

RESEARCH

Open Access



# Improved sex-specific cardiovascular risk prediction with multi-omics data in people with type 2 diabetes

Ruijie Xie<sup>1,2</sup>, Christian Herder<sup>3,4,5</sup>, Sha Sha<sup>1</sup>, Hermann Brenner<sup>1,6</sup>, Sigrid Carlsson<sup>1,7</sup> and Ben Schöttker<sup>1\*</sup>

## Abstract

**Background** To evaluate whether integrating proteomics, metabolomics, and a cardiovascular disease specific polygenic risk score (CVD-PRS) in the SCORE2-Diabetes model improves sex-specific 10-year prediction of major adverse cardiovascular events (MACE) in individuals with type 2 diabetes (T2D).

**Methods** Genome-wide association study (GWAS), plasma proteomics (with the Olink Explore 3072 platform), and metabolomics (with nuclear magnetic resonance spectroscopy by Nightingale Health) data were measured in the UK Biobank. A novel sex-specific protein algorithm was developed using bootstrap-LASSO (Least absolute shrinkage and selection operator) regression. The CVD-PRS and sex-specific metabolite algorithms were used from previous UK Biobank projects. In a subset of 990 participants with T2D, age 40–69 years, with no prior MACE, and complete multi-omics data, we evaluated, which omics data improved the SCORE2-Diabetes model performance using Harrell's C-index.

**Results** Overall 9 proteins were selected for males and 7 for females and adding them to the SCORE2-Diabetes model significantly improved discrimination in the total population (C-index increase from 0.766 to 0.835 ( $P < 0.001$ )). Further adding of metabolites significantly improved model performance (C-index, 0.846,  $P = 0.035$ ), which was mostly attributable to model improvement among males ( $\Delta$ C-index, 0.012,  $P = 0.078$ ) but not among females ( $\Delta$ C-index, 0.004,  $P = 0.723$ ). Further adding the CVD-PRS did not statistically significantly improve the SCORE2-Diabetes + proteomics + metabolomics model further in the total population (C-index, 0.848 ( $P = 0.070$ )).

**Conclusions** Sex-specific proteomic signatures markedly improved 10-year MACE risk prediction in individuals with T2D. In men but not in women, further integration of metabolomics may enhance model performance whereas adding the CVD-PRS is not needed. External validation is warranted.

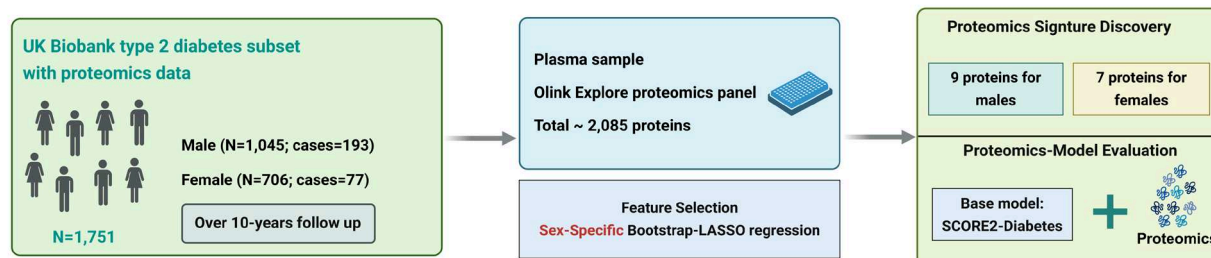
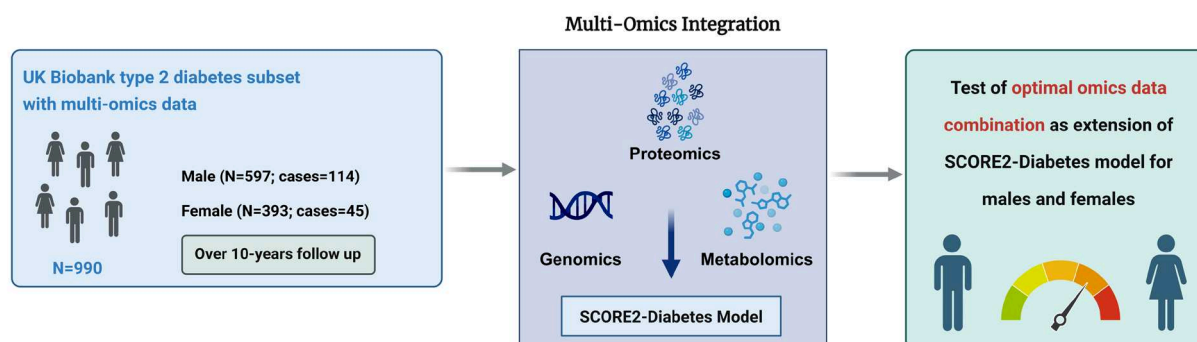
**Keywords** Proteomics, Multi-omics, Cardiovascular risk, Type 2 diabetes, Sex-specific, SCORE2-diabetes

\*Correspondence:  
Ben Schöttker  
b.schoettker@dkfz.de

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Graphical abstract****STAGE A: Proteomics Signature Discovery & Evaluation****STAGE B: Incremental Predictive Value of Multi-Omics Integration****Introduction**

Cardiovascular disease (CVD) remains the leading global cause of morbidity and mortality, and individuals with type 2 diabetes (T2D) are at substantially higher risk than the general population [1]. Despite advances in cardiovascular care, individuals with T2D continue to experience accelerated atherosclerosis and an elevated risk of major adverse cardiovascular events (MACE), underscoring the need for more precise and individualized risk stratification strategies [2, 3]. Additionally, sex-specific differences in cardiovascular outcomes have been increasingly recognized in individuals with T2D [4]. While men generally show higher absolute CVD risk at younger ages, women with T2D face a disproportionately greater relative increase in cardiovascular risk [5, 6].

In response to the need for more personalized prediction tools, the European Society of Cardiology recently introduced SCORE2-Diabetes, which is a sex-specific algorithm developed from large European cohort data to estimate the 10-year risk of MACE [7]. While this represents a significant improvement over previous general-population models, SCORE2-Diabetes primarily incorporates traditional clinical risk factors and does not sufficiently account for subclinical pathophysiological mechanisms or sex-specific variability, potentially

limiting its predictive accuracy and clinical applicability [8, 9].

Omics-based methodologies, encompassing proteomics, metabolomics, and genomics, provide opportunities to improve cardiovascular risk prediction by quantifying disease-associated molecular signatures and uncovering underlying biological pathways not captured by traditional clinical assessments alone [10]. In particular, proteomics enables the identification of circulating proteins that reflect early pathophysiological changes and offers insights into the complex interplay of genetic, environmental, and metabolic factors contributing to CVD, especially in individuals with T2D [11, 12]. However, the application of proteomic biomarkers in cohorts of patients with T2D remains limited, and most previous studies have treated sex as a covariate rather than examining sex-specific biological differences directly [13, 14]. Given accumulating evidence of pronounced sex differences in proteomic signatures and cardiovascular pathogenesis, there is a critical need to develop and validate sex-specific proteomic risk models tailored to patients with T2D [15, 16].

The recent availability of large-scale proteomic data in the UK Biobank (UKB) provides an unprecedented opportunity to identify circulating protein biomarkers

that improve cardiovascular risk prediction in individuals with T2D [17]. In this study, we first aimed to develop sex-specific proteomic signatures that enhance the predictive performance of the SCORE2-Diabetes model for 10-year MACE risk. We then evaluated whether the addition of metabolomic biomarkers and a cardiovascular polygenic risk score (PRS) could further improve prediction. The metabolomic biomarkers were derived from our previous study using the UKB, which demonstrated that 7 sex-specific metabolites significantly enhanced the predictive accuracy of the SCORE2-Diabetes model [18]. The CVD-PRS was provided by the UKB based on external GWAS and internal optimization [19]. Building on these multi-omics components, we investigated whether the integration of multiple omics layers could further refine sex-specific cardiovascular risk models in people with T2D.

## Methods

### Study population

The UKB is a large, population-based prospective cohort comprising 502,414 participants aged 40–69 years at recruitment (with few outliers), conducted between 13 March 2006 and 1 October 2010 across 22 assessment centers in England, Scotland, and Wales [20]. The cohort is predominantly White (approximately 94–95% White overall, and about 80% White British). The cohort includes comprehensive phenotypic and genotypic data, covering biomarkers, imaging, lifestyle, anthropometric, and genomic information.

The present study included two analytical cohorts (Supplemental Fig. S1). For the proteomics-based analysis, we excluded participants without proteomics data ( $n=448,195$ ), with more than 50% missing proteomics data ( $n=1871$ ), and individuals non-randomly selected for proteomics measurements who did not experience a MACE event during follow-up ( $n=5282$ ). Among the remaining 47,066 participants with usable proteomics data, we excluded individuals without diabetes ( $n=44,365$ ), with missing diabetes status ( $n=189$ ), diagnosed with type 1 diabetes ( $n=488$ ), with a history of MACE or missing outcome data ( $n=254$ ), or aged outside the SCORE2-Diabetes target range ( $< 40$  or  $> 69$  years;  $n=19$ ). Ultimately, 1751 participants with T2D were included in the proteomics-based analysis dataset. The study design is a case-cohort design with a randomly selected T2D sub-cohort with  $n=1672$  study participants (of whom  $n=191$  developed MACE during follow-up) and  $n=79$  additional MACE cases [21].

The multi-omics-based analysis used the same exclusion criteria but additionally excluded participants with missing metabolomics data ( $n=750$ ) or PRS data ( $n=11$ ), resulting in 990 participants with T2D and complete multi-omics.

### Base model: SCORE2-diabetes

The SCORE2-Diabetes is a developed and validated sex-specific cardiovascular risk model tailored for adults with T2D aged 40–69 years [7]. The model incorporates the following variables: age, sex, systolic blood pressure (SBP), current smoking status, age at diabetes diagnosis, high-density lipoprotein cholesterol (HDL-C), total cholesterol (TC), glycated hemoglobin (HbA<sub>1c</sub>), and estimated glomerular filtration rate (eGFR). Age, sex, and smoking status were self-reported through standardized touchscreen questionnaires at baseline. SBP was measured in the seated position using an automated digital sphygmomanometer (Omron HEM-7015IT) on the left arm, with the mean of two readings used for analysis. HDL-C and total cholesterol concentrations were determined from non-fasting blood samples using enzymatic colorimetric assays on the Beckman Coulter AU5800 platform. HbA<sub>1c</sub> was measured via high-performance liquid chromatography (Bio-Rad Variant II Turbo). eGFR was calculated using the CKD-EPI 2009 creatinine equation.

### Proteomics data

Proteomic profiling was conducted on EDTA-plasma samples collected at baseline. The assay protocols, including sample handling and selection procedures, have been described previously [22]. In brief, a total of 2923 unique proteins were measured using the Olink Explore 3072 platform (Olink Proteomics, Uppsala, Sweden). Olink utilizes the Proximity Extension Assay (PEA) method that targets proteins via pairs of antibodies linked to complementary oligonucleotides [23]. As outlined in our prior study [24], proteins with more than 20% missing values or with over 25% of values below the detection limit were excluded ( $n=838$ ). Ultimately, 2085 proteins from the Olink Explore platform were included in the proteomic biomarker selection process. In this study, sex-specific biomarker selection was performed using least absolute shrinkage and selection operator (LASSO) with bootstrap resampling to identify the most predictive proteins for cardiovascular risk among individuals with T2D. This bootstrap resampling approach serves as a form of internal validation for the feature selection process, aiming to ensure the stability of the selected protein signature and to mitigate the risk of overfitting.

### Metabolomic data

Plasma metabolite profiling was performed by Nightingale Health using a targeted nuclear magnetic resonance (NMR) spectroscopy platform [25]. This approach captured a broad spectrum of 249 circulating metabolites, including lipid subclasses, fatty acids, amino acids, and other low-molecular-weight compounds. Of these, 168 were quantified as absolute concentrations,

encompassing 61 composite biomarkers derived from 107 directly measured metabolites, while the remaining 81 were reported as ratio measures. In our previous study [18], we applied sex-specific LASSO regression to the 249 metabolites and identified a parsimonious signature comprising 7 metabolites (3 common to both sexes, 3 male-specific and 1 female-specific; Supplemental Table S1). Incorporation of these metabolites into the SCORE2-Diabetes model increased C-index from 0.660 to 0.678 in the total population ( $P=0.037$ ), with  $\Delta C\text{-index} = +0.025$  in men (from 0.632 to 0.657) and  $\Delta C\text{-index} = +0.009$  in women (from 0.670 to 0.679). In the present analysis, we evaluated only these 7 pre-selected metabolites.

### Polygenic risk score for CVD

The PRS for CVD was obtained from the UKB PRS Release [15]. This pre-computed score incorporates 1.42 million genetic variants with non-zero weights. The score was developed using a Bayesian modeling framework, initially trained on meta-analyzed summary statistics from nine external GWAS datasets. These summary statistics

**Table 1** Baseline characteristics of participants included in the proteomics and multi-omics analyses from the UK biobank

| Baseline characteristics                       | Proteomics cohort (N = 1751) | Multi-omics cohort (N = 990) | P-value |
|------------------------------------------------|------------------------------|------------------------------|---------|
| Male sex, N (%)                                | 1045 (59.7%)                 | 597 (60.3%)                  | 0.796   |
| Age (years)                                    | 60.3 (6.6)                   | 60.5 (6.6)                   | 0.364   |
| Current smoker, N (%)                          | 187 (10.7%)*                 | 103 (10.4%)*                 | 0.879   |
| Systolic blood pressure (mmHg)                 | 144.5 (17.7)*                | 144.3 (17.6)*                | 0.757   |
| Total cholesterol (mmol/L)                     | 4.5 (1.0)*                   | 4.5 (1.0)*                   | 0.790   |
| HDL cholesterol (mmol/L)                       | 1.2 (0.3)*                   | 1.2 (0.3)*                   | 0.671   |
| Age at diabetes diagnosis (years)              | 55.0 (6.5)*                  | 55.2 (6.4)*                  | 0.335   |
| HbA <sub>1c</sub> (mmol/mol)                   | 51.8 (12.9)*                 | 51.7 (12.4)*                 | 0.833   |
| eGFR (ml/min/1.73m <sup>2</sup> ) <sup>a</sup> | 88.3 (16.1)*                 | 87.4 (16.4)*                 | 0.176   |
| Lipid-lowering medication, N (%)               | 1308 (74.7%)                 | 750 (75.8%)                  | 0.556   |
| Anti-hypertensive medication, N (%)            | 1087 (62.1%)                 | 632 (63.9%)                  | 0.374   |

Abbreviations: eGFR, estimated Glomerular Filtration Rate; HbA<sub>1c</sub>, glycated hemoglobin; HDL, high-density lipoprotein

<sup>a</sup>The eGFR was calculated with the CKD-EPI 2009 equation

Values are expressed as mean (standard deviation) for continuous variables, and as N (%) for categorical variables. P-values were calculated using t-tests for continuous variables and  $\chi^2$  tests for categorical variables

\*In the proteomics cohort (N = 1751), the following variables contained missing data before imputation: HDL cholesterol (210 missing, 12.0%; data available for 1541 participants), eGFR (105 missing, 6.0%; data available for 1646 participants), HbA<sub>1c</sub> (120 missing, 6.9%; data available for 1631 participants), total cholesterol (101 missing, 5.8%; data available for 1650 participants), age at diabetes diagnosis (115 missing, 6.6%; data available for 1636 participants), current smoking status (14 missing, 0.8%; n = 1737), and systolic blood pressure (5 missing, 0.3%; data available for 1746 participants). All other baseline variables were complete

#In the multiomics cohort (N = 990), missingness was similar: HDL cholesterol (118 missing, 11.9%; n = 872), eGFR (62 missing, 6.3%; n = 928), HbA<sub>1c</sub> (67 missing, 6.8%; n = 923), total cholesterol (56 missing, 5.7%; n = 934), age at diabetes diagnosis (63 missing, 6.4%; n = 927), systolic blood pressure (3 missing, 0.3%; n = 987), and current smoking status (8 missing, 0.8%; n = 982). All other baseline variables were complete

were sourced from large-scale international consortia, primarily CARDIoGRAMplusC4D and MEGASTROKE, where genotyping was conducted on a variety of Illumina and Affymetrix arrays, with Illumina platforms being predominant [26]. Subsequently, it was optimized using individual-level data from the UK Biobank cohort, for which genotyping was conducted using the UK Biobank Axiom and UK BiLEVE Axiom arrays [15]. According to internal validation analyses, this CVD-PRS demonstrates robust predictive performance and is superior to previously published scores [15, 27].

### Outcome ascertainment

The primary outcome was the occurrence of a MACE, defined in accordance with the SCORE2-Diabetes model as a composite of cardiovascular death, non-fatal myocardial infarction, or non-fatal stroke [7]. Non-fatal events were identified through linked primary care records and hospital inpatient data obtained via the Hospital Episode Statistics system. Information on dates and causes of death was obtained from national death registries, specifically the National Health Service (NHS) Information Centre for England and Wales and the NHS Central Register for Scotland. Participants were followed from baseline until the earliest of the following: occurrence of a MACE event, death from any cause, or completion of the 10-year follow-up period. A list of diagnostic codes and outcome definitions is provided in Supplemental Table S2.

### Statistical analyses

#### General remarks

All analyses were performed using R software (version 4.4.0, R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was defined as P-values < 0.05 for two-sided tests. Missing data in SCORE2-Diabetes variables (with HDL-C having the highest missing rate of 12.0%) and multi-omics measurements (mostly complete, with up to 20% missing for a few proteins) were single imputed using the chained equations method with random forest algorithms implemented in the R package *missForest* (version 1.5) [26]. The missing values per variable are shown in the legend of Table 1. To the best of our knowledge, all missing values were missing at random. The variable distributions did not change by missing value imputation. In sensitivity analysis, multiple imputation of 5 data sets led to almost identical associations as *MissForest* imputation (data not shown). All Cox proportional hazards models accounted for the case-cohort design using inverse-probability Barlow weights [27].



### **Evaluation of model performance and incremental value of multi-omics integration**

Model development and evaluation followed a two-stage approach. In the first stage, proteomics-based models were derived and evaluated in models adjusted for the variables of the SCORE2-Diabetes model using the subset of T2D patients with available proteomics data ( $N=1751$ ) for best statistical power. To obtain stable and parsimonious protein signatures, we applied a bootstrap-LASSO stability-selection protocol. Specifically, for each sex, we generated 1000 bootstrap resamples and fitted a LASSO-penalized Cox model in each resample, with the regularization parameter ( $\lambda$ ) determined by ten-fold cross-validation selecting the value that minimized the cross-validated error ( $\lambda_{\min}$ ). Proteins selected in at least 95% of the bootstrap iterations were retained to ensure robustness against stochastic variation. The retained proteins were then refit in unpenalized Cox models adjusted for all SCORE2-Diabetes variables to obtain final coefficients and model performance.

In the second stage, the additional predictive value of signatures derived from genomic, metabolomic, and proteomic data (Supplemental Table S1), when added to the SCORE2-Diabetes model, was assessed in the subset of T2D patients with complete multi-omics data ( $N=990$ ). These three multi-omics layers were added sequentially to the SCORE2-Diabetes model, and model performance was evaluated. Finally, the best-performing omics layers were added one by one to the SCORE2-Diabetes model until the addition of another layer did not result in a statistically significant improvement in model discrimination.

Model discrimination was evaluated using Harrell's C-index and statistical significance for differences between correlated C-indices was assessed using the method proposed by Kang et al., implemented in the R package *compareC* (version 1.3.2) [28]. Moreover, the incremental C-index for each individual biomarker was estimated to assess its independent contribution to model discrimination.

Additionally, risk reclassification was quantified using the net reclassification index (NRI) and integrated discrimination index (IDI) [29]. Predefined cardiovascular risk categories (0–15%, > 15–30%, and > 30%) were used to calculate the NRI, comparing the proportion of correctly reclassified individuals against the SCORE2-Diabetes model [7]. Model calibration was assessed by plotting observed versus predicted MACE event rates across deciles of predicted absolute risk.

### **Associations of selected multi-omics biomarkers with MACE**

To quantify the association between individual biomarkers and 10-year MACE risk, each selected biomarker was entered separately into a Cox proportional hazards

model, adjusted for the SCORE2-Diabetes variables. Hazard ratios (HRs) and 95% confidence intervals (CIs) per one standard deviation increment were reported, stratified by sex. Models were fitted using the *survival* package (version 3.5-5) in R.

### **Correlation matrix**

To assess redundancy among selected biomarkers, Spearman correlation coefficients were calculated for the selected proteins, metabolites, and the CVD-PRS in males and females separately.

## **Results**

### **Baseline characteristics and MACE incidence**

Table 1 summarizes the baseline characteristics of the 1751 UKB participants with T2D included in the proteomics-based analysis cohort and the 990 participants in the multi-omics cohort. In the proteomics cohort, the mean age was  $60.3 \pm 6.6$  years and 59.7% of participants were male. The multi-omics cohort was similar, with a mean age of  $60.5 \pm 6.6$  years and 60.3% male participants. The distributions of smoking status, SBP, TC, HDL-C, age at diabetes diagnosis, HbA<sub>1c</sub>, eGFR, use of lipid-lowering and anti-hypertensive medication were also comparable between the two cohorts, with all  $P$ -values > 0.05. During the 10-year follow-up, 270 participants (15.4%) in the proteomics cohort experienced a MACE, including 193 males and 77 females. In the multi-omics cohort, 159 participants (16.1%) developed MACE, among whom 114 were males and 45 females.

### **Predictive performance of proteomics extended SCORE2-Diabetes model**

Overall, 15 proteomic biomarkers were selected by LASSO regression for their ability to enhance the predictive performance of the SCORE2-Diabetes model: 8 were male-specific, 6 were female-specific, and one, WFDC2 (WAP four-disulfide core domain protein 2), was common to both sexes (see receiver operating characteristic (ROC) curves in Supplemental Fig. S2). Supplemental Table S3 presents the  $\beta$ -coefficients of the proteomics-extended SCORE2-Diabetes model. As shown in Table 2, the inclusion of these proteins markedly improved model discrimination. The C-index increased from 0.725 to 0.810 ( $\Delta$ C-index, 0.085;  $P < 0.001$ ) in the total population, with similarly large improvements in males (from 0.700 to 0.797;  $\Delta$ C-index, 0.097;  $P < 0.001$ ) and females (from 0.740 to 0.835;  $\Delta$ C-index, 0.095;  $P < 0.001$ ). This gain in discrimination was corroborated by a substantial improvement in risk reclassification, highlighted by a categorical NRI of 24.2%, continuous NRI of 31.6%, and a significant increase in the IDI of 0.098 for the overall cohort. The calibration curve illustrated good agreement between observed and predicted risks, highlighting

**Table 2** Metrics of the predictive performance of the SCORE2-Diabetes model for 10-year MACE risk without and with extension by proteins (N=1751)

| Metrics                               | Male <sup>a</sup>           | Female <sup>b</sup>         | Overall                     |
|---------------------------------------|-----------------------------|-----------------------------|-----------------------------|
| C-Statistics (SCORE2-Diabetes)        | 0.700 (0.659, 0.742)        | 0.740 (0.682, 0.799)        | 0.725 (0.697, 0.757)        |
| C-Statistics (+ Proteins)             | 0.797 (0.760, 0.833)        | 0.835 (0.789, 0.881)        | 0.810 (0.782, 0.837)        |
| P-values for C-Statistics comparisons | <0.001                      | <0.001                      | <0.001                      |
| Categorical NRI total (%)             | <b>51.2 (18.3, 67.9)</b>    | <b>29.5 (18.4, 42.1)</b>    | <b>24.2 (15.9, 33.1)</b>    |
| Categorical NRI events (%)            | −12.6 (−17.1, 6.8)          | <b>34.9 (24.5, 47.1)</b>    | 19.1 (−2.1, 37.6)           |
| Categorical NRI non-events (%)        | <b>63.9 (18.6, 78.1)</b>    | −5.4 (−7.6, −2.4)           | 5.1 (−1.6, 28.3)            |
| Continuous NRI (%)                    | <b>32.3 (27.4, 39.6)</b>    | <b>36.8 (25.1, 47.1)</b>    | <b>31.6 (25.6, 38.5)</b>    |
| IDI                                   | <b>0.105 (0.058, 0.146)</b> | <b>0.111 (0.028, 0.210)</b> | <b>0.098 (0.052, 0.130)</b> |

Bold values indicate statistically significant differences between models (P < 0.05)

Abbreviations: IDI, integrated discrimination index; MACE, major cardiovascular events; NRI, net reclassification index

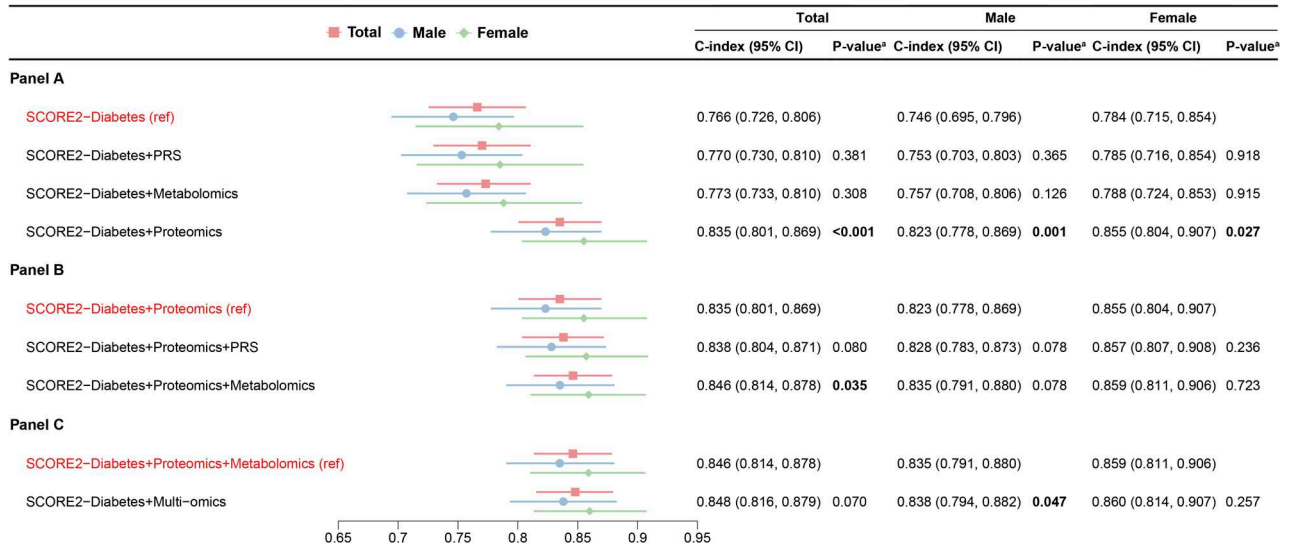
<sup>a</sup>Proteins included in the combined model for males were: ACTA2, CHGA, CRNN, ENPP5, GAST, NT-proBNP, POSTN, TRIM21, WFDC2

<sup>b</sup>Proteins included in the combined model for females were: CTRC, EFHD1, HS3ST3B1, IL22, SMOC1, TNC, WFDC2

the reliable performance of the proteomics extended SCORE2-Diabetes model (Supplemental Fig. S3).

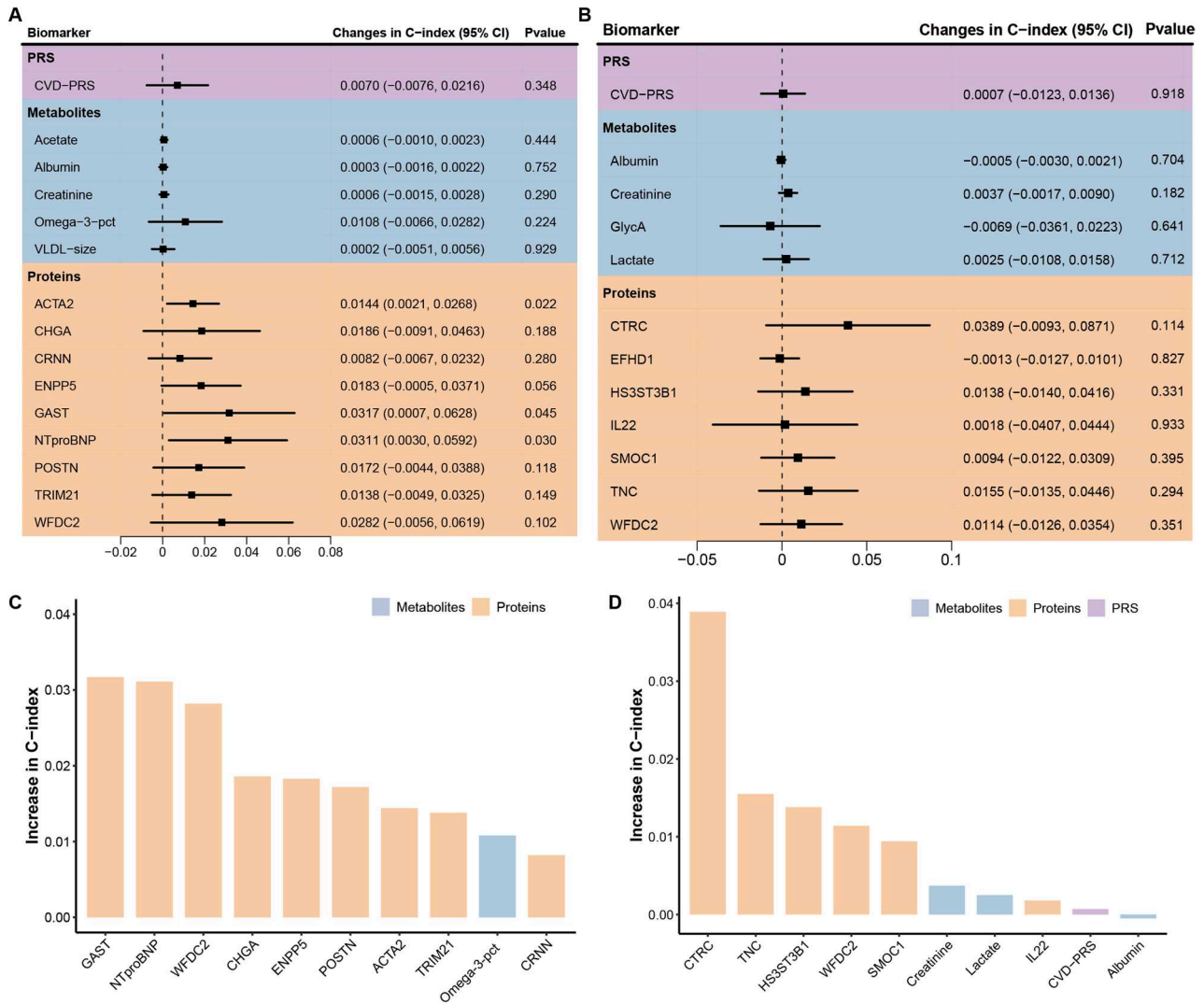
**Predictive performance of multi-omics extended SCORE2-Diabetes model**

Figure 1 evaluates which combination of omics data yielded the highest C-index for 10-year MACE prediction. Supplemental Table S5 adds the reclassification statistics NRI and IDI for these model comparisons. Among all single-omics extensions to the SCORE2-Diabetes model, proteomics led to the largest improvement in the total population (C-index: 0.835 vs. 0.766; ΔC-index, 0.069; *P*<0.001), in males (0.823 vs. 0.746; ΔC-index, 0.077; *P*=0.001), and in females (0.855 vs. 0.784; ΔC-index, 0.071; *P*=0.027), and was therefore selected as the reference model for subsequent comparisons. Using the proteomics-extended SCORE2-Diabetes model as the reference, the addition of the CVD-PRS did not significantly improve model performance in the total population, in males, or in females. The inclusion of metabolomic biomarkers led to a significant increase in the C-index in the total population (C-index: 0.846 vs. 0.835; ΔC-index, 0.011; *P*=0.035; NRI: −0.8%, *P*=0.811; IDI: 0.001, *P*=0.812), which was highly attributable to the model improvement among males, which almost reached statistical significance in C-index increase and NRI but not IDI statistics (C-index: 0.835 vs. 0.823; ΔC-index, 0.012; *P*=0.078; NRI: 9.8%, *P*=0.086; IDI: −0.005, *P*=0.792). Further incorporation of the CVD-PRS into the model combining SCORE2-Diabetes variables, proteomics, and metabolomics did not significantly enhance predictive performance in the total population or females. Although a marginal increase in the



**Fig. 1** Discrimination performance of single-omics and multi-omics extensions to the SCORE2-Diabetes in 10-year MACE risk prediction (N=990). <sup>a</sup>*P*-values indicate statistical comparisons of C-indices against the corresponding reference model

Panel **A** shows the C-index with 95% confidence intervals for models extending SCORE2-Diabetes with individual omics layers (PRS, metabolomics, or proteomics). Panel **B** compares multi-omics combinations using SCORE2-Diabetes+proteomics as the reference. Panel **C** contrasts the full multi-omics model with the SCORE2-Diabetes+proteomics+metabolomics model as the reference. CI, confidence interval; PRS, polygenic risk score; MACE, major adverse cardiovascular event



**Fig. 2** Incremental predictive value of selected multi-omic biomarkers added to the SCORE2-Diabetes model for 10-year MACE risk prediction ( $N=990$ ). Panel **A** and **B** show the changes in C-index when each biomarker is individually added to the SCORE2-Diabetes model in male and female participants. Panel **C** and **D** illustrate the top 10 biomarkers ranked by individual C-index gains in males and females. ACTA2, Actin Alpha 2; CHGA, Chromogranin-A; CRNN, Cornulin; CTRC, Chymotrypsin-C; CVD, cardiovascular disease; EFHD1, EF-hand domain-containing protein D1; ENPP5, Ectonucleotide pyrophosphatase/phosphodiesterase family member 5; GAST, Gastrin; HS3ST3B1, Heparan sulfate glucosamine 3-O-sulfotransferase 3B1; IL22, integrated discrimination improvement; IL22, Interleukin-22; MACE, major adverse cardiovascular event; NRI, net reclassification improvement; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; Omega-3-pct, The Omega-3 fatty acids percentage of total fatty acids; PRS, polygenic risk score; SMOC1, SPARC-related modular calcium-binding protein 1; TNC, Tenascin-C; TRIM21, E3 ubiquitin-protein ligase TRIM21; VLDL-size, Average diameter for very-low-density lipoprotein particles; WFDC2, WAP four-disulfide core domain protein

C-index was statistically significant in males, the reclassification statistics did not indicate a model improvement ( $\Delta$ C-index, 0.003;  $P=0.047$ ; NRI:  $-2.6\%$ ,  $P=0.220$ ; IDI:  $-0.002$ ,  $P=0.495$ ). Supplemental Table S5 presents the  $\beta$ -coefficients for the variables included in the sex specific, multi-omics extended SCORE2-Diabetes models. Supplemental Fig. S4 shows the ROC curves and Supplemental Fig. S5 the model calibration curves of the SCORE2-Diabetes with and without extension by multi-omics data.

**Incremental contributions of individual biomarkers to improved model performance**

Figure 2A and B present the increase in C-index of each individual selected multi-omics biomarker when added to the SCORE2-Diabetes model for males and females, respectively. In males, GAST (Gastrin) and NT-proBNP (N-Terminal Pro-Brain Natriuretic Peptide) conferred the largest individual increase in the C-index ( $\Delta$ C-index =  $+0.0317$  and  $+0.0311$ , respectively). In females, CTRC (Chymotrypsin-C) was the most influential biomarker ( $\Delta$ C-index =  $+0.0389$ ). In males, 9 out

of the 10 most important contributors were proteins (Fig. 2C), while in females, it was 6 out of 10 (Fig. 2D). However, the top 5 predictors among females were all proteins, which further underscores the higher predictive ability of proteomic biomarkers compared to metabolomic and genomic markers for both sexes.

#### Associations of selected biomarkers with MACE incidence

Supplemental Figure S6 shows hazard ratios for the associations between the selected multi-omics biomarkers and 10-year MACE risk in male and female participants, respectively. In both sexes, all selected proteins except one (CRNN (Cornulin) in males) were significantly associated with MACE. In contrast, all selected metabolites except one (omega-3-pct (Omega-3 fatty acids percentage of total fatty acids) in males) were not significantly associated with MACE in both sexes. The CVD-PRS was only significantly associated with MACE among males.

#### Correlations among selected biomarkers

Supplemental Figure S7 displays the Spearman correlation matrices for the selected proteins, metabolites, and the CVD-PRS in males and females, respectively. Correlations between biomarkers across omics layers were weak overall. No correlation coefficient exceeded 0.1 between the CVD-PRS and any selected protein or metabolite in either sex. With one exception (CHGA (Chromogranin-A) and GAST among males ( $r=0.64$ )), no pairwise correlation exceeded  $r=0.47$ .

## Discussion

#### Summary of the findings

In this study of individuals with T2D from the UKB, we developed sex-specific risk models that integrated proteomic biomarkers into the established SCORE2-Diabetes algorithm to improve 10-year prediction of MACE. A total of 15 proteins were selected using a bootstrap-LASSO approach, including one protein (WFDC2) shared across sexes, 8 specific to males, and 6 specific to females. The inclusion of these proteomic signatures markedly enhanced model discrimination for both sexes. In the total population, an additional gain in model performance was observed by further adding metabolites, which was attributable to model improvement among males but not among females. Further adding the CVD-PRS slightly improved the model's C-index but not reclassification statistics among men but not women. This sex difference was supported by a lack of statistically significant associations of all selected metabolites and the CVD-PRS with MACE incidence in females, whereas the CVD-PRS and one metabolite showed statistically significant associations with MACE incidence in males.

#### Comparison with previous studies

To our knowledge, this is the first study to develop sex-specific proteomic signatures for cardiovascular risk prediction in individuals with T2D. Most previous studies with sex-specific analyses have investigated proteomic biomarkers in general populations with limited attention to individuals with T2D [30–33]. These studies often overlooked the distinct cardiovascular risk profiles and pathophysiological mechanisms unique to populations with diabetes. In our prior analysis using the UKB cohort without diabetes ( $N=47,382$ ), we applied the same sex-specific biomarker selection framework and identified 18 proteins associated with MACE [34]. Interestingly, only two proteins overlapped with the current diabetes-specific results (NT-proBNP and WFDC2 in males; WFDC2 alone in females). This limited overlap in selected proteins underscores the need to develop proteomic risk models tailored to individuals with T2D.

Prior proteomics-based cardiovascular risk prediction studies in populations with diabetes have utilized less extensive protein panels without sex-specific biomarker selection [13, 14, 35–38]. For example, the SUMMIT study employed 6 plasma proteins across five European T2D cohorts ( $N=2310$ ) and observed a notable improvement in AUC from 0.59 to 0.72 for 5-year cardiovascular risk prediction [13]. Similarly, Nowak et al. utilized 20 proteins measured via the Olink platform in pooled Swedish T2D cohorts ( $N=1211$ ), and achieved a C-index increase from 0.686 to 0.747 [35]. The EXAMINE trial ( $N=5131$ ), which focused on T2D patients with recent myocardial infarction, demonstrated that adding a protein panel including NT-proBNP and Galectin-9 improved the C-index by over 0.05 and significantly enhanced reclassification metrics [36]. Although these studies support the utility of proteomics in T2D populations, their lack of sex-specific biomarker selection potentially limits their predictive precision.

Our study specifically developed sex-specific proteomic signatures for cardiovascular risk prediction among individuals with T2D. Unlike previous studies that required extensive protein panels, we achieved comparable or superior predictive accuracy with a concise set of biomarkers (9 proteins for males and 7 for females). Our study was the first to test whether additional inclusion of metabolomics and genomic data could further improve MACE prediction in patients with T2D. Further studies are needed to corroborate our results.

#### Potential biological mechanisms for the selected proteins in CVD development

The 15 proteins identified in this study are involved in a range of biological processes relevant to cardiovascular risk in individuals with T2D, including pathways related to cardiac stress, fibrosis, inflammation, and metabolic



dysfunction. Several proteins, including NT-proBNP, CHGA, and EFHD1 (EF-hand domain-containing protein D1), are associated with cardiac stress and neurohumoral signaling [39–41]. NT-proBNP is a well-established biomarker of myocardial wall stress and subclinical heart dysfunction, extensively validated in populations with and without diabetes [42, 43]. CHGA is a prohormone co-released with catecholamines that influences autonomic regulation and vascular tone through active peptides such as catestatin [40]. EFHD1 is a mitochondrial calcium-binding protein linked to oxidative stress and mitochondrial dysfunction, which may contribute to cardiomyocyte injury in metabolically impaired states such as T2D [41].

Several other proteins are involved in fibrosis and extracellular matrix remodeling, processes known to drive plaque instability and adverse ventricular remodeling. WFDC2, the sole protein selected in both sexes in our study, is associated with myocardial and renal fibrosis, potentially via mechanisms involving collagen deposition regulation and protease inhibition [44, 45]. TNC (Tenascin-C), POSTN (Periostin), ACTA2 (Actin Alpha 2), and SMOC1 (SPARC-related modular calcium-binding protein 1) are matricellular proteins or contractile regulators that influence fibroblast activation, vascular smooth muscle cell proliferation, and extracellular matrix turnover [46–49]. These processes contribute to vascular stiffening, neointimal hyperplasia, and fibrotic remodeling, particularly relevant in T2D patients who experience accelerated atherosclerosis and impaired tissue repair [50].

Inflammatory and immune-regulatory mechanisms are also reflected among the selected proteins. IL22 (Interleukin-22) is a pleiotropic cytokine involved in mucosal immunity and tissue regeneration; however, in chronic inflammatory states such as T2D, it has been associated with endothelial dysfunction and atherogenesis [51]. TRIM21 (E3 ubiquitin-protein ligase TRIM21) functions as an intracellular Fc receptor and E3 ligase, regulating immune signaling and oxidative stress pathways associated with vascular inflammation [52]. CRNN is an epithelial stress-response protein associated with inflammation, potentially reflecting systemic inflammatory burden or metabolic stress relevant to cardiovascular risk in T2D [53].

Several proteins are involved in metabolic-vascular regulation and endothelial homeostasis. GAST is primarily known for its role in gastrointestinal function but has also been linked to hypertension and subclinical atherosclerosis, possibly through stimulation of vascular smooth muscle [54]. CTRC is a pancreatic serine protease that may reflect systemic proteolytic activity and metabolic disturbances in diabetes, and has recently been linked to coronary artery disease in a sex-specific manner

[55, 56]. This may reflect a broader disruption of protease-antiprotease balance in T2D, potentially affecting endothelial integrity or vascular remodeling, although the precise mechanisms remain unclear. ENPP5 (Ectonucleotide pyrophosphatase/phosphodiesterase family member 5) is part of the ectonucleotide pyrophosphatase family and has been implicated in cellular senescence, vascular stiffening, and low-grade inflammation [57]. HS3ST3B1 (Heparan sulfate glucosamine 3-O-sulfotransferase 3B1) encodes a 3-O-sulfotransferase that modifies the fine structure of heparan sulfate chains [58]. By altering sulfation patterns on endothelial surfaces, it may indirectly modulate leukocyte adhesion and inflammatory signaling, as heparan sulfate structure is known to influence cytokine binding and vascular homeostasis. These mechanisms may underlie early metabolic-endothelial dysfunction in T2D and help explain the elevated cardiovascular risk observed in this population.

#### Clinical translation and feasibility

Previous large-scale proteomics studies often selected a large number of proteins, significantly increasing measurement complexity, cost, and limiting clinical feasibility [59, 60]. Our approach substantially reduces this burden by utilizing concise, sex-specific protein panels (9 proteins for males and 7 for females). For example, it would be possible to design an Olink Focus panel, which allows simultaneous measurement of up to 21 proteins with absolute concentrations (<https://olink.com/products/olink-focus>). After relatively high investment costs once to develop such a panel, the scaling up of measuring it in clinical routine is feasible at low costs. The current costs are approx. €67 per sample but are expected to decrease if assays are translated from scientific use for a few samples to large-scale use in clinical routine. As our results need to be externally validated anyway before they can be translated in clinical routine, it would be ideal if further studies would use PEA or a comparable technology to measure the selected proteins with absolute concentrations.

Our results suggest that integrating further multi-omics layers may improve the predictive accuracy in males, but the clinical translation of these additional biomarker platforms is challenging with regard to costs and timely measurement, as additional laboratories would need to be involved [61]. Future studies should carefully weigh the incremental predictive gains (which were particularly low for the CVD-PRS) by additional omics biomarker measurements in men against the associated measurement costs and feasibility. In women, the marginal improvements in model accuracy did not reach statistical significance in our study, which suggests that a protein-extended SCORE2-Diabetes model is sufficient for MACE risk assessment in women with T2D.

### Strengths and limitations

This study's strengths include the availability of multi-omics data from proteomics, metabolomics, and genomics in a large sample of people with T2D with 10-year MACE follow-up. The use of sex-specific modeling accounted for biological differences between males and females and represents an incremental step toward personalized cardiovascular disease prevention.

Several limitations must also be considered. First, a limitation is that the number of MACE cases was not sufficient for a separate internal validation using, for instance, a data split into dedicated training and testing sets or k-fold cross-validation. Consequently, the reported improvements in predictive performance may be overly optimistic, and replication in independent cohorts is necessary to confirm generalizability [62]. Second, the metabolomics data derived from the Nightingale Health platform were limited to approximately 250 biomarkers, and more comprehensive, untargeted metabolomics approaches based on mass spectrometry might yield superior predictive performance. Third, UK Biobank participants are almost entirely of European ancestry, with approximately 95% being White and around 80% being White British, and were aged 40–69 years at recruitment. As these volunteers are generally healthier than the broader population with diabetes, the direct extrapolation of our findings to non-European ethnic groups, older individuals, or higher risk secondary prevention cohorts is limited. Fourth, although beta coefficients are provided for the proteomics-extended and multi-omics data extended SCORE2-Diabetes model for external validation, recalibration may be necessary when these models are applied to populations with different baseline CVD risk or if alternative omics measurement platforms were used. Finally, an in-depth cost-benefit or cost-effectiveness analysis about implementing either the proteomics-extended or multi-omics data extended risk models into routine clinical practice was beyond the scope of this study but remains essential.

### Conclusion

In this study, we identified 9 proteins in men and 7 proteins in women that improved 10-year MACE prediction beyond the SCORE2-Diabetes model in individuals with T2D. Among the omics layers, proteomics provided the largest incremental gain in predictive accuracy. In women, adding metabolomics or a CVD-PRS did not further enhance the performance of the protein-extended model. In men, incorporating metabolomics showed additional model improvement, whereas further adding a CVD-PRS did only marginally improve its performance. External validation and cost-effectiveness analyses are now warranted before these improved SCORE2-Diabetes models can be recommended for clinical routine.

We hypothesize that adding the measurement of the few selected proteins specific for males or females might be feasible and cost-effective in clinical practice. These concise, sex-specific biomarker panels would be a first step toward more precise cardiovascular risk stratification.

### Abbreviations

|                   |                                                                  |
|-------------------|------------------------------------------------------------------|
| ACTA2             | Actin Alpha 2                                                    |
| CHGA              | Chromogranin-A                                                   |
| CI                | Confidence intervals                                             |
| CRNN              | Cornulin                                                         |
| CTRC              | Chymotrypsin-C                                                   |
| CVD               | Cardiovascular disease                                           |
| EFHD1             | EF-hand domain-containing protein D1                             |
| ENPP5             | Ectonucleotide pyrophosphatase/phosphodiesterase family member 5 |
| eGFR              | Estimated glomerular filtration rate                             |
| GAST              | Gastrin                                                          |
| GWAS              | Genome-wide association study                                    |
| HbA <sub>1c</sub> | Glycated hemoglobin                                              |
| HDL-C             | High-density lipoprotein cholesterol                             |
| HR                | Hazard ratio                                                     |
| HS3ST3B1          | Heparan sulfate glucosamine 3-O-sulfotransferase 3B1             |
| IDI               | Integrated discrimination index                                  |
| LASSO             | Least absolute shrinkage and selection operator                  |
| MACE              | Major adverse cardiovascular events                              |
| NHS               | National health service                                          |
| NMR               | Nuclear magnetic resonance                                       |
| NRI               | Net reclassification index                                       |
| NT-proBNP         | N-terminal prohormone of brain natriuretic peptide               |
| POSTN             | Periostin                                                        |
| PRS               | Polygenic risk score                                             |
| ROC               | Receiver operating characteristic                                |
| SBP               | Systolic blood pressure                                          |
| SD                | Standard deviation                                               |
| SMOC1             | SPARC-related modular calcium-binding protein 1                  |
| T2D               | Type 2 diabetes                                                  |
| TC                | Total cholesterol                                                |
| TNC               | Tenascin-C                                                       |
| UKB               | UK Biobank                                                       |
| WFDC2             | WAP four-disulfide core domain protein 2                         |

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12933-025-03036-5>.

Supplementary Material 1

### Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 101633. This work uses data provided by patients and collected by the NHS as part of their care and support. We would like to thank all participants of the UK Biobank as well as the staff of the UK Biobank assessment centers for their contributions. The first author (R.X.) was supported by the China Scholarship Council (No. 202208430012).

### Author contributions

R.X. and B.S. generated the idea for the study and formulated the analytical plan. R.X. performed the data analyses and drafted the manuscript. B.S. revised it. All authors contributed valuable intellectual content to the discussion. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

### Funding

Open Access funding enabled and organized by Projekt DEAL. The UK Biobank was established by the Wellcome Trust, Medical Research Council, Department of Health, Scottish government, and Northwest Regional Development Agency, and the Welsh assembly government and the British

Heart Foundation. The German Diabetes Center is supported by the Ministry of Culture and Science of the State of North Rhine-Westphalia (Düsseldorf, Germany) and the German Federal Ministry of Health (Berlin, Germany). The German Diabetes Center is also supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD). The sponsors had no role in data acquisition or the decision to publish the data.

#### Data availability

Data from the UK Biobank are available to bona fide researchers upon application through the UK Biobank Access Management System (<https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>).

#### Declarations

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Division of Clinical Epidemiology of Early Cancer Detection, German Cancer Research Center, Im Neuenheimer Feld 581, 69120 Heidelberg, Germany

<sup>2</sup>Faculty of Medicine, Heidelberg University, 69115 Heidelberg, Germany

<sup>3</sup>Institute for Clinical Diabetology, German Diabetes Center (DDZ), Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany

<sup>4</sup>German Center for Diabetes Research (DZD), Partner Düsseldorf, Munich-Neuherberg, Germany

<sup>5</sup>Department of Endocrinology and Diabetology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

<sup>6</sup>Network Aging Research, Heidelberg University, Heidelberg, Germany

<sup>7</sup>Division of Urological Cancers, Department of Translational Medicine, Medical Faculty, Lund University, Lund, Sweden

Received: 9 October 2025 / Accepted: 2 December 2025

Published online: 24 December 2025

#### References

- Wong ND, Sattar N. Cardiovascular risk in diabetes mellitus: epidemiology, assessment and prevention. *Nat Rev Cardiol*. 2023;20(10):685–95.
- Joseph JJ, Deedwania P, Acharya T, Aguilar D, Bhatt DL, Chyun DA, Di Palo KE, Golden SH, Sperling LS, Lifestyle, AHADCoTCo, et al. Comprehensive management of cardiovascular risk factors for adults with type 2 diabetes: a scientific statement from the American heart association. *Circulation*. 2022;145(9):e722–59.
- Wright AK, Suarez-Ortega MF, Read SH, Kontopantelis E, Buchan I, Emsley R, Sattar N, Ashcroft DM, Wild SH, Rutter MK. Risk factor control and cardiovascular event risk in people with type 2 diabetes in primary and secondary prevention settings. *Circulation*. 2020;142(20):1925–36.
- O'Meara E, Clayton T, McEntegart MB, McMurray JJ, Piña IL, Granger CB, Östergren J, Michelson EL, Solomon SD, Pocock S. Sex differences in clinical characteristics and prognosis in a broad spectrum of patients with heart failure: results of the Candesartan in heart failure: assessment of reduction in mortality and morbidity (CHARM) program. *Circulation*. 2007;115(24):3111–20.
- Peters SA, Huxley RR, Woodward M. Diabetes as risk factor for incident coronary heart disease in women compared with men: a systematic review and meta-analysis of 64 cohorts including 858,507 individuals and 28,203 coronary events. *Diabetologia*. 2014;57(8):1542–51.
- Huxley R, Barzi F, Woodward M. Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies. *BMJ*. 2006;332(7533):73–8.
- Pennells L, Kaptoge S, Ostergaard HB, Read SH, Carinci F, Franch-Nadal J, Petitjean C, Taylor O, Hageman SH, Xu Z. SCORE2-Diabetes: 10-year cardiovascular risk Estimation in type 2 diabetes in Europe. *Eur Heart J*. 2023;44(28):2544–56.
- Tokgozoglul L, Torp-Pedersen C. Redefining cardiovascular risk prediction: Is the crystal ball clearer now? *Eur Heart J*. 2021;42(25):2468–71.
- Matsushita K, Kaptoge S, Hageman SH, Sang Y, Ballew SH, Grams ME, Surapaneni A, Sun L, Arnlov J, Bozic M, et al. Including measures of chronic kidney disease to improve cardiovascular risk prediction by SCORE2 and SCORE2-OP. *Eur J Prev Cardiol*. 2022;30(1):8–16.
- Tahir UA, Gerszten RE. Omics and cardiometabolic disease risk prediction. *Annu Rev Med*. 2020;71:163–75.
- Carrasco-Zanini J, Pietzner M, Davitte J, Surendran P, Croteau-Chonka DC, Robins C, Torralbo A, Tomlinson C, Grünschläger F, Fitzpatrick N, et al. Proteomic signatures improve risk prediction for common and rare diseases. *Nat Med*. 2024;30(9):2489–98.
- Chen ZZ, Gerszten RE. Metabolomics and proteomics in type 2 diabetes. *Circ Res*. 2020;126(11):1613–27.
- Looker HC, Colombo M, Agakov F, Zeller T, Groop L, Thorand B, Palmer CN, Hamsten A, de Faire U, Nogoceke E, et al. Protein biomarkers for the prediction of cardiovascular disease in type 2 diabetes. *Diabetologia*. 2015;58(6):1363–71.
- Luo H, Huemer MT, Petrera A, Hauck SM, Rathmann W, Herder C, Koenig W, Hoyer A, Peters A, Thorand B. Association of plasma proteomics with incident coronary heart disease in individuals with and without type 2 diabetes: results from the population-based KORA study. *Cardiovasc Diabetol*. 2024;23(1):53.
- Lau ES, Paniagua SM, Guseh JS, Bhambhani V, Zanni MV, Courchesne P, Lyass A, Larson MG, Levy D, Ho JE. Sex differences in circulating biomarkers of cardiovascular disease. *J Am Coll Cardiol*. 2019;74(12):1543–53.
- Wright AK, Kontopantelis E, Emsley R, Buchan I, Mamas MA, Sattar N, Ashcroft DM, Rutter MK. Cardiovascular risk and risk factor management in type 2 diabetes mellitus: a population-Based cohort study assessing sex disparities. *Circulation*. 2019;139(24):2742–53.
- Garg M, Karpinski M, Matelska D, Middleton L, Burren OS, Hu F, Wheeler E, Smith KR, Fabre MA, Mitchell J. Disease prediction with multi-omics and biomarkers empowers case-control genetic discoveries in the UK biobank. *Nat Genet*. 2024;56(9):1821–31.
- Xie R, Seum T, Sha S, Trares K, Holleczeck B, Brenner H, Schöttker B. Improving 10-year cardiovascular risk prediction in patients with type 2 diabetes with metabolomics. *Cardiovasc Diabetol*. 2025;24(1):18.
- Thompson DJ, Wells D, Selzam S, Peneva I, Moore R, Sharp K, Tarran WA, Beard EJ, Riveros-Mckay F, Giner-Delgado C, et al. A systematic evaluation of the performance and properties of the UK biobank polygenic risk score (PRS) release. *PLoS ONE*. 2024;19(9):e0307270.
- Littlejohns TJ, Sudlow C, Allen NE, Collins R. UK biobank: opportunities for cardiovascular research. *Eur Heart J*. 2019;40(14):1158–66.
- O'Brien KM, Lawrence KG, Keil AP. The case for case-Cohort: an applied epidemiologist's guide to reframing case-Cohort studies to improve usability and flexibility. *Epidemiology*. 2022;33(3):354–61.
- Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG. Plasma proteomic associations with genetics and health in the UK biobank. *Nature*. 2023;622(7982):329–38.
- Wik L, Nordberg N, Broberg J, Björkstén J, Assarsson E, Henriksson S, Grundberg I, Pettersson E, Westerberg C, Liljeroth E. Proximity extension assay in combination with next-generation sequencing for high-throughput proteome-wide analysis. *Mol Cell Proteom* 2021, 20.
- Xie R, Vlaski T, Trares K, Herder C, Holleczeck B, Brenner H, Schöttker B. Large-Scale proteomics improve risk prediction for type 2 diabetes. *Diabetes Care* 2025.
- Ritchie SC, Surendran P, Karthikeyan S, Lambert SA, Bolton T, Pennells L, Danesh J, Di Angelantonio E, Butterworth AS, Inouye M. Quality control and removal of technical variation of NMR metabolic biomarker data in ~ 120,000 UK biobank participants. *Sci Data*. 2023;10(1):64.
- Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112–8.
- Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. *J Clin Epidemiol*. 1999;52(12):1165–72.
- Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med*. 2015;34(4):685–703.
- Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk-prediction instruments: a critical review. *Epidemiol (Cambridge Mass)*. 2014;25(1):114.
- Hoogeveen RM, Pereira JPB, Nurmoheamed NS, Zampolieri V, Bom MJ, Baragetti A, Boekholdt SM, Knaapen P, Khaw KT, Wareham NJ, et al. Improved cardiovascular risk prediction using targeted plasma proteomics in primary prevention. *Eur Heart J*. 2020;41(41):3998–4007.

31. Helgason H, Eiriksdottir T, Ulfarsson MO, Choudhary A, Lund SH, Ivarsdottir EV, Hjorleifsson Eldjarn G, Einarsson G, Ferkingstad E, Moore KHS, et al. Evaluation of Large-Scale proteomics for prediction of cardiovascular events. *JAMA*. 2023;330(8):725–35.
32. Nurmohamed NS, Belo Pereira JP, Hoogeveen RM, Kroon J, Kraaijenhof JM, Waisai F, Timmerman N, Bom MJ, Hoefer IE, Knaapen P, et al. Targeted proteomics improves cardiovascular risk prediction in secondary prevention. *Eur Heart J*. 2022;43(16):1569–77.
33. Royer P, Björnson E, Adiels M, Josefson R, Hagberg E, Gummesson A, Bergström G. Large-scale plasma proteomics in the UK biobank modestly improves prediction of major cardiovascular events in a population without previous cardiovascular disease. *Eur J Prev Cardiol*. 2024;31(14):1681–9.
34. Xie R, Vlaski T, Sha S, Brenner H, Schöttker B. Sex-specific proteomic signatures improve cardiovascular risk prediction for the general population without cardiovascular disease or diabetes. *J Adv Res* 2025.
35. Nowak C, Carlsson AC, Östgren CJ, Nyström FH, Alam M, Feldreich T, Sundström J, Carrero JJ, Leppert J, Hedberg P, et al. Multiplex proteomics for prediction of major cardiovascular events in type 2 diabetes. *Diabetologia*. 2018;61(8):1748–57.
36. Ferreira JP, Sharma A, Mehta C, Bakris G, Rossignol P, White WB, Zannad F. Multi-proteomic approach to predict specific cardiovascular events in patients with diabetes and myocardial infarction: findings from the EXAMINE trial. *Clin Res Cardiol*. 2021;110(7):1006–19.
37. Li Y, Li D, Lin J, Zhou L, Yang W, Yin X, Xu C, Cao Z, Wang Y. Proteomic signatures of type 2 diabetes predict the incidence of coronary heart disease. *Cardiovasc Diabetol*. 2025;24(1):120.
38. Yu H, Zhang J, Qian F, Yao P, Xu K, Wu P, Li R, Qiu Z, Li R, Zhu K, et al. Large-Scale plasma proteomics improves prediction of peripheral artery disease in individuals with type 2 diabetes: a prospective cohort study. *Diabetes Care*. 2025;48(3):381–9.
39. Bayes-Genis A, Docherty KF, Petrie MC, Januzzi JL, Mueller C, Anderson L, Bozkurt B, Butler J, Chioncel O, Cleland JG. Practical algorithms for early diagnosis of heart failure and heart stress using NT-proBNP: a clinical consensus statement from the heart failure association of the ESC. *Eur J Heart Fail*. 2023;25(11):1891–8.
40. Gayen JR, Zhang K, Ramachandra Rao SP, Mahata M, Chen Y, Kim H-S, Naviaux RK, Sharma K, Mahata SK, O'Connor DT. Role of reactive oxygen species in hyperadrenergic hypertension: biochemical, physiological, and Pharmacological evidence from targeted ablation of the chromogranin A (Chga) gene. *Circ Cardiovasc Genet*. 2010;3(5):414–25.
41. Hou T, Jian C, Xu J, Huang AY, Xi J, Hu K, Wei L, Cheng H, Wang X. Identification of EFHD1 as a novel  $Ca^{2+}$  sensor for mitoflash activation. *Cell Calcium*. 2016;59(5):262–70.
42. Malachias MVB, Wijkman MO, Bertoluci MC. NT-proBNP as a predictor of death and cardiovascular events in patients with type 2 diabetes. *Diabetol Metab Syndr*. 2022;14(1):64.
43. Bibbins-Domingo K, Gupta R, Na B, Wu AH, Schiller NB, Whooley MA. N-terminal fragment of the prohormone brain-type natriuretic peptide (NT-proBNP), cardiovascular events, and mortality in patients with stable coronary heart disease. *JAMA*. 2007;297(2):169–76.
44. Yamamoto M, Hanatani S, Araki S, Izumiya Y, Yamada T, Nakanishi N, Ishida T, Yamamura S, Kimura Y, Arima Y, et al. HE4 predicts progressive fibrosis and cardiovascular events in patients with dilated cardiomyopathy. *J Am Heart Assoc*. 2021;10(15):e021069.
45. LeBleu VS, Teng Y, O'Connell JT, Charytan D, Müller GA, Müller CA, Sugimoto H, Kalluri R. Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis. *Nat Med*. 2013;19(2):227–31.
46. Jones PL, Crack J, Rabinovitch M. Regulation of tenascin-C, a vascular smooth muscle cell survival factor that interacts with the  $\alpha\beta3$  integrin to promote epidermal growth factor receptor phosphorylation and growth. *J Cell Biol*. 1997;139(1):279–93.
47. Li G, Jin R, Norris RA, Zhang L, Yu S, Wu F, Markwald RR, Nanda A, Conway SJ, Smyth SS. Periostin mediates vascular smooth muscle cell migration through the integrins  $\alpha\beta3$  and  $\alpha\beta5$  and focal adhesion kinase (FAK) pathway. *Atherosclerosis*. 2010;208(2):358–65.
48. Guo D-C, Papke CL, Tran-Fadulu V, Regalado ES, Avidan N, Johnson RJ, Kim DH, Pannu H, Willing MC, Sparks E. Mutations in smooth muscle  $\alpha$ -actin (ACTA2) cause coronary artery disease, stroke, and Moyamoya disease, along with thoracic aortic disease. *Am J Hum Genet*. 2009;84(5):617–27.
49. Awwad K, Hu J, Shi L, Mangels N, Abdel Malik R, Zippel N, Fisslthaler B, Eble JA, Pfeilschifter J, Popp R. Role of secreted modular calcium-binding protein 1 (SMOC1) in transforming growth factor  $\beta$  signalling and angiogenesis. *Cardiovasc Res*. 2015;106(2):284–94.
50. Dhulekar J, Simionescu A. Challenges in vascular tissue engineering for diabetic patients. *Acta Biomater*. 2018;70:25–34.
51. Herder C, Kannenberg JM, Carstensen-Kirberg M, Huth C, Meisinger C, Koenig W, Peters A, Rathmann W, Roden M, Thorand B. Serum levels of interleukin-22, cardiometabolic risk factors and incident type 2 diabetes: KORA F4/FF4 study. *Cardiovasc Diabetol*. 2017;16:1–8.
52. Gong X, Xu L, Cai P. Friend or foe of tripartite motif-containing protein 21 in cardiovascular disease: a review. *Int J Biol Macromol*. 2025;308(Pt 4):142682.
53. Sundkvist A, Myte R, Bodén S, Enroth S, Gyllenstein U, Harlid S, van Guelpen B. Targeted plasma proteomics identifies a novel, robust association between Cornulin and Swedish moist snuff. *Sci Rep*. 2018;8(1):2320.
54. Shan J, Bai X, Han L, Yuan Y, Yang J, Sun X. Association between atherosclerosis and gastric biomarkers concerning *Helicobacter pylori* infection in a Chinese healthy population. *Exp Gerontol*. 2018;112:97–102.
55. Whitcomb DC. Genetic risk factors for pancreatic disorders. *Gastroenterology*. 2013;144(6):1292–302.
56. Schuermans A, Pournamdari AB, Lee J, Bhukar R, Ganesh S, Darosa N, Small AM, Yu Z, Hornsby W, Koyama S, et al. Integrative proteomic analyses across common cardiac diseases yield mechanistic insights and enhanced prediction. *Nat Cardiovasc Res*. 2024;3(12):1516–30.
57. Takaya K, Kishi K. Regulation of ENPP5, a senescence-associated secretory phenotype factor, prevents skin aging. *Biogerontology*. 2024;25(3):529–42.
58. Zhang L, Song K, Zhou L, Xie Z, Zhou P, Zhao Y, Han Y, Xu X, Li P. Heparan sulfate D-glucosaminyl 3-O-sulfotransferase-3B1 (HS3ST3B1) promotes angiogenesis and proliferation by induction of VEGF in acute myeloid leukemia cells. *J Cell Biochem*. 2015;116(6):1101–12.
59. Yin X, Subramanian S, Hwang S-J, O'Donnell CJ, Fox CS, Courchesne P, Muntendam P, Gordon N, Adourian A, Juhasz P. Protein biomarkers of new-onset cardiovascular disease: prospective study from the systems approach to biomarker research in cardiovascular disease initiative. *Arterioscler Thromb Vasc Biol*. 2014;34(4):939–45.
60. Lygirou V, Latosinska A, Makridakis M, Mullen W, Delles C, Schanstra JP, Zoidakis J, Pieske B, Mischak H, Vlahou A. Plasma proteomic analysis reveals altered protein abundances in cardiovascular disease. *J Transl Med*. 2018;16(1):104.
61. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*. 2020;12(1):44.
62. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: What, why, how, when and where? *Clin Kidney J*. 2021;14(1):49–58.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.