), and all of the six SNPs were in Hardy–Weinberg equilibrium.

The alleles were coded 0, 1 or 2 according to the number of CA19-9-raising alleles. A simple GRS was calculated by summing the number of risk alleles of the remaining five SNPs. Considering that effect sizes of each SNP were different, we further calculated a weighted GRS by weighing the individual SNPs by their effects on serum CA19-9 levels using estimates from the published GWAS study²¹.

Statistical analysis

The multivariate adjusted hazard ratios (HRs) and 95% confidence intervals (CIs) were computed for incident diabetes using the Cox proportional hazards regression model to evaluate the

relationship between serum CA19-9 levels and incident diabetes risk. In multivariate model 1, adjusted variables were age and sex. Then, we further adjusted body mass index (BMI), smoking status, alcohol consumption status, education, physical activity, hypertension, hyperlipidemia and family history of diabetes in model 2. We ran sensitivity analysis by exclusion of the incident diabetes cases diagnosed during the first 2 years of follow up. To eliminate the influence of the hepatobiliary diseases and tumor, we excluded participants who developed hepatobiliary diseases or tumor during the follow-up period. Furthermore, we excluded participants with CA19-9 in the higher than normal range (>37 U/mL; $n = 284$) from the sensitivity analyses. Finally, considering that serum CA19-9 might be associated with liver function, we further adjusted for AST and ALT levels in the multivariate model.

Associations of individual SNPs with clinical parameters were examined with Spearman's correlation test. We also investigated the associations of the CA19-9-related weighted GRS with potential confounders using linear regression and logistic regression for the continuous traits and dichotomous confounders, respectively.

The associations of individual SNP with serum CA19-9 levels were assessed with a linear regression model. A logistic regression model was used to examine the association between the SNPs/GRS and diabetes risk. To examine whether the CA19-9 levels were a mediator of associations of individual SNP or GRS with diabetes mellitus risk, we further adjusted for CA19-9 levels in the multivariate model. The expected effect size (β_E) of the individual SNP or GRS on diabetes mellitus risk was obtained by multiplying β_{GB} (effect sizes of individual SNP or GRS on CA19-9 levels) and β_{BD} (the effect sizes of CA19-9 on diabetes mellitus risk)²⁴. Then, we used the Student's *t*-test to detect the differences between expected effects sizes (β_E) and observed effect sizes (β_o)²⁵.

The simple median-based method, weighted median based, inverted variance weighted and MR-Egger method were used to verify the causal relationship between CA19-9 and diabetes risks^{26–28}. Furthermore, MR-Egger regression was carried out to test the potential bias from pleiotropy²⁹. Analyses were carried out using SPSS software (version 22.0; SPSS Inc., Chicago, IL, USA) software and R (version 3.2.3). A two-sided *P*-value < 0.05 was considered statistically significant.

RESULTS

Baseline characteristics of the study population in the observational analyses

Baseline characteristics of the participants according to the quartiles of serum CA19-9 levels are summarized in Table 1. During 57,616.23 person-years of follow up, we identified 1,004 incident diabetes cases. Among the 12,700 participants, 43.2% were men and the mean age was 62.19 years. Compared with participants in quartile (Q)1, individuals with higher levels of CA19-9 were more likely to be older, men, physically active, with higher levels of total cholesterol, high-density lipoprotein

Table 1 | Baseline characteristics of the study population

	Cancer antigen 19-9				P trend
	Q1	Q2	Q3	Q4	
Variables	<3.67	3.67–7.45	7.45–14.36	>14.36	
Participants	3,176	3,177	3,172	3,175	
Men (%)	40.2	44.6	45.7	42.3	0.063
Age (years)	61.66 (7.45)	61.99 (7.43)	62.41 (7.99)	62.71 (8.01)	<0.001
Education (%)					
Primary or below	24.1	29.2	27.5	28.7	0.187
Junior high school	36.8	37.9	36.2	35.9	
High school	27.9	24.2	25.3	24.9	
College or above	11.3	8.7	11	10.4	
Body mass index (kg/m ²)	24.16 (3.33)	24.19 (3.23)	23.90 (3.21)	23.73 (3.31)	<0.001
Current smoking, yes (%)	18.3	19.9	20.1	17.5	0.477
Current drinking, yes (%)	22.3	24.9	24.3	21	0.159
Physical activity, yes (%)	87.3	89.7	88.6	88.6	0.284
Family history of diabetes, yes (%)	5.1	3.9	3.6	4.1	0.071
Hypertension, yes (%)	44.7	45.1	46.8	47	0.032
Hyperlipidemia, yes (%)	36.4	40.8	37.9	39.2	0.151
Incident diabetes, yes (%)	8.2	7.6	8.5	8.3	0.739
Systolic blood pressure (mmHg)	127.47 (18.83)	127.47 (18.37)	128.46 (18.73)	128.30 (18.95)	0.02
Diastolic blood pressure (mmHg)	78.02 (10.78)	77.36 (10.75)	77.84 (10.89)	77.59 (10.88)	0.343
Total cholesterol (mmol/L)	5.21 (0.92)	5.15 (0.92)	5.19 (0.92)	5.22 (0.90)	0.868
Triglycerides (mmol/L)	1.39 (0.87)	1.33 (0.83)	1.35 (0.95)	1.36 (1.13)	0.599
HDL cholesterol (mmol/L)	1.48 (0.42)	1.47 (0.43)	1.49 (0.42)	1.51 (0.42)	0.001
LDL cholesterol (mmol/L)	2.99 (0.95)	3.01 (0.77)	3.02 (0.77)	3.05 (0.81)	0.005
Fasting blood glucose (mmol/L)	5.49 (0.57)	5.53 (0.57)	5.52 (0.58)	5.52 (0.58)	0.194
Alanine transaminase (U/L)	22.22 (10.63)	22.59 (10.92)	22.22 (10.81)	21.98 (10.16)	0.224
Aspartate transaminase (U/L)	23.89 (7.42)	24.34 (7.71)	24.42 (7.52)	24.90 (8.03)	<0.001
Total bilirubin (μmol/L)	13.60 (4.56)	13.78 (4.95)	14.02 (4.92)	13.73 (4.85)	0.135
Cancer antigen 19-9 (U/mL)	1.47 (1.31)	5.39 (1.08)	10.37 (1.94)	25.86 (15.63)	–

Total $n = 12,700$. Continuous variables were presented as mean (SD). Dichotomous variables were presented as n (%). HDL, high-density lipoprotein; LDL, low-density lipoprotein; Q1, quartile 1; Q2, quartile 2; Q3, quartile 3; Q4, quartile 4.

(HDL) cholesterol, AST and TB. Also, participants with higher levels of CA19-9 were more likely to be non-current smokers and non-current drinkers, and with lower BMI. The percentage of hyperlipidemia in Q4 participants was higher, whereas the percentage of family history of diabetes mellitus was lower in contrast to that in Q1 participants. As CA19-9 levels increased, the age, levels of HDL cholesterol, low-density lipoprotein cholesterol, AST and systolic blood pressure increased; in contrast, BMI decreased (all P trend <0.05).

Association of serum CA19-9 levels and incident diabetes risk

In the observational analysis, compared with participants in Q1, the HRs and 95% CIs of incident diabetes for individuals in Q2, Q3 and Q4 were 1.05 (0.87–1.26), 1.25 (1.05–1.50) and 1.28 (1.07–1.54), respectively, after adjustment for potential confounders (P for trend = 0.003). When CA19-9 entered the model as a continuous variable, the fully adjusted HR was 1.12 (95% CI 1.06–1.18) per standard deviation (12.17 U/mL) increase. When type 2 diabetes was defined without HbA1c

levels as criteria, the results did not alter materially (Table S2). When we excluded the participants who were diagnosed with diabetes during the first 2 years of follow up, or diagnosed with hepatobiliary diseases or tumor during the follow-up period, similar associations were obtained (Table 2). We further excluded participants with CA19-9 higher than the normal range and similar results were obtained (Table S3). Moreover, further adjusted for AST and ALT in the multivariate model the association attenuated, but still remained with a per standard deviation CA19-9 increase, the incident diabetes risk increased by 8% (95% CI 1.01–1.15; $P = 0.028$; Table S3).

General characteristics of the study population in the Mendelian randomization analyses

Characteristics of the study population in the Mendelian randomization analysis are shown in Table 3. A total of 11,690 individuals (3,349 diabetes patients and 8,341 controls) were included. Compared with participants without diabetes, diabetes patients were more likely to be older, men and non-current

Table 2 | Multivariate adjusted hazard ratios with 95% confidence intervals for incident diabetes mellitus according to quartiles of cancer antigen 19-9

	Cancer antigen 19-9 (U/mL)				P-value*	HR (95% CI) for per-SD	P-value per-SD
	Q1 <3.67	Q2 3.67–7.45	Q3 7.45–14.36	Q4 >14.36			
No. patients/person-years	255/14,541.18	235/14,425.53	260/14,336.27	254/14,313.25			
Model 1	Ref	1.17 (0.98, 1.40)	1.31 (1.09, 1.56)	1.29 (1.08, 1.54)	0.008	1.09 (1.04, 1.15)	0.001
Model 2	Ref	1.05 (0.87, 1.26)	1.25 (1.05, 1.50)	1.28 (1.07, 1.54)	0.003	1.12 (1.06, 1.18)	<0.001
Sensitivity analysis [†]							
Model 3 [‡]	Ref	1.08 (0.89, 1.31)	1.29 (1.07, 1.57)	1.32 (1.09, 1.59)	0.003	1.12 (1.05, 1.18)	<0.001
Model 4 [§]	Ref	1.06 (0.84, 1.34)	1.27 (1.01, 1.59)	1.33 (1.06, 1.66)	0.008	1.13 (1.05, 1.20)	<0.001
Model 5	Ref	1.04 (0.86, 1.26)	1.22 (1.02, 1.47)	1.29 (1.06, 1.55)	0.003	1.13 (1.07, 1.19)	<0.001

Model 1: adjusted for age and sex. Model 2: further adjusted for age, sex, body mass index, smoke, drink, physical activity, education, hypertension, hyperlipidemia and family history of diabetes. *P-value when assigning the median value to each quartile and entered as a continuous variable in the models. †Model 2 was used in sensitivity analysis. ‡Excluding participants who developed diabetes during the first 2 years of follow up (n = 180). §Excluding participants who developed hepatobiliary diseases during the follow up (n = 4072). ||Excluding participants who developed tumor during the follow up (n = 433).

drinkers. Also, the diabetes patients were more likely to have higher levels of BMI, systolic blood pressure, diastolic blood pressure, total cholesterol, triglycerides, low-density lipoprotein cholesterol, FBG and AST levels, and lower HDL cholesterol levels. The percentage of family history of diabetes in patients was higher in contrast to that in controls.

Association of individual SNP and GRS with potential confounders or mediators

We did not find any association between individual SNPs with clinical parameters, except for rs1047781 and rs3760775. The rs1047781 was significantly and negatively associated with FBG levels (P < 0.05; Table S4). The rs3760775 was significantly and negatively associated with total cholesterol, HDL cholesterol, low-density lipoprotein cholesterol and TB levels (all P < 0.05; Table S4). For SNP rs3760775, we carried out a sensitivity analysis in the following analysis by excluding SNP rs3760775 from the genetic score. For the GRS, except for the TB concentrations (β = -0.052, 95% CI -0.104, -0.001; P = 0.04), no associations were found between GRS and any other potential confounders or mediators (Table S5).

Associations of individual SNPs and GRS with serum CA19-9 levels and diabetes risk

As Table 4 showed, the effect sizes of individual SNPs on serum CA19-9 levels ranged from 0.05 to 0.40 U/mL. Each additional CA19-9-increasing allele in the GRS was associated with 0.12 (95% CI 0.11–0.14) U/mL CA19-9 (P < 0.001). To investigate the causal association of CA19-9 levels and diabetes risk, we further explored the associations of the six SNPs with diabetes risk. Among the six SNPs, rs17271883 (OR 1.05, 95% CI 0.99–1.11; P = 0.12) and rs3760776 (OR 1.09, 95% CI 0.99–1.19; P = 0.05) were associated with an increased risk of diabetes without statistical significance. Further adjustment for CA19-9 in the model slightly attenuated the association. The remaining four SNPs tended to be associated with a decreased risk of diabetes, which was directionally opposite to what would be expected based on their CA19-9-increasing effect (OR ranged 0.91–0.96; P-value ranged 0.005–0.36). Similarly, the GRS was not associated with diabetes risk (OR 0.99, 95% CI 0.94–1.04; P = 0.61; Table 4). The associations of rs3760775, rs265548, rs1047781 and GRS with the diabetes risk were different from the expected associations based on the observed associations between these SNPs and CA19-9 levels, and the association between CA19-9 levels and diabetes mellitus (P-value range <0.001–0.029; Table 4). We then carried out sensitivity analysis by excluding SNP rs3760775 from GRS, and the result showed that the associations of GRS with the diabetes mellitus risk were not different from the expected associations based on the observed associations between these SNPs and CA19-9 levels, and the association between CA19-9 levels and diabetes mellitus (P = 0.45; Table 4). Sensitivity analyses were also carried out with the unweighted GRS and we obtained similar results (Tables S6,S7).

Table 3 | General characteristics of the study population in the Mendelian randomization study

Variable	Cases	Controls	P-value
Participants	3,349	8,341	
Age (years)	64.42 (7.38)	62.99 (7.82)	<0.001
Men (%)	48.8	45.4	0.001
Body mass index (kg/m ²)	25.57 (3.35)	24.13 (3.28)	<0.001
Current smoking, yes (%)	18.0	19.1	0.09
Current drinking, yes (%)	19.7	22.6	<0.001
Physical activity, yes (%)	89.1	88.5	0.18
Family history of diabetes, yes (%)	7.8	3.7	<0.001
Systolic blood pressure (mmHg)	133.68 (18.68)	127.87 (18.49)	<0.001
Diastolic blood pressure (mmHg)	78.35 (11.2)	77.39 (10.91)	<0.001
Total cholesterol (mmol/L)	5.25 (0.99)	5.17 (0.94)	<0.001
Triglycerides (mmol/L)	1.71 (1.31)	1.33 (0.81)	<0.001
HDL cholesterol (mmol/L)	1.36 (0.39)	1.44 (0.38)	<0.001
LDL cholesterol (mmol/L)	3.07 (0.83)	3.02 (0.85)	0.001
Fasting blood glucose (mmol/L)	7.47 (2.61)	5.51 (0.54)	<0.001
Alanine transaminase (U/L)	27.58 (21.13)	23.54 (20.19)	<0.001
Aspartate transaminase (U/L)	25.87 (17.65)	25.37 (15.82)	0.15
Total bilirubin (μmol/L)	14.25 (5.64)	14.37 (5.66)	0.29
CA19-9 (U/mL)	13.22 (19.22)	11.21 (16.18)	<0.001

Total $n = 11,690$. Continuous variables were presented as mean (standard deviation). Dichotomous variables were presented as n (%). HDL, high-density lipoprotein; LDL, low-density lipoprotein; Q1, quartile 1; Q2, quartile 2; Q3, quartile 3; Q4, quartile 4.

In sensitivity analyses, we used four different methods (the simple median-based, weighted median-based, inverse-variance weighted and MR-Egger method) to estimate the causal effect of CA19-9 on the risk of diabetes. The results consistently showed a non-causal association between CA19-9 and diabetes (All $P > 0.05$; Table S8), except for the weighted median-based method, which yielded suggestive evidence of a negative association between CA19-9 and diabetes ($P = 0.03$). We further tested whether any of the selected SNPs were influenced by pleiotropy. Using the MR-Egger method, the beta coefficient of the MR-Egger regression provided pleiotropy-corrected causal estimates, and an intercept distinct from the origin provided evidence for pleiotropic effects. We found that the intercept term estimated from MR-Egger regression was centered at the origin with a confidence interval including the null (0.015, 95% CI $-0.092, 0.121$; $P = 0.79$; Table S8), suggesting the results were not influenced by pleiotropy.

DISCUSSION

The present large prospective study found that higher serum CA19-9 levels were positively associated with higher incident diabetes risk, with a 12% higher diabetes risk per standard deviation (12.17 U/mL) of CA19-9 increase. However, the Mendelian randomization analysis provided did not support the causal relevance of serum CA19-9 levels to diabetes risk.

The findings of the observational analysis are in line with most of the previous studies including cross-sectional^{10,13} and prospective studies¹⁵. The earliest study carried out by Nakamura *et al.* found that the level of CA19-9 increased in patients

with diabetes, especially those with poorly controlled conditions or complications, when compared with healthy participants³⁰. After that, an increasing number of studies found a positive correlation between CA19-9 and diabetes status^{10,31} or parameters related to glycemic control³². Recently, a prospective study¹⁵ carried out among the middle-aged and elderly Chinese population found that serum CA19-9 levels were significantly associated with an increased risk of incident diabetes mellitus. In the current prospective longitudinal study with a large sample size, the findings showed that serum CA19-9 levels were independently associated with an increased incident risk of diabetes.

Although the positive association between serum CA19-9 levels and diabetes risk were found in the present longitudinal cohort study and most of the previous studies, whether the positive associations were causal or not still remained to be examined. Mendelian randomization analysis can provide information on the potential causal relationship between biomarkers and diseases, because the natural randomization of genetic variants that make the genetic variants can be free from confounding factors³³. When we explore evidence for causality using the Mendelian randomization analysis, certain rules must be followed. First, the instrumental variable must be associated with the risk factor of interest. In the present study, all selected SNPs were strongly associated with serum CA19-9 levels and no SNPs were in linkage disequilibrium with each other. Meanwhile, the serum CA19-9 levels were associated with increased diabetes risk; therefore, the six variants of CA19-9 could serve as instruments in a Mendelian randomization study. Second,

Table 4 | Association of the genetic risk score with cancer antigen 19-9 levels and diabetes mellitus risk

SNP	Gene	MAF	EAF	CA19-9 increasing allele/other allele	Effect size on CA19-9 levels		Observed association with DM		Expected association with DM		P-value**	Observed association with DM adjusted for CA19-9 levels
					β (95% CI)	P-value*	OR (95% CI)	P-value*	OR (95% CI)	P-value		
rs17271883	FUT6	0.48	0.48	C/T	0.12 (0.09–0.16)	<0.001	1.05 (0.99–1.11)	0.12	1.017 (1.017–1.017)	0.32	1.04 (0.98–1.11)	0.18
rs3760776	FUT6	0.12	0.88	A/T	0.13 (0.07–0.18)	<0.001	1.09 (0.99–1.19)	0.05	1.017 (1.017–1.017)	0.12	1.09 (0.99–1.19)	0.07
rs3760775	FUT6	0.25	0.75	A/T	0.40 (0.36–0.44)	<0.001	0.96 (0.89–1.02)	0.21	1.057 (1.056–1.057)	0.002	0.95 (0.89–1.02)	0.15
rs11880333	CA11	0.23	0.23	T/C	0.16 (0.12–0.19)	<0.001	0.96 (0.90–1.04)	0.36	1.021 (1.021–1.021)	0.13	0.96 (0.88–1.03)	0.21
rs265548	B3GN73	0.24	0.76	T/C	0.05 (0.01–0.09)	0.01	0.94 (0.87–1.00)	0.07	1.007 (1.007–1.007)	0.04	0.92 (0.87–1.00)	0.05
rs1047781	FUT2	0.39	0.39	T/A	0.27 (0.24–0.31)	<0.001	0.91 (0.86–0.97)	0.005	1.038 (1.038–1.039)	<0.001	0.89 (0.84–0.95)	<0.001
Weighted GRS					0.14 (0.13–0.15)	<0.001	0.99 (0.94–1.04)	0.61	1.019 (1.019–1.019)	0.03	0.99 (0.97–1.01)	0.23
Weighted GRS [†]					0.14 (0.13–0.15)	<0.001	0.99 (0.99–1.00)	0.64	1.019 (1.019–1.019)	0.03	0.99 (0.97–1.02)	0.64
Weighted GRS [‡]					0.11 (0.09–0.12)	<0.001	1.021 (0.98–1.03)	0.68	1.014 (1.014–1.014)	0.45	0.99 (0.97–1.02)	0.77

*Adjusted for age, sex, body mass index, smoke, drink, and physical activity. **P-value for difference between expected and observed associations with diabetes mellitus (DM). †Further adjusted for total bilirubin. ‡Excluding the contribution of rs3760775.

the instrumental variable must be independent of potential confounders (confounders in the association between CA19-9 and diabetes). For the individual SNPs reported in the previous GWAS, we did not find any association between individual SNP with potential confounders except for rs3760775, which was significantly and negatively associated with lipid profile and TB levels. However, when we excluded this SNP from the genetic score, the null association of GRS and diabetes risk still remained. For TB, when we additionally adjusted for TB levels in the multivariate model to investigate the instrumental variable estimate of CA19-9 on diabetes risk, the null effect remained ($P = 0.64$). Finally, the selected SNPs and GRS must affect the outcome only through CA19-9. In the present study, the trend of two SNPs showing a positive association with diabetes risk attenuated after further adjusting for CA19-9. However, the lack of statistical significance between the genetic variants and the diabetes risk did not provide support for the causal association between the CA19-9 levels and diabetes risk.

Possible explanations for the discrepant results from observational study and Mendelian randomization analysis might be attributed to the following reasons. On the one hand, a previous study¹⁹ showed that insulin might be related to the increment of the activity of intestinal galactosyltransferase activity involved in the process of CA19-9 biosyntheses³⁴. Also, a negative association between serum CA19-9 and insulin secretion in individuals with prediabetes was found³⁵, indicating that an increase in serum CA19-9 levels might reflect the disorders of insulin secretion among individuals with prediabetes. On the other hand, the small contribution of the GRS to the variation of CA19-9 levels might partly explain the null association. In the present study, the weighted GRS only explained 4.7% of the total variation of CA19-9 levels (data not shown). Therefore, to find more CA19-9-related loci in future studies, larger-scale studies including more CA19-9-related SNPs are required to validate whether there are causal associations of CA19-9 and diabetes risk.

Several limitations of the present study need to be considered. First, although we adjusted for the major lifestyle and other confounders in the analyses, we could not rule out the possible influence of unmeasured or residual confounders. Second, the instrument analysis is a case-control study, and more cohort studies are warranted to validate the present findings. In addition, we carried out a Mendelian randomization analysis in a one-sample setting, which might induce bias in the direction of the confounded association between the exposure and outcome in a finite sample³⁶. Although we used the method of equal weights to reduce the potential weak instrument bias, we still found that the GRS was not associated with diabetes risk (OR 0.96, 95% CI 0.89–1.04). Diverse data sources should be used in future studies. Third, even though we defined diabetes based on physician diagnosis, antidiabetic medications, FBG level and HbA1c level, the postprandial blood glucose in an oral glucose tolerance test was not examined and this might result in misclassification.

Fourth, the analyses were restricted to the middle- and old-aged Chinese population; therefore, the generalizability of the present findings to populations of young people and other ancestries is limited. Finally, the Mendelian randomization approach has its own limitations. For example, we cannot account for some feedback loops that might potentially exist. Also, we cannot completely rule out the potential pleiotropy of the genes we selected.

In conclusion, the present study does not support the hypothesis that circulating CA19-9 has a causal effect on diabetes risk, and CA19-9 might be a potential biomarker of incident diabetes mellitus risk. Further studies with larger sample sizes are warranted to validate these findings.

ACKNOWLEDGMENTS

The authors thank all study subjects for participating in the DFTJ cohort study, as well as all volunteers for assisting in collecting the samples and data. This study was funded by the grants from the National Natural Science Foundation (grants NSFC-81522040 and 81473051), the Program for HUST Academic Frontier Youth Team and the National Key R&D Program of China (2017YFC0907501).

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

- Ogurtsova K, da Rocha Fernandes JD, Huang Y, *et al.* IDF Diabetes Atlas: global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract* 2017; 128: 40–50.
- Wang L, Gao P, Zhang M, *et al.* Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. *JAMA* 2017; 317: 2515–2523.
- Abdulreda MH, Berggren PO. Islet inflammation in plain sight. *Diabetes Obes Metab* 2013; 15: 105–116.
- Ezirik DL, Miani M, Cardozo AK. Signalling danger: endoplasmic reticulum stress and the unfolded protein response in pancreatic islet inflammation. *Diabetologia* 2013; 56: 234–241.
- Andersen DK, Korc M, Petersen GM, *et al.* Diabetes, pancreatogenic diabetes, and pancreatic cancer. *Diabetes* 2017; 66: 1103–1110.
- Basso D, Greco E, Fogar P, *et al.* Pancreatic cancer-associated diabetes mellitus: an open field for proteomic applications. *Clin Chim Acta* 2005; 357: 184–189.
- Pannala R, Leirness JB, Bamlet WR, *et al.* Prevalence and clinical profile of pancreatic cancer-associated diabetes mellitus. *Gastroenterology* 2008; 134: 981–987.
- Pfeffer F, Koczan D, Adam U, *et al.* Expression of connexin26 in islets of Langerhans is associated with impaired glucose tolerance in patients with pancreatic adenocarcinoma. *Pancreas* 2004; 29: 284–290.
- Scara S, Bottoni P, Scatena R. CA 19–9: biochemical and clinical aspects. *Adv Exp Med Biol* 2015; 867: 247–260.
- Uygur-Bayramicli O, Dabak R, Orbay E, *et al.* Type 2 diabetes mellitus and CA 19–9 levels. *World J Gastroenterol* 2007; 13: 5357–5359.
- Gul K, Nas S, Ozdemir D, *et al.* CA 19–9 level in patients with type 2 diabetes mellitus and its relation to the metabolic control and microvascular complications. *Am J Med Sci* 2011; 341: 28–32.
- Shang X, Song C, Du X, *et al.* The serum levels of tumor marker CA19-9, CEA, CA72-4, and NSE in type 2 diabetes without malignancy and the relations to the metabolic control. *Saudi Med J* 2017; 38: 204–208.
- Huang Y, Xu Y, Bi Y, *et al.* Relationship between CA 19–9 levels and glucose regulation in a middle-aged and elderly Chinese population. *J Diabetes* 2012; 4: 147–152.
- Zhang D, Hou W, Liu F, *et al.* Metformin reduces serum CA199 levels in type 2 diabetes Chinese patients with time-effect and gender difference. *Diabetes Technol Ther* 2015; 17: 72–79.
- Du R, Sun W, Lin L, *et al.* Serum CA 19–9 and risk of incident diabetes in middle-aged and elderly Chinese: a prospective cohort study. *Acta Diabetol* 2017; 54: 201–208.
- Banfi G, Ardemagni A, Bravi S, *et al.* Are diabetic metabolic compensation and CA19.9 really correlated? *Int J Biol Markers* 1996; 11: 207–210.
- Ritts RE Jr, Del Villano BC, Go VL, *et al.* Initial clinical evaluation of an immunoradiometric assay for CA 19–9 using the NCI serum bank. *Int J Cancer* 1984; 33: 339–345.
- Williams JA, Goldfine ID. The insulin-pancreatic acinar axis. *Diabetes* 1985; 34: 980–986.
- Lenoir D, Gréco S, Louisot P, *et al.* Implication of insulin and nutritional factors in the regulation of intestinal galactosyltransferase activity during postnatal development. *Metabolism* 2000; 49: 526–531.
- Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; 32: 1–22.
- He M, Wu C, Xu J, *et al.* A genome wide association study of genetic loci that influence tumour biomarkers cancer antigen 19–9, carcinoembryonic antigen and alpha fetoprotein and their associations with cancer risk. *Gut* 2014; 63: 143–151.
- Wang F, Zhu J, Yao P, *et al.* Cohort Profile: the Dongfeng-Tongji cohort study of retired workers. *Int J Epidemiol* 2013; 42: 731–740.
- Standards of medical care in diabetes–2010. *Diabetes Care* 2010; 33: S11–S61.
- Li S, Zhao JH, Luan J, *et al.* Genetic predisposition to obesity leads to increased risk of type 2 diabetes. *Diabetologia* 2011; 54: 776–782.
- Freathy RM, Timpson NJ, Lawlor DA, *et al.* Common variation in the FTO gene alters diabetes-related metabolic

- traits to the extent expected given its effect on BMI. *Diabetes* 2008; 57: 1419–1426.
26. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013; 37: 658–665.
 27. Bowden J, Davey Smith G, Haycock PC, *et al.* Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; 40: 304–314.
 28. Yavorska OO, Burgess S. Mendelian randomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017; 46: 1734–1739.
 29. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; 44: 512–525.
 30. Nakamura N, Aoji O, Yoshikawa T, *et al.* Elevated serum CA19-9 levels in poorly controlled diabetic patients. *Jpn J Med* 1986; 25: 278–280.
 31. Petit JM, Vaillant G, Olsson NO, *et al.* Elevated serum CA19-9 levels in poorly controlled diabetic patients. Relationship with Lewis blood group. *Gastroenterol Clin Biol* 1994; 18: 17–20.
 32. Kim SH, Baek CO, Lee KA, *et al.* Clinical implication of elevated CA 19–9 level and the relationship with glucose control state in patients with type 2 diabetes. *Endocrine* 2014; 46: 249–255.
 33. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007; 16: 309–330.
 34. Shiraishi N, Natsume A, Togayachi A, *et al.* Identification and characterization of three novel beta 1,3-N-acetylglucosaminyltransferases structurally related to the beta 1,3-galactosyltransferase family. *J Biol Chem* 2001; 276: 3498–3507.
 35. Shimodaira M, Niwa T, Nakajima K, *et al.* The relation between CA 19–9 level and early-phase insulin secretion in normoglycemic and prediabetic subjects. *Int J Biol Markers* 2015; 30: e169–73.
 36. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016; 40: 597–608.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 | Matrix of r^2 and D' for each pair of the six single-nucleotide polymorphisms.

Table S2 | Multivariate adjusted hazard ratios with 95% confidence intervals for incident diabetes according to quartiles of cancer antigen 19-9.

Table S3 | Multivariate adjusted hazard ratios with 95% confidence intervals for incident diabetes according to quartiles of cancer antigen 19-9.

Table S4 | Correlation analysis between individual single-nucleotide polymorphisms with clinical parameters at baseline.

Table S5 | Association of the weighted genetic risk score with potential confounders and glycemic traits.

Table S6 | Association of the unweighted genetic risk score with potential confounders and glycemic traits.

Table S7 | Association of the unweighted genetic risk score with cancer antigen 19-9 levels and diabetes risk.

Table S8 | Mendelian randomization of cancer antigen 19-9 and diabetes mellitus.