BMC Medicine

RESEARCH

Open Access

Exploring the pathways linking fasting insulin to coronary artery disease: a proteome-wide Mendelian randomization study

Xin Huang[1] and Jie V. Zhao[1,2*]

Abstract

Background Insulin is known to be associated with a higher risk of coronary artery disease (CAD), but molecular mechanisms remain unclear. This study aimed to explore protein-mediated pathways linking fasting insulin to CAD using Mendelian randomization (MR).

Methods This MR study examined the association between fasting insulin and CAD using genome-wide association study (GWAS) data from MAGIC and CARDIoGRAMplusC4D. To investigate underlying mechanisms, a two-step proteome-wide MR analysis was conducted. First, associations of fasting insulin with 2940 circulating proteins were assessed using GWAS of proteomics from UKB-PPP. Proteins affected by insulin were then analyzed for their association with CAD risk. Proteins selected in both steps were considered as potential mediators. Sensitivity analyses to test whether associations are robust to pleiotropy and replication using other GWAS data, including GWAS of proteomics from deCODE and GWAS of CAD from FinnGen Biobank, were performed.

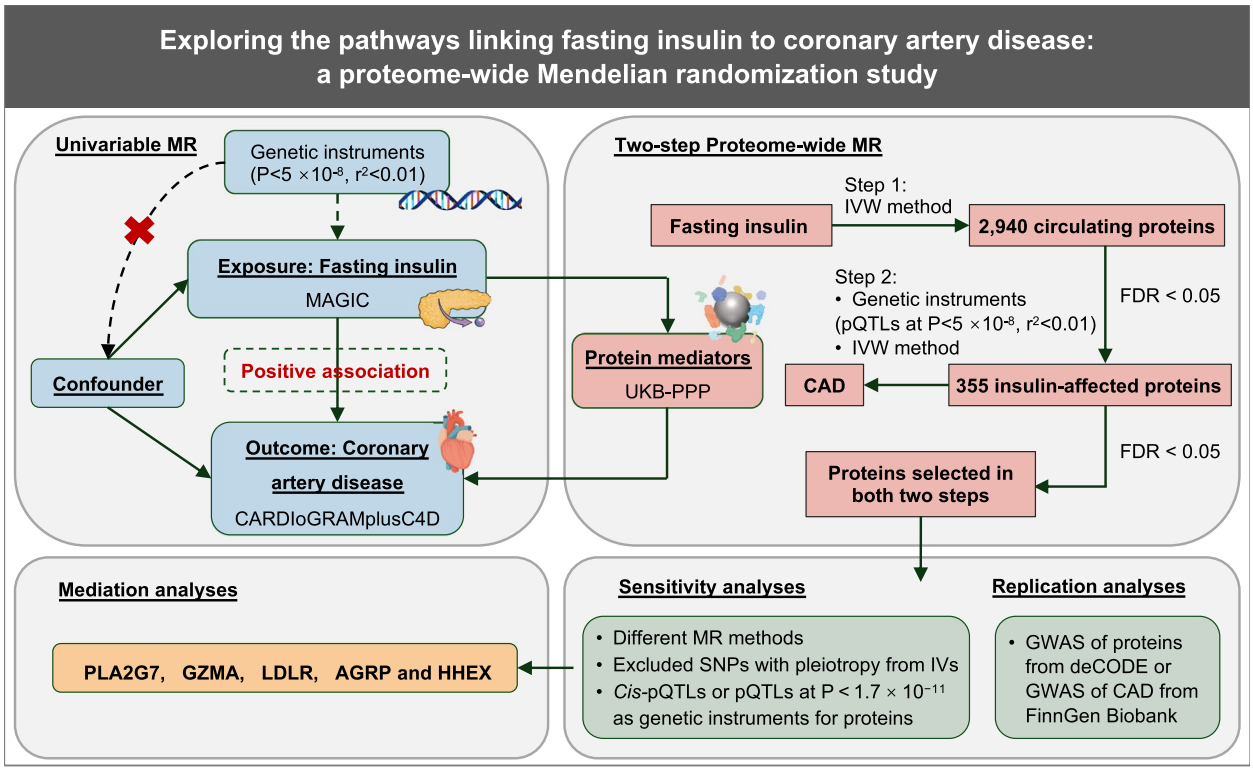
Results Genetically predicted insulin was associated with a higher risk of CAD (odds ratio 1.79, 95% confidence interval 1.34 to 2.40). At a false discovery rate of 0.05, insulin affected 355 proteins, ten of which were both increased by insulin and linked to a higher risk of CAD. After sensitivity and replication analyses, PLA2G7, GZMA, LDLR, AGRP, and HHEX were identified as reliable mediators. Mediation analyses using non-pleiotropic instruments showed that PLA2G7, GZMA, LDLR, and AGRP explained 19.50%, 6.91%, 19.31%, and 29.66% of insulin's total effect on CAD, respectively.

Conclusions This study identified five protein mediators linking insulin to CAD. These proteins could be considered as potential targets to mitigate insulin-related cardiovascular risk, providing novel insights for drug repurposing.

Keywords Insulin, Proteomics, Coronary artery disease, Mendelian randomization

*Correspondence:
Jie V. Zhao
janezhao@hku.hk
Full list of author information is available at the end of the article

Graphical Abstract



Background

Insulin is a key hormone regulating glucose metabolism. As early as the 1980s, Reaven's hypothesis indicated that hyperinsulinemia arising from insulin resistance contributed to metabolic syndrome and hence involved in the etiology of coronary artery disease (CAD) [1]. Later epidemiologic studies added to the evidence by showing that high fasting insulin concentrations were associated with increased CAD risk and mortality, such as the Busselton study [2], the Helsinki study [3], and the Paris Prospective Study [4].

Despite insulin's recognized role in CAD, the specific mechanisms remain unelucidated. The physiological function of insulin is usually seen in association with its role in maintaining glucose homeostasis, but as a major anabolic hormone, insulin also stimulates various metabolic processes, such as the promotion of protein synthesis [5]. Proteins, which are downstream products of metabolic processes and modifiable targets for dietary or pharmacological interventions, serve as critical biomarkers and potential drug targets for diseases, including CAD [6]. The use of proteins as intermediate phenotypes could offer valuable insights into the mechanistic pathways linking insulin to CAD. However, findings from observational studies are often confounded by environmental factors, limiting robust causal inference.

Mendelian randomization (MR) offers a powerful approach to address this concern. By leveraging genetic variants as instrumental variables which were randomly allocated at conception, MR can infer causal relationships that are less vulnerable to the influence of environmental confounders [7], which are difficult to tackle in traditional observational studies. For example, previous observational studies have the limitation of unmeasured confounding or residual confounding by BMI, health status, or mediation uses [8–10]. Existing MR studies gave supported evidence for the unfavorable associations of insulin with cardiovascular events and its risk factors, such as blood pressure and lipid profile [11–13], implicating insulin's broader pathophysiological effects, but the mechanisms via proteins have not been clarified. The availability of large GWAS of proteins provides us the opportunity to comprehensively examine the mediating role of proteins in the insulin-CAD relationship.

In this study, we first examined the association between insulin and CAD risk using the latest genome-wide

association study (GWAS) data. Then to explore the molecular pathways underlying insulin's effects, we conducted a two-step proteome-wide MR analysis leveraging GWAS of 2940 proteins from the UK Biobank Pharma Proteomics Project (UKB-PPP). We aimed to identify circulating proteins associated with fasting insulin and explore the mediating proteins linking insulin to the risk of CAD.

Methods

The association of genetically predicted fasting insulin with the risk of CAD

We examined the association of fasting insulin on the risk of CAD using two-sample MR by applying genetic instruments predicting insulin to the GWAS of CAD. GWAS summary statistics for fasting insulin were obtained from the Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC) [14]. A total of 151,013 individuals of European ancestry from 69 cohorts contributed to the meta-analysis of insulin levels. Individuals with a physician's diagnosis of diabetes, reported use of diabetes-relevant medication(s), or abnormal levels of fasting plasma glucose, 2-h plasma glucose, or HbA1c were excluded from the analysis. Measures for fasting insulin were taken from whole blood in pmol/l and were corrected to plasma level using the correction factor 1.13 [15]. Genetic associations were obtained from linear regression adjusting for BMI, study-specific covariates (such as age and sex), and principal components. GWAS summary statistics for CAD were obtained from a GWAS meta-analysis in CARDIoGRAMplusC4D (including UK Biobank), which involved 122,733 cases and 424,528 controls, primarily of European ancestry [16]. Prevalence and incidence data on CAD were collected at the Assessment Centre in-patient Health Episode Statistics (HES) using the following ICD 10 codes: I21–I25 and the following Office of Population Censuses and Surveys Classification of Interventions and Procedures, version 4 (OPCS-4) codes: K40–K46, K49, K50, and K75, which includes therapeutic operations on coronary artery. Self-reported CAD was also included. The genetic associations were obtained from logistic regression adjusting for age, sex, the first 30 principal components, and genotyping array.

Genetically predicted fasting insulin was defined as fasting insulin levels proxied by genetic variants that were strongly associated with fasting insulin concentrations identified from GWAS. Specifically, genome-wide significant ($P < 5 \times 10^{-8}$) and uncorrelated ($r^2 < 0.01$) single nucleotide polymorphisms (SNPs) were selected as instrumental variables (IVs) for fasting insulin. We calculated the *F*-statistic using an established approximation to check the strength of each SNP [17]. A cut-off of 10 was used as a "rule of thumb" to distinguish between strong and weak instruments. SNP with an *F*-statistic < 10 was removed from the IVs. In the main analysis, MR estimates were derived from the meta-analysis of the Wald estimates (genetic association with CAD divided by the genetic association with fasting insulin) for each SNP, using inverse variance weighting (IVW) with multiplicative random effects [18].

To account for potential pleiotropy, we conducted several sensitivity analyses. First, we used several MR methods with different assumptions, including MR-Egger, weighted median, MR Robust Adjusted Profile Score (MR-RAPS), and MR Pleiotropy RESidual Sum and Outlier (MR-PRESSO) [19–22]. Second, we examined the genetic associations of each genetic instrument for insulin with the risk of CAD using GWASs from CARDIoGRAMplusC4D, the FinnGen Biobank, and the UK Biobank. We repeated the MR analysis by excluding those SNPs associated with CAD directly at genome-wide significance from IVs. Third, we also checked whether the selected IVs for insulin were associated with potential confounders using GWASs from the UK Biobank, including Townsend index, education, smoking status, alcohol drinking, physical activity, and BMI. We repeated the MR analysis excluding those SNPs associated with any confounders at genome-wide significance from IVs.

Two-step proteome-wide MR analysis to identify protein mediators

Data sources

We obtained genetic associations for 2940 circulating protein levels from the UK Biobank Pharma Proteomics Project (UKB-PPP) [23]. The UK Biobank (UKB) is a large-scale biomedical database with genetic and health data from over 500,000 UK participants aged 40–69, with a median follow-up of 11.1 years [24]. UKB-PPP, involving 54,219 participants from UKB, aimed to expand proteogenomics, creating a comprehensive protein quantitative trait locus (pQTL) database to aid biomarker and drug discovery [23]. Circulating levels of 2940 proteins were measured from blood plasma samples across eight protein panels (2940 available), including cardiometabolic, cardiometabolic II, inflammation, inflammation II, neurology, neurology II, oncology, and oncology II. Genetic associations with these proteins were derived using linear regression, adjusting for age, age^2, sex, age × sex, age^2 × sex, batch, UKB center, UKB genetic array, time between blood sampling and measurement, and the first 20 genetic principal components. We used GWAS data from the discovery cohort which were restricted to participants of European ancestry ($n = 34{,}557$).

Two-step MR analysis

In step 1, we performed a two-sample MR analysis to assess the causal associations of fasting insulin with 2940 proteins using the genetic instruments for insulin obtained from the above analysis. IVW method was used in the main analysis. A false discovery rate (FDR) of 5% was used as the threshold. The results of insulin with 2940 proteins were visualized using volcano plots. In step 2, for proteins affected by fasting insulin identified in step 1, we used *cis*- and *trans*-pQTL variants that were significant ($P < 5 \times 10^{-8}$) and uncorrelated ($r^2 < 0.01$) as genetic instruments for the identified proteins and examined their associations with the risk of CAD. IVW was used in the main analysis. Wald ratio was used if only one SNP was available. MR Steiger directionality test was used to infer the causal direction between exposure and outcome by comparing the total proportion of variance explained by all genetic instruments in each trait. The test was implemented using the "mr_steiger" function in the "TwoSampleMR" package. When instruments account for a greater proportion of variance in the exposure than in the outcome (*P* value for Steiger test < 0.05), this supports a causal direction from exposure to outcome [25].

We used heatmaps to visualize the proteins affected by fasting insulin and the associations between these proteins and CAD risk. Proteins mediating the positive effect of fasting insulin on CAD were identified based on the following criteria: (1) increased by insulin and also elevated CAD risk or decreased by insulin and simultaneously reduced CAD risk and (2) robust in the sensitivity analyses and validation analyses.

Sensitivity analyses

For the proteins met the criteria (1), we conducted several sensitivity analyses. First, we repeated two-step MR analyses using weighted median, MR-PRESSO, MR-RAPS, and MR-Egger methods to detect the directional consistency of effects. Second, we removed SNPs associated with outcome (CAD) directly at genome-wide significance in any of the CARDIoGRAMplusC4D, the FinnGen Biobank, or the UK Biobank and repeated the two-step MR. Third, we removed SNPs associated with Townsend index, education, smoking status, alcohol drinking, physical activity, or BMI in the UK Biobank at genome-wide significance and repeated the two-step MR. Fourth, for protein mediators, we examined the genetic associations of each genetic instrument with all other proteins in the UKB-PPP, using a significance threshold of $P < 1.7 \times 10^{-11}$ ($5 \times 10^{-8}/2923$ unique proteins). We repeated the analyses only using non-pleiotropic pQTLs (i.e., pQTLs associated with no other proteins) as IVs. Additionally, following a previous MR study [26], we also used less stringent criteria (i.e., pQTLs associated with fewer than 5 or 7 proteins) to select IVs. Fifth, we replicated the analyses using *cis*-pQTLs, i.e., SNPs located within ± 500 kb of the gene encoding the protein [6], as IVs for each protein; we also used the more stringent threshold $P < 1.7 \times 10^{-11}$ to select genetic instruments for each protein. Proteins passed sensitivity analyses were included in validation analyses.

Validation analyses

Several validation analyses were performed to check whether the insulin-protein-CAD associations could be replicated. First, to address concerns that adjusting BMI for fasting insulin may lead to possible collider bias, we used genetic instruments for insulin unadjusted for BMI from another MAGIC effort, including 108,557 participants of European ancestry [27], and repeated the MR analyses of insulin with CAD and proteins. Second, we used GWAS summary statistics from different data sources. Specifically, we obtained genetic associations with proteins from deCODE genetics, conducted in 35,559 Icelanders [28]. We also obtained genetic associations with the risk of CAD from FinnGen Biobank, which involved 21,012 CAD cases and 197,780 controls.

Mediation analyses

Mediation analyses were conducted for protein mediators selected from two-step MR and robust in the sensitivity analyses and validation analyses. We used the product of coefficients method to obtain the indirect effect, i.e., the effect of fasting insulin on the risk of CAD via the mediator. Specifically, we multiplied the causal effect of insulin on the protein from step 1 MR with the direct effect of the protein mediator on the CAD derived from multivariable MR using the IVW method. To account for potential pleiotropy, we only used non-pleiotropic pQTLs (i.e., pQTLs associated with no other proteins) as genetic instruments for protein mediators. The delta method was used to calculate the standard error (SE) and 95% CI of the indirect effect [29]. The proportion mediated was calculated by dividing the indirect effect by the total effect.

All statistical analyses were performed using R (Foundation for Statistical Computing, Vienna, Austria;

Table 1 Univariable MR analyses of genetically predicted fasting insulin with the risk of CAD

Method	OR	95% CI	*P* value
IVW	1.7933	(1.34, 2.40)	8.16E − 05
MR-Egger	1.4530	(0.59, 3.56)	4.19E − 01
Weighted median	2.2690	(1.80, 2.86)	2.98E − 12
MR-RAPS	1.9246	(1.42, 2.60)	2.03E − 05
MR-PRESSO	2.1087	(1.68, 2.64)	1.78E − 07

version 4.3.1). MR analyses were performed using the "TwoSampleMR," "MendelianRandomization," "MRPRESSO," and "mr.raps" R packages; figures were created using "forestplot," "ggplot2," "ggVolcano," "ComplexHeatmap," "circlize," "RColorBrewer," and "gridExtra" R packages.

Ethics statement

This study used only published summary data from studies involving human participants, with written informed consent and approval by their respective institutional ethics review committees. No additional ethical approval was required. Written informed consent was obtained from all individual participants for each of the studies included in the analysis and can be found in the original publications.

Results

The effect of fasting insulin on the risk of CAD

Forty-four SNPs were identified as genetic instruments for fasting insulin (Additional file 1: Table S1). Based on *F*-statistic, all these SNPs were considered as strong instruments. Univariable MR analysis using IVW showed that genetically predicted fasting insulin was associated with a higher risk of CAD (OR 1.79, 95% CI 1.34 to 2.40, *P* value 8.16×10^{-5}). In the sensitivity analyses, the significant association was robust to the weighted median, MR-RAPS, and MR-PRESSO method (Table 1). rs998584 was associated with CAD directly in CARDIoGRAMplusC4D (Additional file 2: Fig. S1), and the result was unchanged after removing rs998584 (OR 1.72, 95% CI 1.30 to 2.28, *P* value 1.63×10^{-4}). rs1260326, rs10865959, rs13280813, rs7012814, rs7903146, and rs12454712 were related to alcohol drinking and BMI in the UK Biobank (Additional file 2: Fig. S2), and the significant association remained after removing these SNPs (OR 2.24, 95% CI 1.76 to 2.86, *P* value 1.02×10^{-10}).

Protein mediators on the associations of fasting insulin with CAD

After accounting for FDR, fasting insulin was associated with 355 circulating levels of proteins, including 97 proteins in the cardiometabolic panel, 112 proteins in the inflammation panel, 70 proteins in the neurology panel, and 76 proteins in the oncology panel (Fig. 1). Detailed results of insulin with 2940 proteins are shown in the Additional file 1: Table S2.

As shown in Fig. 2, among these 355 proteins affected by insulin, 17 proteins were shown to be associated with the risk of CAD after accounting for FDR, including ZHX2, CD164L2, PLA2G7, GZMA, LDLR, AGRP, GAST, KRT8, SCRIB, MICB, CCL4, PTPRC, CPXM2, CSF2, CD5, HHEX, and STAB2 (detailed results are presented in Additional file 1: Table S3). Genetic predicted fasting insulin was associated with a higher circulating level of these 17 proteins. Ten proteins (PLA2G7, GZMA, LDLR, AGRP, GAST, CCL4, PTPRC, CPXM2, CD5, and HHEX) also showed a positive association with the risk of CAD (Fig. 3). Steiger test supported the causal direction for these insulin-proteins and proteins-CAD associations (Additional file 1: Table S4). These ten proteins met the criteria (1) for mediating the positive effects of insulin on CAD and were included in the following sensitivity analyses and validation analyses.

Genetic instruments for these ten proteins are presented in Additional file 1: Table S5. Eight proteins (PLA2G7, GZMA, LDLR AGRP, CCL4, PTPRC, CPXM2, and CD5) had both *cis*-pQTLs and *trans*-pQTLs, while GAST and HHEX had *trans*-pQTLs only. In the sensitivity analyses, MR-Egger, weighted median, MR-RAPS, and MR-PRESSO gave directionally consistent estimates with results from two-step MR using IVW, except for the association of insulin with CPXM2 using MR-Egger method (Additional file 1: Table S6). Associations of genetic instruments for ten proteins with the risk of CAD are shown in Additional file 2: Figs. S3–S12. After removing SNPs associated with CAD directly, associations of insulin with these ten proteins, as well as PLA2G7, GZMA, LDLR, AGRP, GAST, CCL4, CPXM2, and HHEX with CAD, remained significant (Additional file 1: Table S7). Associations of genetic instruments for ten proteins with potential confounders, i.e., Townsend index, education, smoking status, alcohol drinking, physical activity, or BMI, are shown in Additional file 2: Figs. S13–S22. After removing SNPs associated with confounders, except for insulin with CPXM2, all associations remained significant (Additional file 1: Table S8). The number of proteins associated with each genetic instrument for these proteins are shown in Additional file 1: Table S5. When using IVs associated with no other proteins, the associations of all proteins except for CD5 and HHEX with the risk of CAD remained. The association of HHEX with CAD remained when using IVs associated with fewer than 7 proteins (Additional file 1: Table S9). When we restricted the analyses to *cis*-pQTLs, we found consistent directions of associations of PLA2G7, GZMA, AGRP, CCL4, CPXM2, and CD5 with CAD, despite with wider confidence intervals (Additional file 1: Table S10). For LDLR, only one *cis*-pQTL was identified (rs41307025) and it was unavailable in the CARDIoGRAMplusC4D. When using pQTLs at $P<1.7\times10^{-11}$ as IVs, the causal effects of PLA2G7, GZMA, LDLR, AGRP, GAST, and HHEX on the risk of CAD remained significant, while those of CCL4, PTPRC, CPXM2, and CD5 did not (Additional file 1: Table S10).

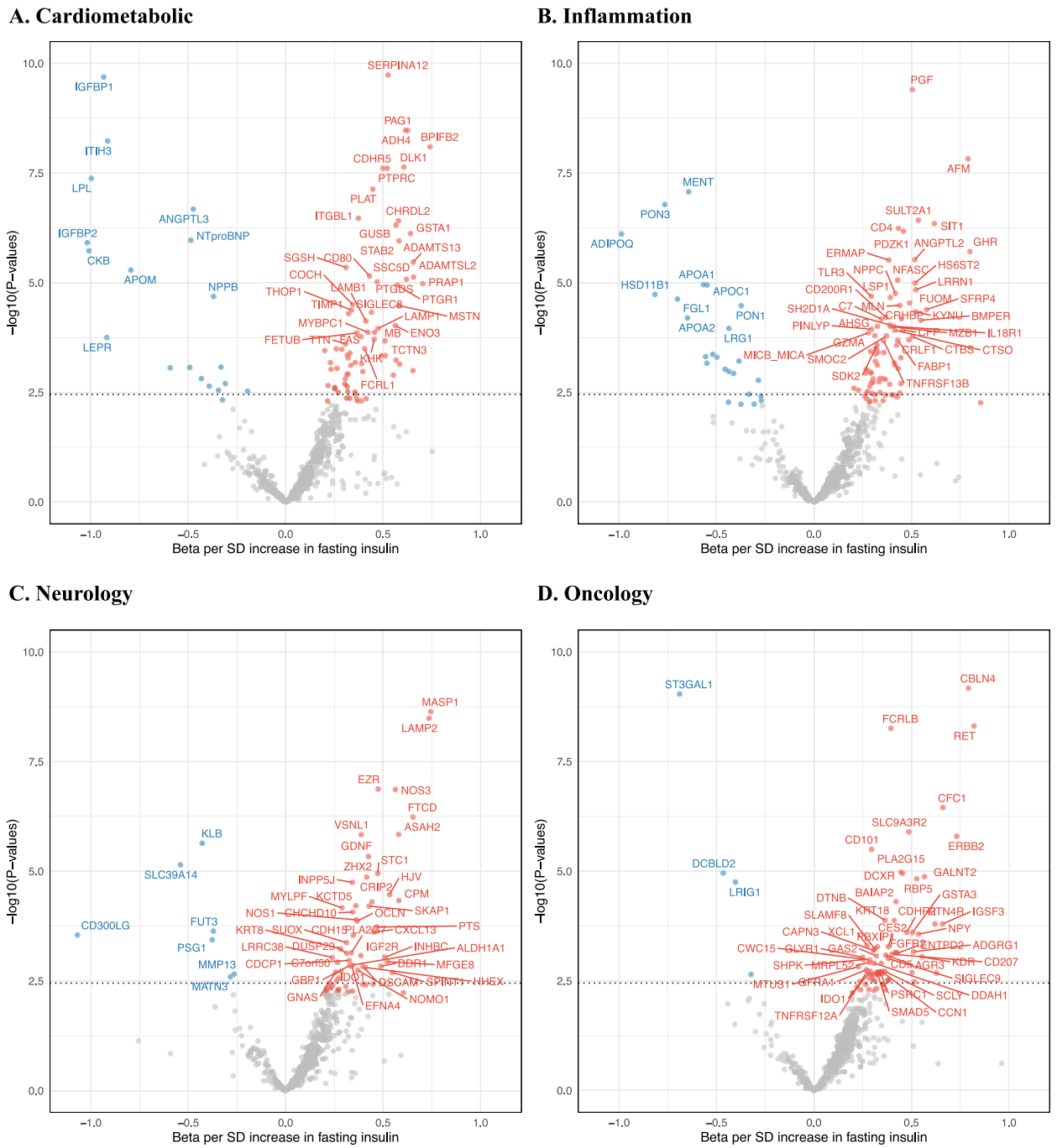


Fig. 1 Volcano plot on the associations of genetically predicted fasting insulin with 2940 proteins using the inverse-variance weighted method. **A** Volcano plot of insulin with proteins in cardiometabolic panel. **B** Volcano plot of insulin with proteins in inflammation panel. **C** Volcano plot of insulin with proteins in neurology panel. **D** Volcano plot of insulin with proteins in oncology panel. The x-axis represents beta estimates, and the y-axis represents − log10(*P*) values from MR analyses. Red dots represent proteins associated with fasting insulin with FDR < 5% and beta > 0; blue dots represent proteins associated with fasting insulin with FDR < 5% and beta < 0. Gray dots represent proteins that failed to pass FDR = 5%. The most significant 50 proteins are labeled with their names

We included PLA2G7, GZMA, LDLR, AGRP, GAST, and HHEX, which passed the sensitivity analyses in the validation analyses. Results of validation analyses are shown in Table 2. When using 12 genetic instruments predicting insulin unadjusted for BMI (Additional file 1: Table S11), the associations of insulin with CAD, GZMA,

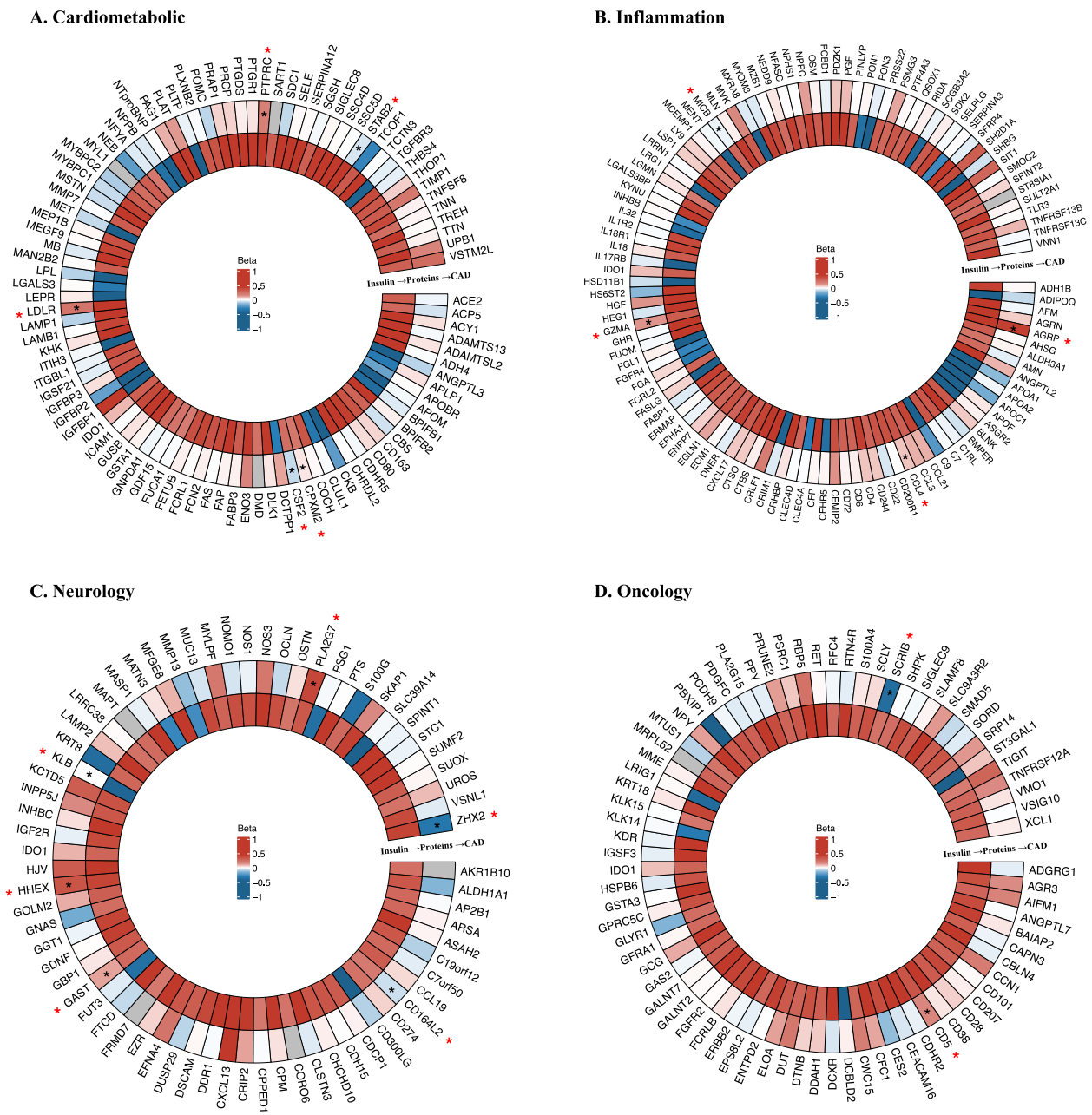


Fig. 2 Heatmap for the effects of insulin-affected proteins on the risk of CAD using the inverse-variance weighted method. **A** Heatmap for proteins in the cardiometabolic panel. **B** Heatmap for proteins in the inflammation panel. **C** Heatmap for proteins in the neurology panel. **D** Heatmap for proteins in the oncology panel. The inner circle presented the associations of fasting insulin with 355 proteins selected from step 1 MR. The outer circle presented the associations of these insulin-affected proteins with the risk of CAD derived from step 2 MR. Significance is marked with * for FDR < 5%

LDLR, and HHEX remained significant. The deCODE genetics consortium provided GWAS summary statistics only for PLA2G7, GZMA, LDLR, and AGRP. Replication analysis using data from deCODE showed directionally consistent estimates with those using data from UKB-PPP, and the associations of PLA2G7 with CAD, insulin with GZMA, and GZMA with CAD remained significant. In replication analysis using GWAS data of CAD

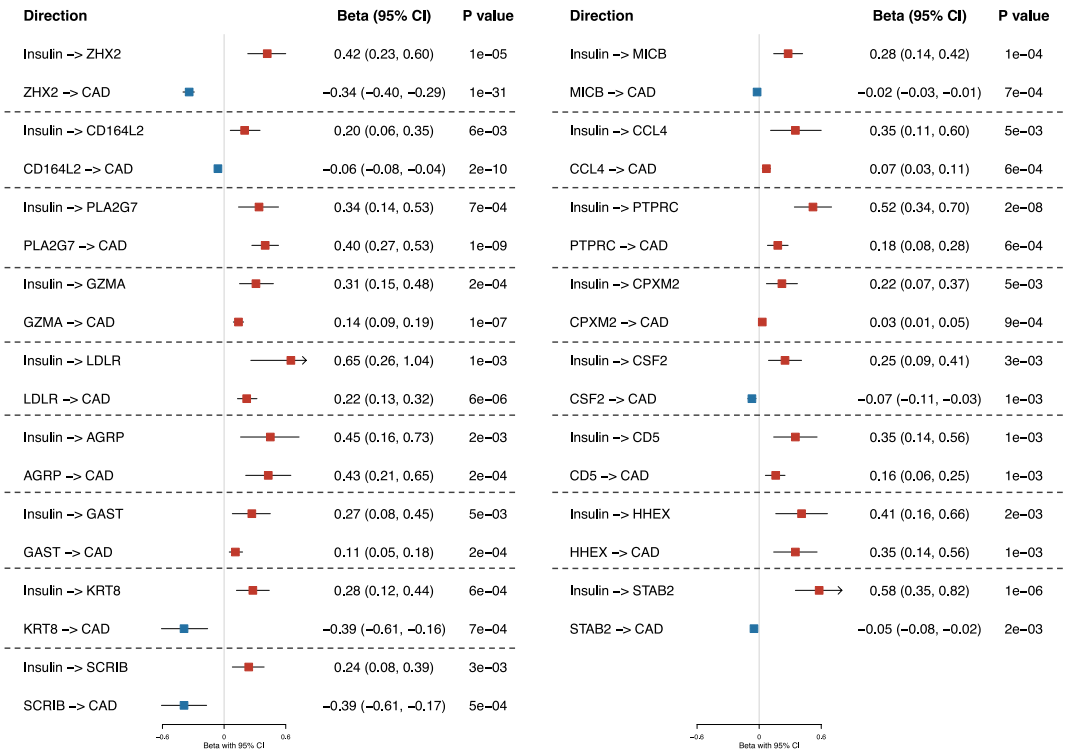


Fig. 3 Insulin-proteins-CAD associations selected from two-step MR

from FinnGen Biobank, except for GAST, all associations were directionally consistent with results using data from CARDIoGRAMplusC4D, and the associations of insulin with CAD, and PLA2G7, GZMA, LDLR. and AGRP with CAD remained significant. We therefore excluded GAST from mediators.

Mediation analyses

PLA2G7, GZMA, LDLR, AGRP, and HHEX were robust in sensitivity analyses and validation analyses and were identified as reliable mediators. However, when using non-pleiotropy pQTLs (i.e., pQTLs associated with no other proteins) as IVs, the association of HHEX with CAD was not statistically significant; therefore, we removed HHEX from mediation analyses. Results of multivariable MR are shown in Additional file 1: Table S12. Mediation analyses showed via protein mediators, insulin had a positive indirect effect on the risk of CAD. The mediation proportions of PLA2G7, GZMA, LDLR, and AGRP are 19.50%, 6.91%, 19.31%, and 29.66%, respectively (Table 3).

Discussion

Our findings, together with previous observational and MR studies [5, 11–13, 30], support a causal positive association of fasting insulin with the risk of CAD. Using proteomics in the MR analysis, we found 355 proteins affected by fasting insulin and identified PLA2G7, GZMA, LDLR, AGRP, and HHEX as potential mediators linking insulin to CAD. Genetically predicted fasting insulin increased the levels of proteins PLA2G7, GZMA, LDLR, AGRP, and HHEX and these proteins were associated with a higher risk of CAD, with replication of these findings in different data sources. Mediation analysis supported the mediating role of PLA2G7, GZMA, LDLR, and AGRP, adding evidence to the biological mechanisms behind the effects of insulin on CAD.

Existing evidence showed that insulin accelerates atherosclerosis development possibly by inhibiting lipolysis and increasing the synthesis of cholesterol, phospholipids, and triglyceride, which results in the accumulation of lipids and attraction of inflammatory cells in the vascular wall [30]. In this study, we identified PLA2G7 as an actionable mediator of the effect of insulin on the risk of CAD. PLA2G7 is a phospholipase involved in phospholipid catabolism during inflammatory and oxidative stress responses. In vitro experiment has shown that insulin acts as a positive regulator of PLA2G7 activity [31]. Our findings provide genetic evidence supporting a link between insulin and PLA2G7 levels in humans, though the causality requires further experimental validation. Observational studies and MR studies using

Table 2 Validation analysis for significant insulin-protein-CAD associations

	Beta	95% CI	*P* value
Genetic instruments for insulin unadjusted for BMI from MAGIC[a]			
Insulin → CAD	0.6074	(0.08, 1.13)	2.40E − 02
Insulin → PLA2G7	0.2148	(−0.26, 0.69)	3.80E − 01
Insulin → GZMA	0.4018	(0.08, 0.72)	1.38E − 02
Insulin → LDLR	0.7640	(0.23, 1.30)	5.25E − 03
Insulin → AGRP	0.2822	(−0.15, 0.72)	2.02E − 01
Insulin → GAST	0.1846	(−0.17, 0.54)	3.10E − 01
Insulin → HHEX	0.5575	(0.24, 0.88)	6.10E − 04
Genetic associations with proteins from deCODE[b]			
Insulin → PLA2G7	0.1600	(−0.07, 0.39)	1.64E − 01
PLA2G7 → CAD	0.0928	(0.03, 0.15)	3.18E − 03
Insulin → GZMA	0.2102	(0.01, 0.41)	3.62E − 02
GZMA → CAD	0.0631	(0.02, 0.10)	2.92E − 03
Insulin → LDLR	0.0921	(−0.11, 0.30)	3.79E − 01
LDLR → CAD	0.0406	(−0.19, 0.28)	7.34E − 01
Insulin → AGRP	0.1909	(−0.05, 0.43)	1.15E − 01
AGRP → CAD	0.0033	(−0.03, 0.04)	8.45E − 01
Genetic associations with CAD from FinnGen Biobank[c]			
Insulin → CAD	0.6754	(0.31, 1.04)	2.66E − 04
PLA2G7 → CAD	0.4107	(0.25, 0.58)	1.14E − 06
GZMA → CAD	0.0988	(0.01, 0.18)	2.47E − 02
LDLR → CAD	0.2185	(0.12, 0.32)	3.42E − 05
AGRP → CAD	0.2667	(0.06, 0.47)	1.19E − 02
GAST → CAD	−0.1784	(−0.30, −0.05)	5.36E − 03
HHEX → CAD	0.2962	(−0.02, 0.61)	6.73E − 02

[a] Genetic association with proteins and CAD were obtained from UKB-PPP and CARDIoGRAMplusC4D, respectively

[b] Genetic association with insulin and CAD were obtained from MAGIC and CARDIoGRAMplusC4D, respectively

[c] Genetic association with insulin and proteins were obtained from MAGIC and UKB-PPP, respectively

Table 3 Mediation analyses of protein mediators in the association of fasting insulin with the risk of CAD

Protein	Indirect effect of insulin on CAD via protein (beta (95% CI))[a]	Mediation proportion (% (95% CI))[b]
PLA2G7	0.11 (0.04, 0.19)	19.50% (6.79%, 32.20%)
GZMA	0.04 (0.01, 0.07)	6.91% (2.19%, 11.63%)
LDLR	0.11 (0.01, 0.21)	19.31% (2.45%, 36.17%)
AGRP	0.17 (0.03, 0.31)	29.66% (5.68%, 53.62%)

[a] Beta and its 95% CI of indirect effects were obtained using the product of coefficients method and the delta method

[b] The mediation proportions were calculated as indirect effect divided by the total effect

different genetic instruments showed PLA2G7 were positively associated with proatherogenic lipids and the risk of coronary heart disease, expected to be a potential therapeutic target for coronary atherosclerosis, which is consistent with our results [32, 33]. However, evidence from RCTs with consistent results is lacking; exploring its efficacy especially in hyperinsulinemia populations or in patients with type 2 diabetes may be warranted.

We also found that insulin may elevate the risk of CAD by increasing circulating levels of LDLR. LDLR is a cell surface glycoprotein with the ability of binding and up taking of plasma low-density lipoprotein (LDL) particles [34]. Consistent with our findings, insulin was found to be a stimulant of LDLR, associated with increased expression of LDLR [35]. Circulating LDLR is thought to be a product of LDLR shedding from the cell surface. Existing experimental and epidemiological studies showed that elevated circulating levels of LDLR were associated with poor lipid profiles and a higher risk of atherosclerosis and CAD, suggesting that alteration of LDLR shedding can emerge as a target to treat dyslipidemia [36–38]. Insulin may facilitate this process and increase LDLR levels in circulation, thereby promoting CAD.

In addition to proteins relevant to lipid metabolism, we also found three novel proteins GZMA, AGRP, and HHEX affected by fasting insulin and associated with an increased risk of CAD. GZMA is the most abundant protease in the cytotoxic granules of killer cells and possesses pro-inflammatory activity [39]. GZMA is implicated in several inflammatory diseases, and early studies have reported its elevation in murine atherosclerotic plaques, potentially due to the increased inflammation and immune system overactivity it induces [40, 41]. While direct evidence on how insulin influences GZMA is lacking, previous studies have shown significant correlations between insulin and the secretion or activity of cytotoxic T cells and NK cells, the primary producers of GZMA, suggesting that insulin may indirectly affect GZMA levels [42, 43]. Interestingly, AGRP is a neuropeptide produced in the hypothalamic arcuate nucleus and is well-known for its potent orexigenic effects [44]. Activation of AGRP neurons leads to decreased energy expenditure, weight gain, and increased blood pressure, indicating its potential role in cardiovascular regulation [45]. AGRP neuron function is regulated by various hormones, including insulin, and in turn, efferent signals from AGRP neurons modulate insulin action on hepatic glucose production [46, 47]. The identification of this mediator may improve the understanding of the neural actions of insulin. We also identified HHEX linking insulin to CAD. *HHEX* is reported to be a diabetes gene whose expression is required for δ-cell maintenance and islet function [48]. And mutations in *HHEX* gene were

found to be associated with abnormalities of cardiac and vascular development [49]. However, this finding should be interpreted with caution, as the genetic instruments for HHEX are only from *trans*-pQTLs, some of which are highly pleiotropic, and genetic instruments for HHEX were not available in other datasets for replication. Drug targets developed against these proteins may offer new therapeutic strategies for CAD, particularly in populations at high risk for diabetes. However, to our knowledge, no relevant clinical trials have been conducted to explore this potential.

This is the first study using MR analyses and proteomics to explore the mechanisms via proteins of insulin's effect on CAD. By identifying proteins involved in lipid metabolism, such as PLA2G7 and LDLR, along with three novel proteins—GZMA, AGRP, and HHEX—as mediators, this study provides novel insights for the development of targeted therapies against the cardiovascular disease risk associated with high insulin. Despite the novelty, we acknowledge several limitations. First, MR estimates may be less precise since the genetic instruments capture only a small proportion of variance. To address this, we used GWAS summary statistics from large consortia or meta-analyses and performed replication analyses when available. Second, MR relies on three key assumptions, i.e., the relevance, independence, and exclusion-restriction assumption [7], which may be subject to violation. To satisfy the relevance assumptions, we only used strong instruments for fasting insulin or proteins. There is a concern that *trans*-pQTLs are more likely to have pleiotropic effects. However, the directions of associations are consistent when using *cis*-pQTLs, despite with wider confidence intervals. This is possibly due to the less genetic instruments and therefore lower power. Moreover, we obtained consistent associations when using a range of sensitivity analyses, including using different analytic methods robust to pleiotropy, excluding pleiotropic SNPs associated with outcome (CAD) or potential confounders, limiting instruments to those associated with none or few other proteins, applying a more stringent significance threshold ($P < 1.7 \times 10^{-11}$) to select instruments, and replication analyses using different independent data sources with different confounding structures. Third, sample overlap may bias two-sample MR results. However, a simulation study suggested that MR results from overlapping samples in large cohorts like the UK Biobank are valid [50]. And we also used data from the Icelanders and FinnGen Biobank which does not overlap to validate our results. Fourth, although the protein mediators in this study were strictly selected, it is possible that some other protein mediators were not selected due to the lack of statistical significance [51]. And we cannot rule out the possibility that other unmeasured proteins are involved in the mechanism of insulin effects. Further studies in larger sample sizes and larger proteomics datasets would be needed, but we used by far the largest dataset available. To increase the validity, we also conducted several replication analyses and only included the proteins that remained after replication as mediators. Fifth, we only used data from people of European ancestry considering population stratification might confound MR estimates, so the findings cannot be generalizable to other populations. Future studies assessing the effects in non-European populations, like East Asians, are needed, but we cannot find large GWAS of proteins with comparable sample sizes in non-European populations by far. Lastly, the mediation proportion, such as 6.91% for GZMA, is modest. However, it offers valuable insights into the biological mechanisms linking fasting insulin to CAD and should not be underestimated. In addition, the roles of epigenetic and environmental factors in mediating insulin's effects also warrant investigation in future studies.

Conclusions

In conclusion, we found 355 proteins affected by fasting insulin and identified PLA2G7, GZMA, LDLR, AGRP, and HHEX as key mediators on the pathway from insulin to CAD. These findings offer new insights into the biological mechanisms linking insulin to CAD and suggest potential therapeutic targets for mitigating insulin-related cardiovascular risk by modulating the specific proteins identified. Further research is needed to validate these findings, and caution is advised in their clinical application.

Abbreviations

CAD	Coronary artery disease
FDR	False discovery rate
GWAS	Genome-wide association study
IVs	Instrumental variables
IVW	Inverse variance weighting
MAGIC	Meta-Analysis of Glucose and Insulin-related Traits Consortium
MR	Mendelian randomization
MR-PRESSO	MR Pleiotropy RESidual Sum and Outlier
MR-RAPS	MR Robust Adjusted Profile Score
PQTL	Protein quantitative trait locus
RCTs	Randomized controlled trials
SE	Standard error
UKB-PPP	UK Biobank Pharma Proteomics Project

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-04127-6.

Additional file 1: Tables S1–S12. Table S1 Genome-wide significant ($P < 5 \times 10^{-8}$) and uncorrelated ($r^2 < 0.01$) genetic instruments for fasting insulin. Table S2 Associations of genetically predicted fasting insulin with circulating levels of 2940 proteins using inverse-variance weighted method. Table S3 Associations of genetically predicted circulating levels of insulin-driven proteins with the risk of CAD using inverse-variance weighted method. Table S4 MR Steiger directionality test for significant

insulin-proteins-CAD associations obtained from two-step MR. Table S5 Genome-wide significant ($P < 5 \times 10^{-8}$) and uncorrelated ($r^2 < 0.01$) genetic instruments for ten proteins. Table S6 Sensitivity analyses for significant insulin-proteins-CAD associations using different analytic methods. Table S7 Sensitivity analyses for significant insulin-proteins-CAD associations after removing SNPs associated with CAD directly. Table S8 Sensitivity analyses for significant insulin-proteins-CAD associations after removing SNPs associated with potential confounders. Table S9 Sensitivity analyses for proteins-CAD associations with the full and non-pleiotropic instruments. Table S10 Sensitivity analyses for selected proteins-CAD associations using cis-pQTL only or using pQTLs at $P < 1.7E{-}11$ as instrumental variables. Table S11 Genetic instruments for fasting insulin unadjusted for BMI. Table S12 Direct effects of protein mediators on the risk of CAD derived from multivariable MR.

Additional file 2: Figures S1–S22. Fig. S1 Associations of genetic instruments for fasting insulin with the risk of CAD. Fig. S2 Associations of genetic instruments for fasting insulin with potential confounders in the UK Biobank. Fig. S3 Associations of genetic instruments for PLA2G7 with the risk of CAD. Fig. S4 Associations of genetic instruments for GZMA with the risk of CAD. Fig. S5 Associations of genetic instruments for LDLR with the risk of CAD. Fig. S6 Associations of genetic instruments for AGRP with the risk of CAD. Fig. S7 Associations of genetic instruments for GAST with the risk of CAD. Fig. S8 Associations of genetic instruments for CCL4 with the risk of CAD. Fig. S9 Associations of genetic instruments for PTPRC with the risk of CAD. Fig. S10 Associations of genetic instruments for CPXM2 with the risk of CAD. Fig. S11 Associations of genetic instruments for CD5 with the risk of CAD. Fig. S12 Associations of genetic instruments for HHEX with the risk of CAD. Fig. S13 Associations of genetic instruments for PLA2G7 with potential confounders in the UK Biobank. Fig. S14 Associations of genetic instruments for GZMA with potential confounders in the UK Biobank. Fig. S15 Associations of genetic instruments for LDLR with potential confounders in the UK Biobank. Fig. S16 Associations of genetic instruments for AGRP with potential confounders in the UK Biobank. Fig. S17 Associations of genetic instruments for GAST with potential confounders in the UK Biobank. Fig. S18 Associations of genetic instruments for CCL4 with potential confounders in the UK Biobank. Fig. S19 Associations of genetic instruments for PTPRC with potential confounders in the UK Biobank. Fig. S20 Associations of genetic instruments for CPXM2 with potential confounders in the UK Biobank. Fig. S21 Associations of genetic instruments for CD5 with potential confounders in the UK Biobank. Fig. S22 Associations of genetic instruments for HHEX with potential confounders in the UK Biobank.

Additional file 3. STROBE-MR checklist.

Acknowledgements
The authors would like to thank the MAGIC, CARDIoGRAMplusC4D, UK Biobank, deCODE genetics consortium, and FinnGen Biobank for sharing the valuable data.

Authors' contributions
The authors' responsibilities were as follows— JVZ conceived the idea; XH and JVZ designed the study; XH and JVZ analyzed the data and interpreted the findings; XH wrote the first draft; XH did the data visualization; JVZ critically revised the manuscript for important intellectual content. Both authors read and approved the final manuscript.

Authors' Twitter handles
Twitter handles: @jvzhao410 (Jie V. Zhao), @hkusph, @hkumed.

Funding
The authors reported no funding received for this study.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
This study used only published summary data that does not require ethical approval.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]School of Public Health, Li Ka Shing, Faculty of Medicine, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China. [2]State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong, Hong Kong SAR, China.

Received: 21 February 2025 Accepted: 13 May 2025
Published online: 30 May 2025

References

1. Reaven GM. Role of insulin resistance in human disease. Diabetes. 1988;37(12):1595–607.
2. Welborn TA, Wearne K. Coronary heart disease incidence and cardiovascular mortality in busselton with reference to glucose and insulin concentrations. Diabetes Care. 1979;2(2):154–60.
3. Pyörälä K, Savolainen E, Kaukola S, Haapakoski J. Plasma insulin as coronary heart disease risk factor: relationship to other risk factors and predictive value during 9 1/2-year follow-up of the Helsinki Policemen Study population. Acta Med Scand Suppl. 1985;701:38–52.
4. Ducimetiere P, Eschwege E, Papoz L, Richard JL, Claude JR, Rosselin G. Relationship of plasma insulin levels to the incidence of myocardial infarction and coronary heart disease mortality in a middle-aged population. Diabetologia. 1980;19(3):205–10.
5. Kolb H, Kempf K, Röhling M, Martin S. Insulin: too much of a good thing is bad. BMC Med. 2020;18(1):224.
6. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet. 2020;52(10):1122–31.
7. Davies NM, Holmes MV, Davey SG. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362: k601.
8. Feskens EJ, Kromhout D. Hyperinsulinemia, risk factors, and coronary heart disease. the Zutphen Elderly Study. Arteriosclerosis and Thrombosis: A Journal of Vascular Biology. 1994;14(10):1641–7.
9. de Boer IH, Katz R, Chonchol MB, Fried LF, Ix JH, Kestenbaum B, et al. Insulin resistance, cystatin C, and mortality among older adults. Diabetes Care. 2012;35(6):1355–60.
10. Lewis JG, Shand BI, Frampton CM, Elder PA, Scott RS. Plasma retinol-binding protein is not a marker of insulin resistance in overweight subjects: a three year longitudinal study. Clin Biochem. 2008;41(13):1034–8.
11. Zhao JV, Luo S, Schooling CM. Sex-specific Mendelian randomization study of genetically predicted insulin and cardiovascular events in the UK Biobank. Communications Biology. 2019;2(1):332.
12. Wang Z, Zhu M, Huang Y, Cao J, Xiong Z. High blood pressure mediated the effect of fasting insulin level on nonalcoholic fatty liver disease risk: a Mendelian randomization study. Digit Health. 2023;9:20552076231216680.
13. Tikkanen E, Pirinen M, Sarin A-P, Havulinna AS, Männistö S, Saltevo J, et al. Genetic support for the causal role of insulin in coronary heart disease. Diabetologia. 2016;59(11):2369–77.
14. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, et al. The trans-ancestral genomic architecture of glycemic traits. Nat Genet. 2021;53(6):840–60.

15. D'Orazio P, Burnett RW, Fogh-Andersen N, Jacobs E, Kuwa K, Külpmann WR, et al. Approved IFCC recommendation on reporting results for blood glucose (abbreviated). Clin Chem. 2005;51(9):1573–6.
16. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. Circ Res. 2018;122(3):433–43.
17. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int J Epidemiol. 2016;45(6):1961–74.
18. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol. 2015;30(7):543–52.
19. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40(4):304–14.
20. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(5):693–8.
21. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. The Annals of Statistics. 2020;48(3):1742–69, 28.
22. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32(5):377–89.
23. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023;622(7982):329–38.
24. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3): e1001779.
25. Hemani G, Tilling K, Davey SG. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 2017;13(11): e1007081.
26. Karjalainen MK, Karthikeyan S, Oliver-Williams C, Sliz E, Allara E, Fung WT, et al. Genome-wide characterization of circulating metabolic biomarkers. Nature. 2024;628(8006):130–8.
27. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan Ja, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature Genetics. 2012;44(9):991–1005.
28. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet. 2021;53(12):1712–21.
29. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. Sociol Methodol. 1982;13:290–312.
30. Jandeleit-Dahm KAM, Gray SP. Insulin and cardiovascular disease: biomarker or association? Diabetologia. 2012;55(12):3145–51.
31. Schliefsteiner C, Hirschmugl B, Kopp S, Curcic S, Bernhart EM, Marsche G, et al. Maternal gestational diabetes mellitus increases placental and foetal lipoprotein-associated phospholipase A2 which might exert protective functions against oxidative stress. Sci Rep. 2017;7(1):12628.
32. Sun L, Zhu Z, Shi M, Jia Y, Yang P, Wang Y, et al. Causal effect of lipoprotein-associated phospholipase A2 activity on coronary artery disease and myocardial Infarction: a two-sample Mendelian randomization study. Clin Chim Acta. 2021;523:491–6.
33. Thompson A, Gao P, Orfei L, Watson S, Di Angelantonio E, Kaptoge S, et al. Lipoprotein-associated phospholipase A(2) and risk of coronary disease, stroke, and mortality: collaborative analysis of 32 prospective studies. Lancet. 2010;375(9725):1536–44.
34. Mineo C. Lipoprotein receptor signalling in atherosclerosis. Cardiovasc Res. 2020;116(7):1254–74.
35. Chandra NC. A comprehensive account of insulin and LDL receptor activity over the years: a highlight on their signaling and functional role. J Biochem Mol Toxicol. 2021;35(9): e22840.
36. Mayne J, Ooi TC, Tepliakova L, Seebun D, Walker K, Mohottalage D, et al. Associations between soluble LDLR and lipoproteins in a white cohort and the effect of PCSK9 loss-of-function. J Clin Endocrinol Metab. 2018;103(9):3486–95.
37. Shimohiro H, Taniguchi S, Koda M, Sakai C, Yamada S. Association between serum soluble low-density lipoprotein receptor levels and metabolic factors in healthy Japanese individuals. J Clin Lab Anal. 2015;29(1):52–6.
38. Peng J, Xing CY, Zhao K, Deng J, Olmedo DA, Ma Z, et al. Associations of pro-protein convertase subtilisin-like kexin type 9, soluble low-density lipoprotein receptor and coronary artery disease: a case-control study. Int J Cardiol. 2022;350:9–15.
39. Lieberman J. Granzyme A activates another way to die. Immunol Rev. 2010;235(1):93–104.
40. Zeglinski MR, Granville DJ. Granzymes in cardiovascular injury and disease. Cell Signal. 2020;76: 109804.
41. Hiebert PR, Boivin WA, Zhao H, McManus BM, Granville DJ. Perforin and granzyme B have separate and distinct roles during atherosclerotic plaque development in apolipoprotein E knockout mice. PLoS ONE. 2013;8(10): e78939.
42. Kim JH, Park K, Lee SB, Kang S, Park JS, Ahn CW, et al. Relationship between natural killer cell activity and glucose control in patients with type 2 diabetes and prediabetes. J Diabetes Investig. 2019;10(5):1223–8.
43. Amer J, Salhab A, Noureddin M, Doron S, Abu-Tair L, Ghantous R, et al. Insulin signaling as a potential natural killer cell checkpoint in fatty liver disease. Hepatol Commun. 2018;2(3):285–98.
44. Tallam LS, Kuo JJ, da Silva AA, Hall JE. Cardiovascular, renal, and metabolic responses to chronic central administration of agouti-related peptide. Hypertension (Dallas, Tex : 1979). 2004;44(6):853–8.
45. Jiang J, Morgan DA, Cui H, Rahmouni K. Activation of hypothalamic AgRP and POMC neurons evokes disparate sympathetic and cardiovascular responses. Am J Physiol Heart Circ Physiol. 2020;319(5):H1069–77.
46. Farias Quipildor G, Mao K, Beltran PJ, Barzilai N, Huffman DM. Modulation of glucose production by central insulin requires IGF-1 receptors in AgRP neurons. Diabetes. 2021;70(10):2237–49.
47. Könner AC, Janoschek R, Plum L, Jordan SD, Rother E, Ma X, et al. Insulin action in AgRP-expressing neurons is required for suppression of hepatic glucose production. Cell Metab. 2007;5(6):438–49.
48. Zhang J, McKenna LB, Bogue CW, Kaestner KH. The diabetes gene Hhex maintains δ-cell differentiation and islet function. Genes Dev. 2014;28(8):829–34.
49. Hallaq H, Pinter E, Enciso J, McGrath J, Zeiss C, Brueckner M, et al. A null mutation of Hhex results in abnormal cardiac development, defective vasculogenesis and elevated Vegfa levels. Development. 2004;131(20):5197–209.
50. Minelli C, Del Greco MF, van der Plaat DA, Bowden J, Sheehan NA, Thompson J. The use of two-sample methods for Mendelian randomization analyses on single large datasets. Int J Epidemiol. 2021;50(5):1651–9.
51. Zhao JV, Zhang J. Using proteomics to identify the mechanisms underlying the benefits of statins on ischemic heart disease. npj Cardiovascular Health. 2024;1(1):15.

Publisher's Note