TABLE OF CONTENTS

(**a**), not receiving lipid-lowering medication (**b**), receiving antihypertensive medication (**c**), and not receiving antihypertensive medication (**d**).

Supplementary Figure 15: Predictive performance of multiomics for cardiovascular diseases after excluding participants who experienced the event within the first two years of follow-up.

Supplementary Figure 16: Predictive performance of multiomics for cardiovascular diseases after considering competing risk of death (N = 24,287).

Supplementary Figure 17: Performance comparison of CardiOmicScore framework against other statistical machine learning models (N = 24,287).

Supplementary Figure 18: Feature importance from MetNet refitted after excluding after excluding participants on lipid-lowering medication.

Supplementary Figure 19: Predictive performance of traditional lipid-related biomarkers for cardiovascular diseases (N = 24,287).

Supplementary Figure 20: Predictive performance of traditional lipid-related biomarkers for cardiovascular diseases among participants not receiving lipid-lowering medication (N = 19,895).

Supplementary Figure 21: The architecture of MetNet (**a**) and ProNet (**b**).

References

Supplementary Methods

Sample-level and task-level weighting

In our multi-task learning framework, the loss function incorporates both sample-level weighting and task-level weighting to account for variations in data distribution and task importance.

At the sample level, we applied the positive class weight used in PyTorch's BCEWithLogitsLoss:

$$\text{pos}_\text{w}\text{eight}_t = \frac{N_t - P_t}{P_t} \quad (1)$$

where N_t and P_t denote the number of samples and positive samples in task t ($t = 1, \ldots, T$), respectively. The weighted binary cross-entropy loss for task t is:

$$\mathcal{L}_t^{\text{sample}} = -\frac{1}{N_t}\sum_{t=1}^{N_t}\left[\text{pos}_\text{w}\text{eight}_t \cdot y_i^{(t)} \cdot \log\sigma\left(x_i^{(t)}\right) + \left(1 - y_i^{(t)}\right) \cdot \log\left(1 - \sigma\left(x_i^{(t)}\right)\right)\right] \quad (2)$$

where $x_i^{(t)}$ is the model's predicted logit of sample i for task t; $y_i^{(t)} \in \{0,1\}$ is the binary label of sample i in task t; $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function. This increases the penalty for misclassifying positive samples, encouraging the model to focus on rare events.

At the task level, we assigned each task t a weight:

$$\omega_t = \frac{1/P_t}{\sum_{k=1}^{T}(1/P_k)} \quad (3)$$

where T is the number of tasks. Tasks with fewer positive samples thus contribute proportionally more to the total loss, preventing them from being dominated by more prevalent diseases. For the ProNet, the task-level weights are 0.06 for coronary artery disease (CAD), 0.19 for stroke, 0.14 for heart failure (HF), 0.07 for atrial fibrillation (AF), 0.38 for peripheral artery disease (PAD), and 0.15 for venous thromboembolism (VTE). The MetNet

model uses a comparable weighting scheme with values of 0.06 for CAD, 0.20 for stroke, 0.15 for HF, 0.07 for AF, 0.36 for PAD, and 0.16 for VTE.

The total weighted loss is:

$$\mathcal{L}_{\text{total}} = \sum_{t=1}^{T} \omega_t \cdot \mathcal{L}_t^{\text{sample}} \quad (4)$$

Calculation of ASCVD and SCORE2

We calculated the ASCVD and SCORE2 using the published algorithms[1–3]. For the ASCVD, the predictors included age, black race, current smoking, total cholesterol, high-density lipoprotein (HDL) cholesterol, systolic blood pressure (SBP), antihypertensive therapy, and diabetes[1,2]. The sex-specific ASCVD was calculated using equations (5)– (7).

$$\begin{aligned}\text{Logit}_{\text{Male}} = {} & -11.679980 + 0.064200 \times \text{Age} + 0.482835 \times \text{Race} + 0.038950 \times \text{SBP} \\ & -0.000061 \times \text{SBP}^2 + 2.055533 \times \text{HTNTherapy} + 0.842209 \times \text{Diabetes} \\ & +0.895589 \times \text{Smoking} + 0.193307 \times \left(\frac{\text{TChol}}{\text{HDL}}\right) - 0.014207 \times \text{SBP} \times \text{HTNTherapy} \\ & +0.011609 \times \text{SBP} \times \text{Race} - 0.119460 \times \text{HTNTherapy} \times \text{Race} + 0.000025 \times \text{Age} \times \text{SBP} \\ & -0.077214 \times \text{Race} \times \text{Diabetes} - 0.226771 \times \text{Race} \times \text{Smoking} - \\ & -0.117749 \times \text{Race} \times \left(\frac{\text{TChol}}{\text{HDL}}\right) + 0.004190 \times \text{Race} \times \text{HTNTherapy} \times \text{SBP} \\ & -0.000199 \times \text{Race} \times \text{Age} \times \text{SBP}\end{aligned} \quad (5)$$

$$\begin{aligned}\text{Logit}_{\text{Female}} = {} & -12.823110 + 0.106501 \times \text{Age} + 0.432440 \times \text{Race} + 0.017666 \times \text{SBP} \\ & +0.000056 \times \text{SBP}^2 + 0.731678 \times \text{HTNTherapy} + 0.943970 \times \text{Diabetes} \\ & +1.009790 \times \text{Smoking} + 0.151318 \times \left(\frac{\text{TChol}}{\text{HDL}}\right) - 0.003647 \times \text{SBP} \times \text{HTNTherapy} \\ & +0.006208 \times \text{SBP} \times \text{Race} + 0.152968 \times \text{HTNTherapy} \times \text{Race} - 0.000153 \times \text{Age} \times \text{SBP} \\ & +0.115232 \times \text{Race} \times \text{Diabetes} - 0.092231 \times \text{Race} \times \text{Smoking} - \\ & +0.070498 \times \text{Race} \times \left(\frac{\text{TChol}}{\text{HDL}}\right) - 0.000173 \times \text{Race} \times \text{HTNTherapy} \times \text{SBP} \\ & -0.000094 \times \text{Race} \times \text{Age} \times \text{SBP}\end{aligned} \quad (6)$$

$$\text{ASCVD} = \frac{100}{1 + e^{-\text{Logit}}} \quad (7)$$

where TChol, HDL, and HTNTherapy indicate total cholesterol, HDL cholesterol, and antihypertensive therapy, respectively. The sex-specific SCORE2 was calculated based on age, current smoking, total cholesterol, HDL cholesterol, and SBP using equations (8)–(11)[3].

$$\begin{aligned}\text{Linear predictor}_{\text{Male}} = {} & 0.3742 \times \text{Age} + 0.6012 \times \text{Smoking} + 0.2777 \times \text{SBP} + 0.6457 \times \text{Diabetes} \\ & + 0.1458 \times \text{TChol} - 0.2698 \times \text{HDL} - 0.0755 \times \text{Age} \times \text{Smoking} \\ & - 0.0255 \times \text{Age} \times \text{SBP} - 0.0281 \times \text{Age} \times \text{TChol} + 0.0426 \times \text{Age} \times \text{HDL} \\ & - 0.0983 \times \text{Age} \times \text{Diabetes}\end{aligned} \quad (8)$$

$$\begin{aligned}\text{Linear predictor}_{\text{Female}} = {} & 0.4648 \times \text{Age} + 0.7744 \times \text{Smoking} + 0.3131 \times \text{SBP} + 0.8096 \times \text{Diabetes} \\ & + 0.1002 \times \text{TChol} - 0.2606 \times \text{HDL} - 0.1088 \times \text{Age} \times \text{Smoking} \\ & - 0.0277 \times \text{Age} \times \text{SBP} - 0.0226 \times \text{Age} \times \text{TChol} + 0.0613 \times \text{Age} \times \text{HDL} \\ & - 0.1272 \times \text{Age} \times \text{Diabetes}\end{aligned} \quad (9)$$

$$\text{SCORE2}_{Male} = 1 - 0.9605 \times e^{-\text{Linear predictor}_{\text{Male}}} \quad (10)$$

$$\text{SCORE2}_{Female} = 1 - 0.97765 \times e^{-\text{Linear predictor}_{\text{Female}}} \quad (11)$$

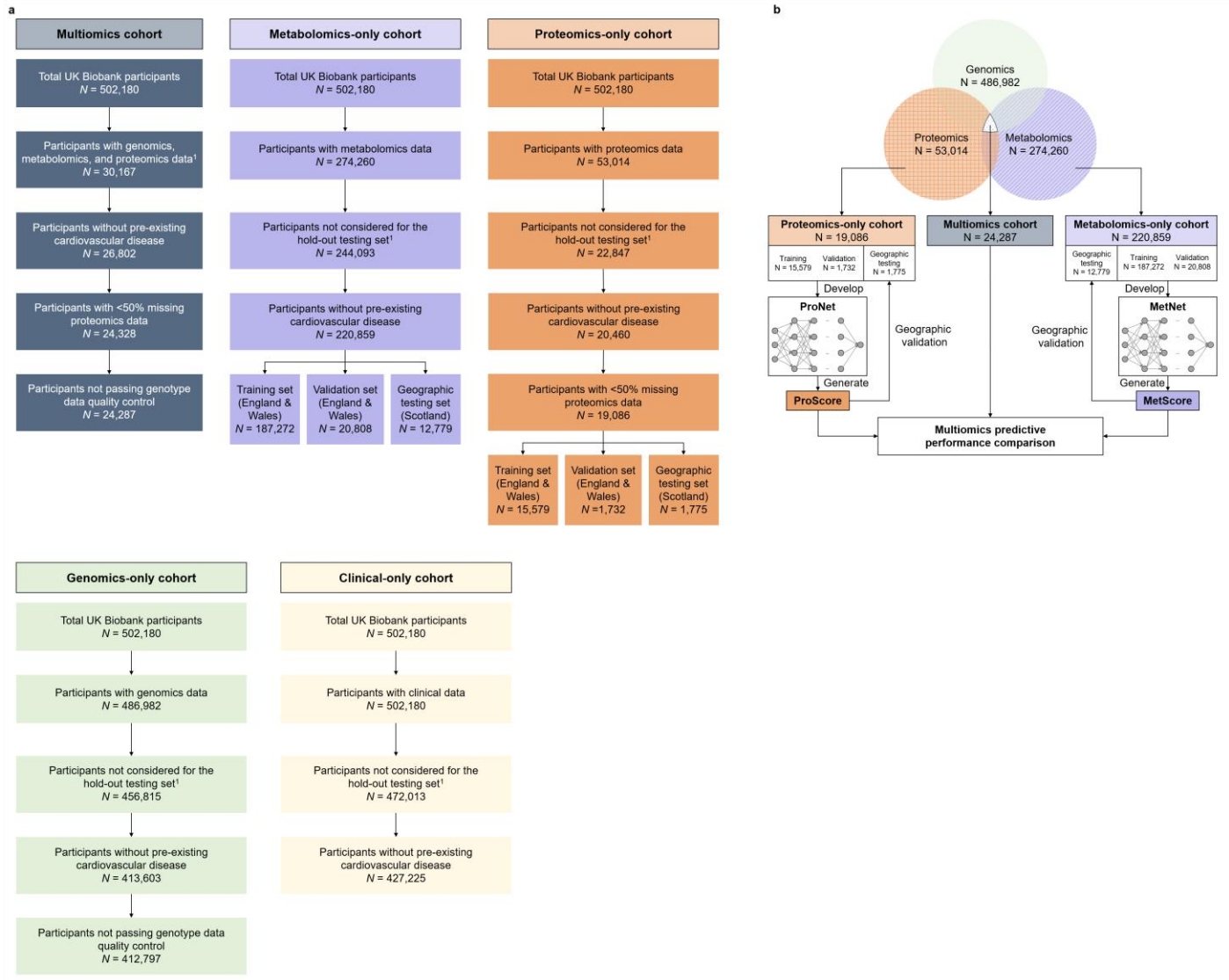
where TChol and HDL indicate total cholesterol and HDL cholesterol, respectively. Prior to the calculation, age, total cholesterol, HDL cholesterol, and SBP were standardized using the following transformations:

$$\begin{aligned}\text{Age} &= (\text{Age} - 60)/5 \\ SBP &= (SBP - 120)/20 \\ TChol &= (\text{Totalcholesterol} - 6)/1 \\ HDL &= (HDL\text{cholesterol} - 1.3)/0.5\end{aligned} \quad (12)$$

Development of machine learning models

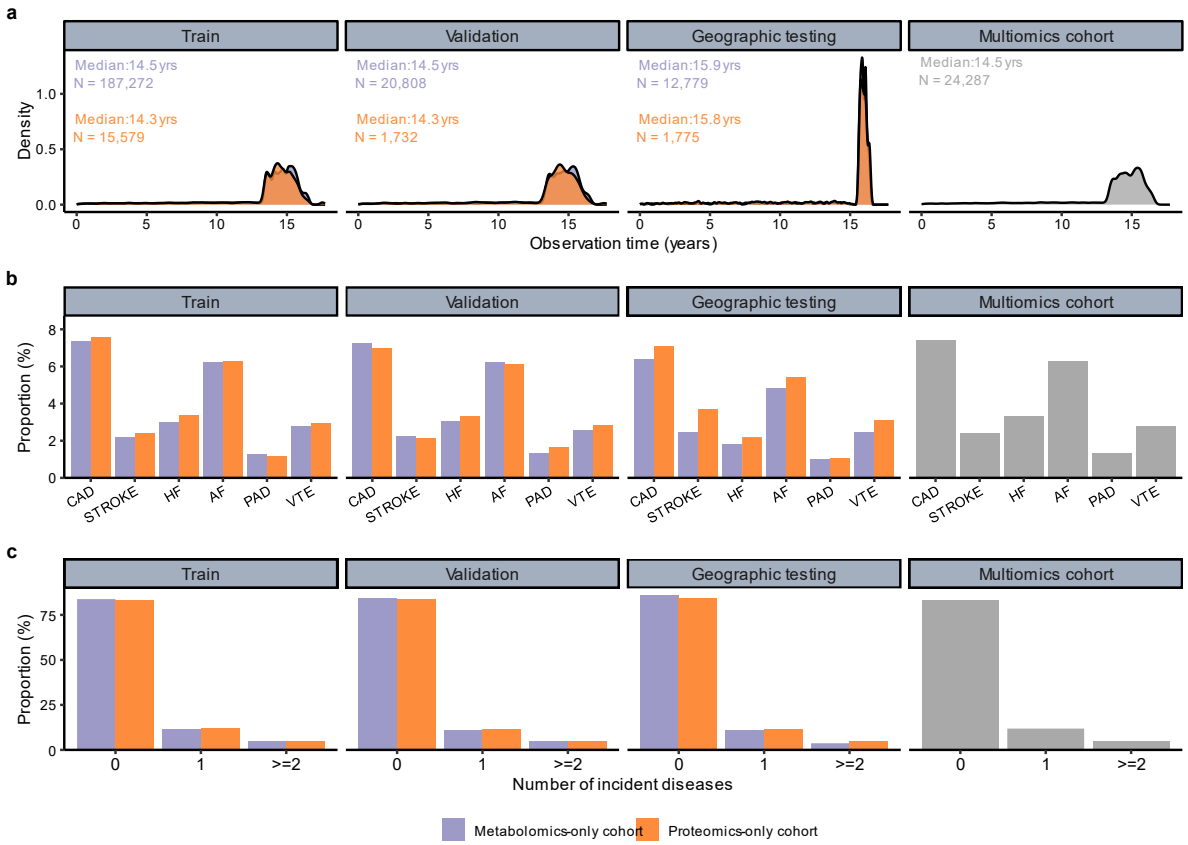
We developed alternative MetScore and ProScore using four widely-used machine learning algorithms: Extreme Gradient Boosting (XGBoost), LightGBM, random forest, and logistic regression. We trained a separate, independent model for each of the six cardiovascular diseases (CVDs) using all individual metabolomic or proteomic biomarkers as input features. Data splitting and preprocessing followed the same procedures as in the main analysis. For

each of the 24 models (4 algorithms × 6 CVDs), we conducted a separate hyperparameter optimization using a random search of 50 combinations. The optimal hyperparameter set was selected based on the combination that achieved the highest area under the receiver operating characteristic curve (AUC) on the validation set. Key hyperparameters were tuned for each model (Supplementary Data 15). For XGBoost and LightGBM, this included the number of estimators, learning rate, maximum tree depth, and subsampling ratios, while for random forest, we tuned the number of estimators, maximum tree depth, maximum features per split, and minimum samples for splits and leaves. For logistic regression, we tuned the inverse of regularization strength. To address class imbalance, we employed a sample weighting strategy. For XGBoost and LightGBM, we used the scale_pos_weight parameter; specifically, this was calculated as the ratio of the number of negative cases to positive cases in the training data. For random forest and logistic regression, the class_weight='balanced' setting was used. After identifying the optimal hyperparameters, a final model was retrained on the complete training dataset. To prevent overfitting in XGBoost and LightGBM, early stopping was applied during the tuning phase, halting training if the validation AUC did not improve for 20 consecutive rounds. The resulting risk predictions from these models constituted the alternative MetScores and ProScores used for comparison.

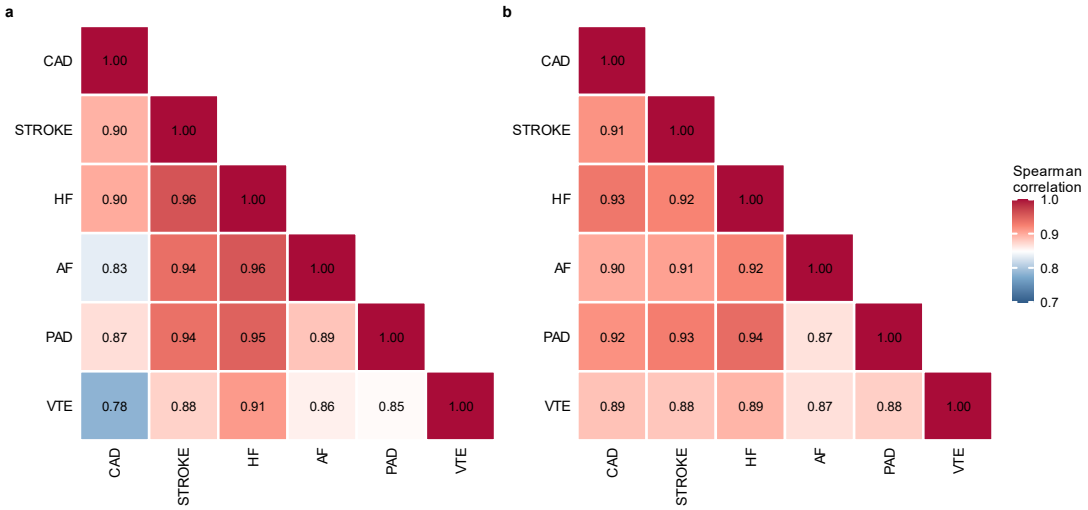


Supplementary Figure 1: Two-stage individual-level data partitioning strategy. **a.** Sample exclusion flowchart. Genotype data quality control excluded participants with discordant self-reported and genetic sex, sex chromosome aneuploidy, and individuals having more than ten third-degree genetic relatives. **b.** Study framework.

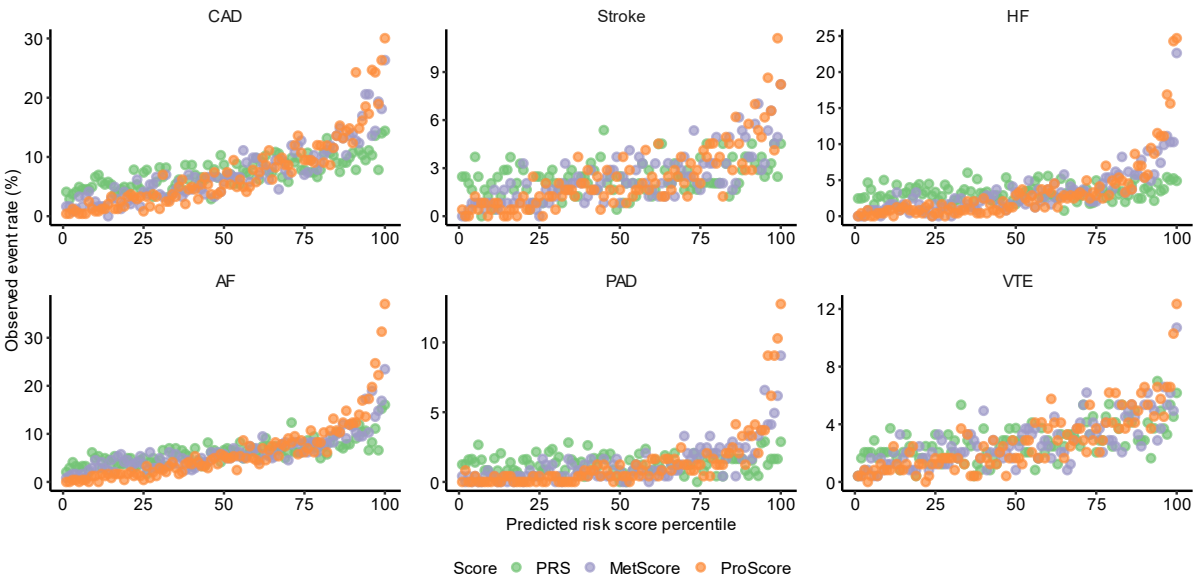
[1]"Participants not considered for the Multiomics cohort" in the Metabolomics-only, Proteomics-only, Genomics-only, and Clinical-only cohorts refer to those with available metabolomics, proteomics, genomics, or clinical data, respectively, after excluding the 30,167 individuals with all three omics data, who were eligible for inclusion in the Multiomics cohort.



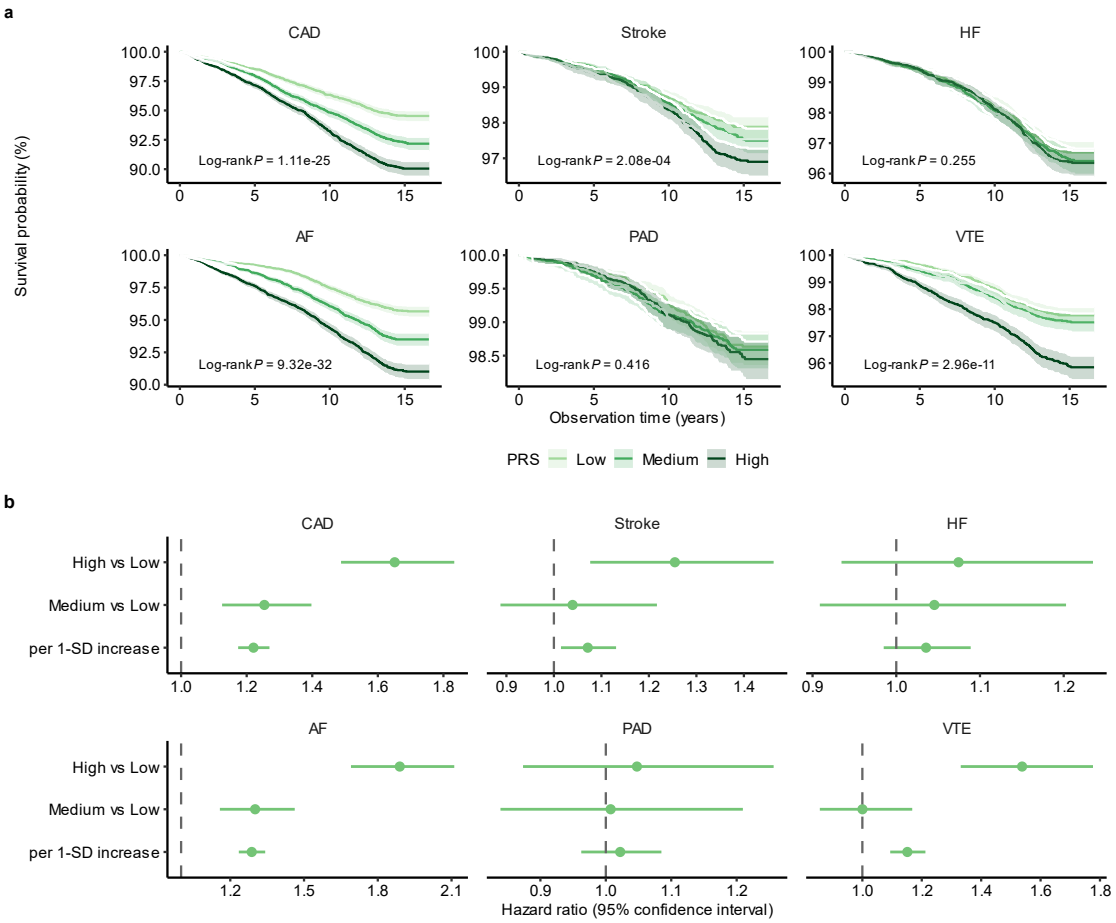
Supplementary Figure 2: Follow-up duration and distribution of incident cardiovascular disease (CVD) cases across datasets. **a.** Distribution of follow-up duration. Here, we defined the follow-up duration as the period from the date of baseline assessment to the date of any incident cardiovascular diseases, death, loss to follow-up, or end of available registry follow-up (November 30 2023 for England & Wales and December 31 2023 for Scotland), whichever came first. Text annotations provide the median observation time and the total number of participants for each dataset. **b.** Proportion of participants with incident cases of each CVD. **c.** Proportion of participants grouped by the number of incident CVD.



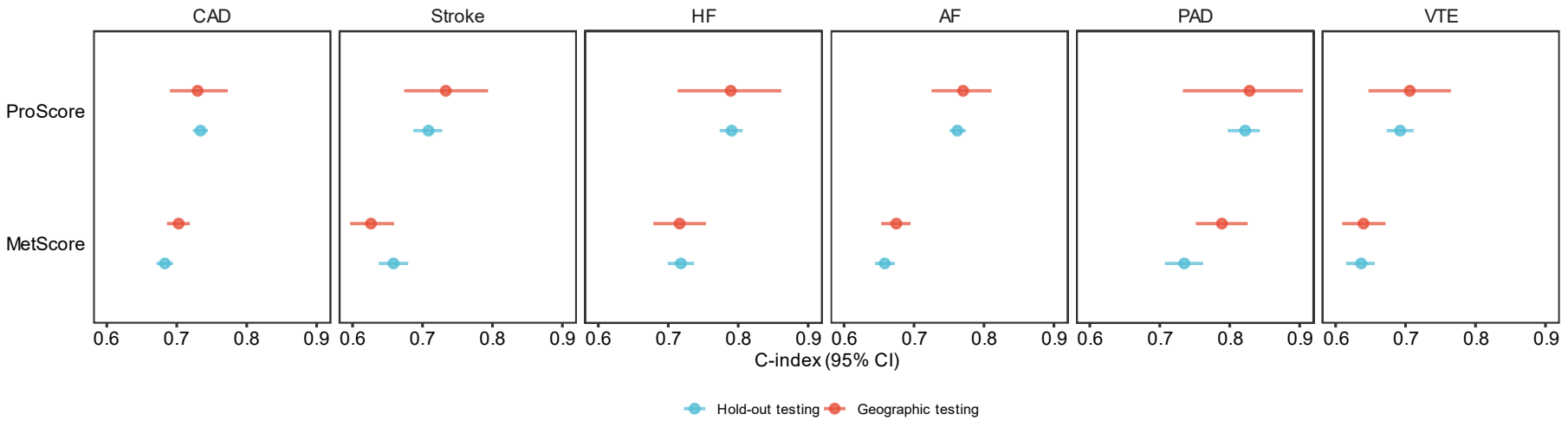
Supplementary Figure 3: Spearman correlation between MetScore (**a**) and ProScore (**b**) across cardiovascular diseases (N = 24,287).



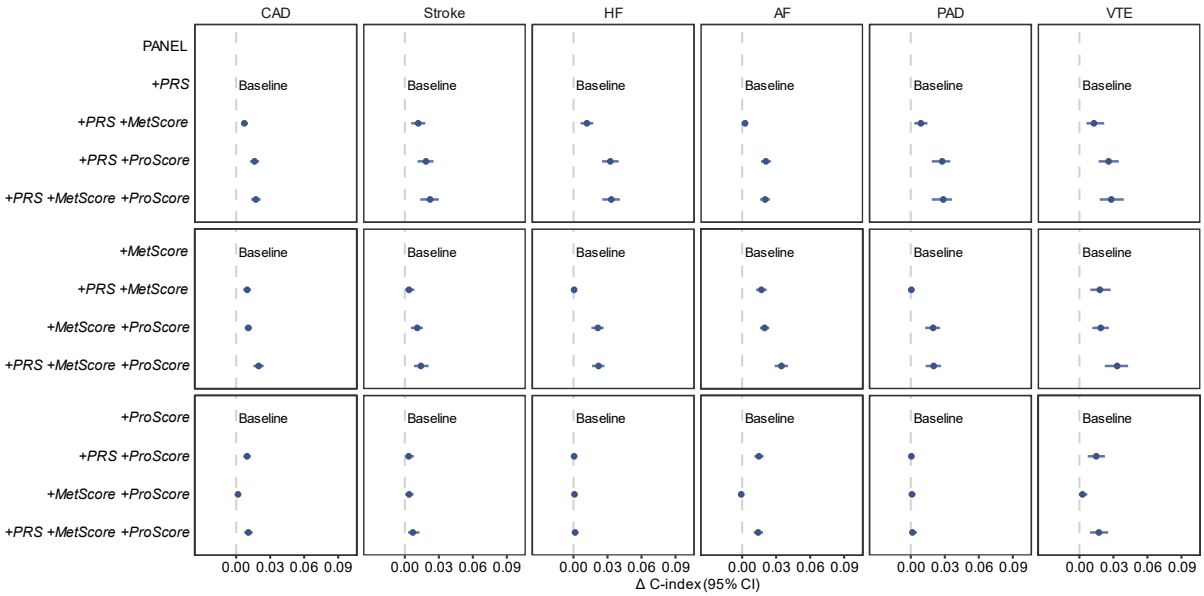
Supplementary Figure 4: Observed event rates for incident cardiovascular diseases across the percentiles of polygenic risk score (PRS), MetScore, and ProScore (*N* = 24,287).



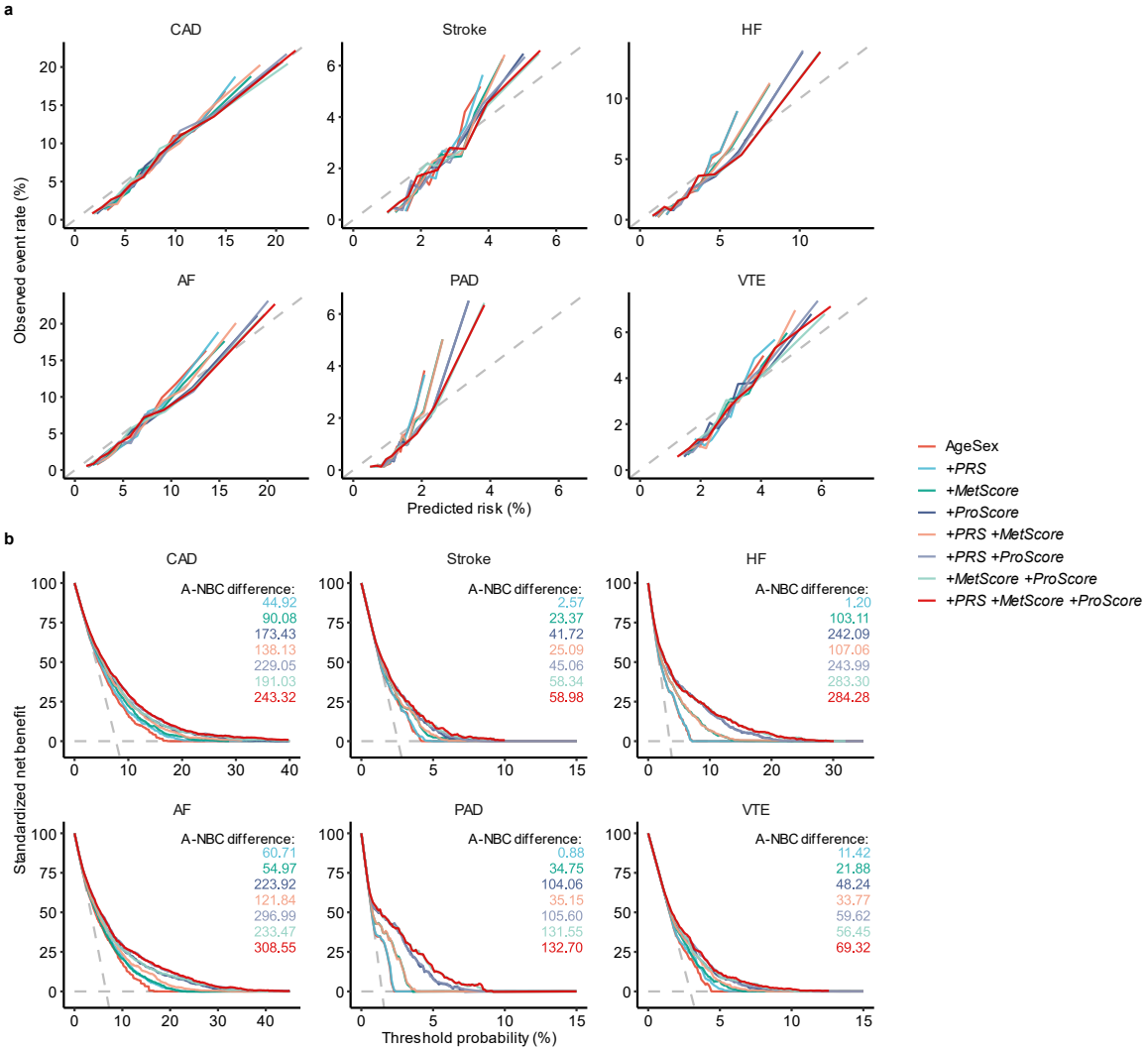
Supplementary Figure 5: Polygenic risk score (PRS) stratify the risk of cardiovascular disease (CVD) (*N* = 24,287). **a.** Survival probabilities over time for six cardiovascular diseases, stratified by the tertiles of PRS. Solid lines represent survival probabilities (measure of centre), estimated using the Kaplan-Meier method, and shaded areas represent 95% exponential Greenwood confidence intervals (error bands). **b.** Associations of CVD risk with PRS. Each score was analyzed both as a categorical variable (tertile groups: low, medium, high; low as reference) and as a continuous variable (per 1-standard deviation [SD] increase after standardization). Data are presented as hazard ratios (measure of centre) estimated by Cox proportional hazard models with error bars representing 95% confidence intervals. All models were adjusted for demographic information, healthy lifestyles, family disease history, disease and medication history, physical measurements, and blood count.



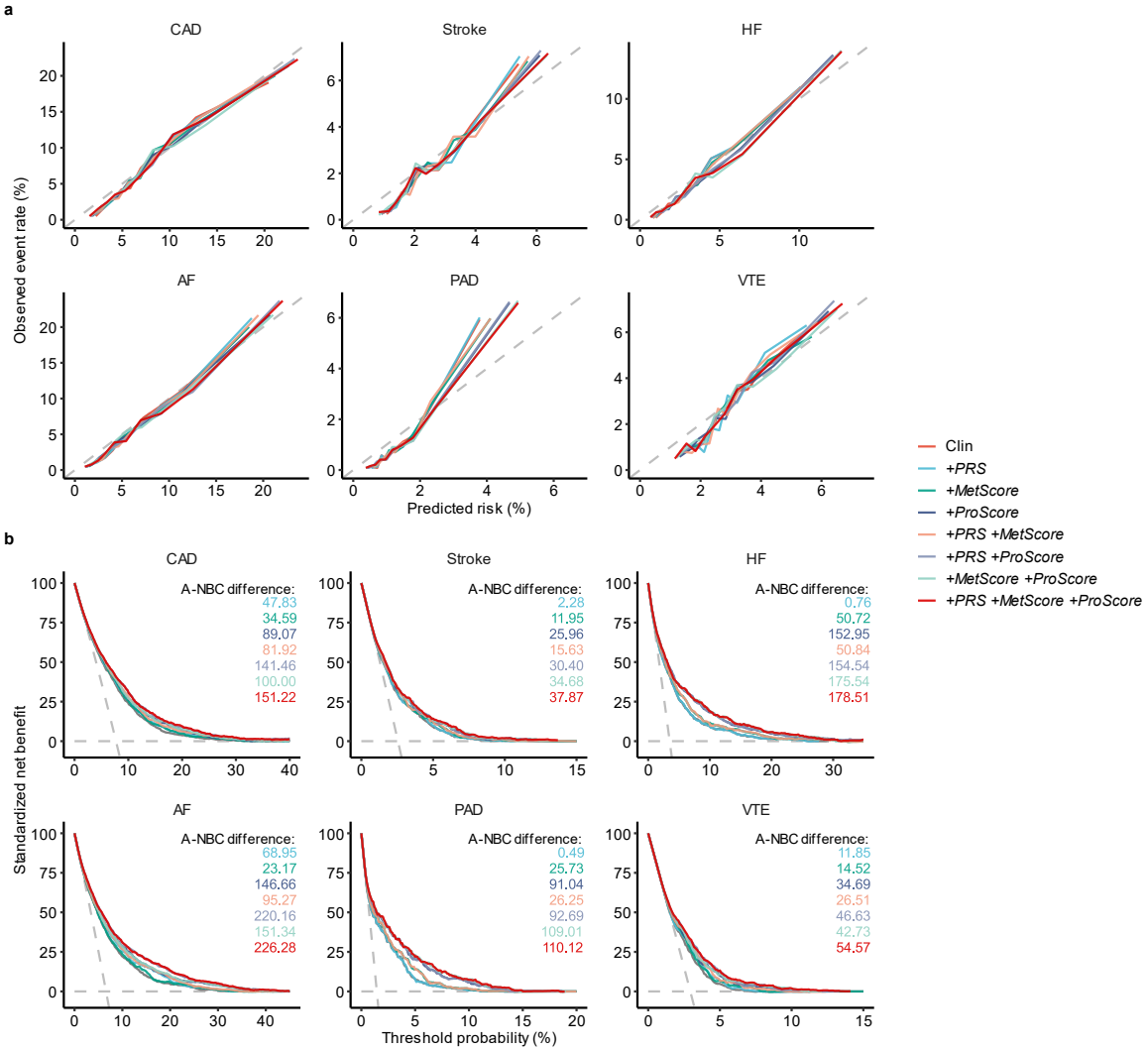
Supplementary Figure 6: Discriminative performance of MetScore and ProScore. Discriminative performance was evaluated separately for MetScore and ProScore in both the geographic testing and the Multiomics cohort. The Multiomics cohort included 24,287 participants, while the geographic testing sets included 12,779 for MetScore and 1,775 for ProScore. Performance was assessed using Harrell's C-index. Data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



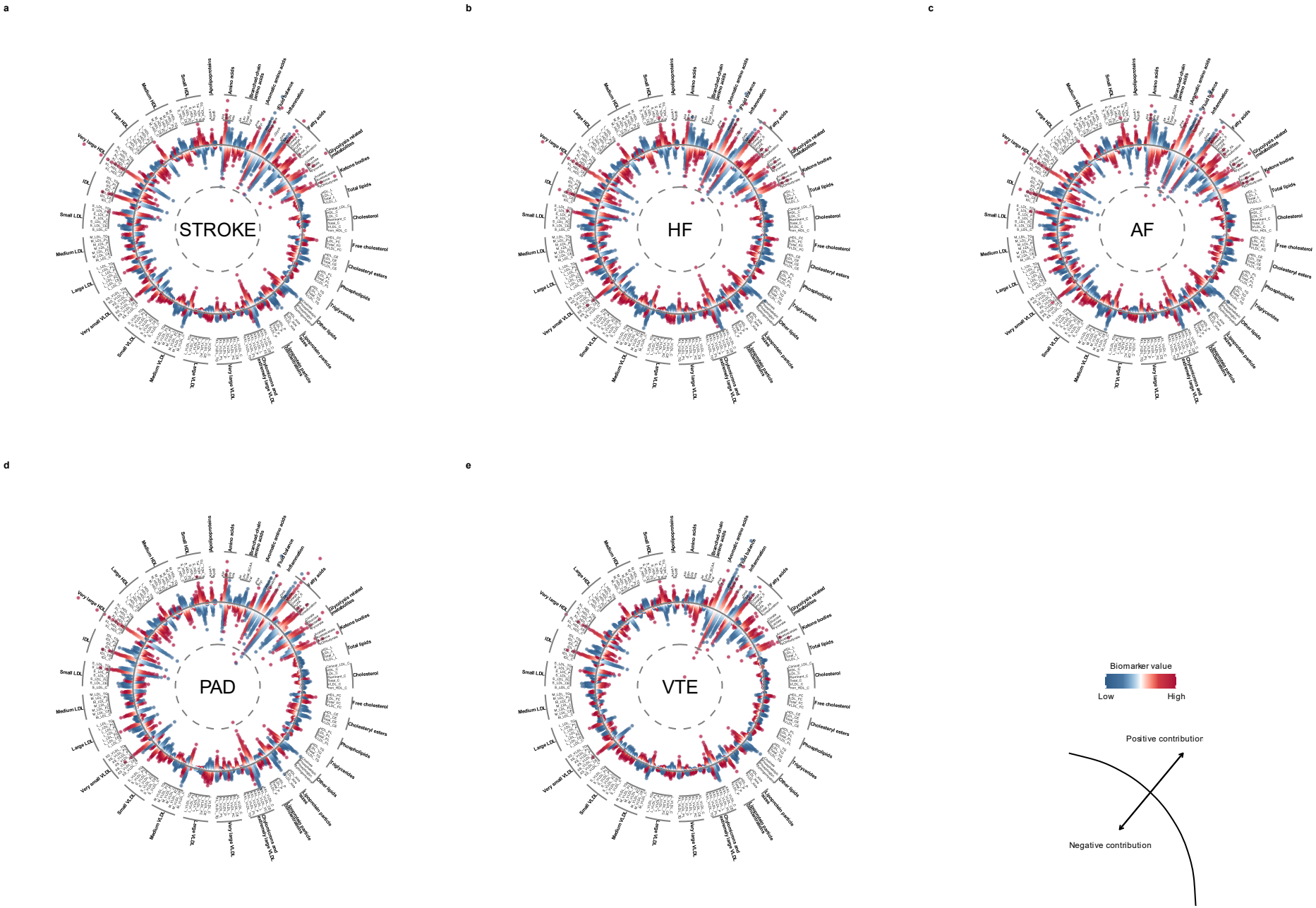
Supplementary Figure 7: Differences in discriminative performance for cardiovascular diseases by adding additional omics features (N = 24,287). Performance was assessed using Harrell's C-index. Data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



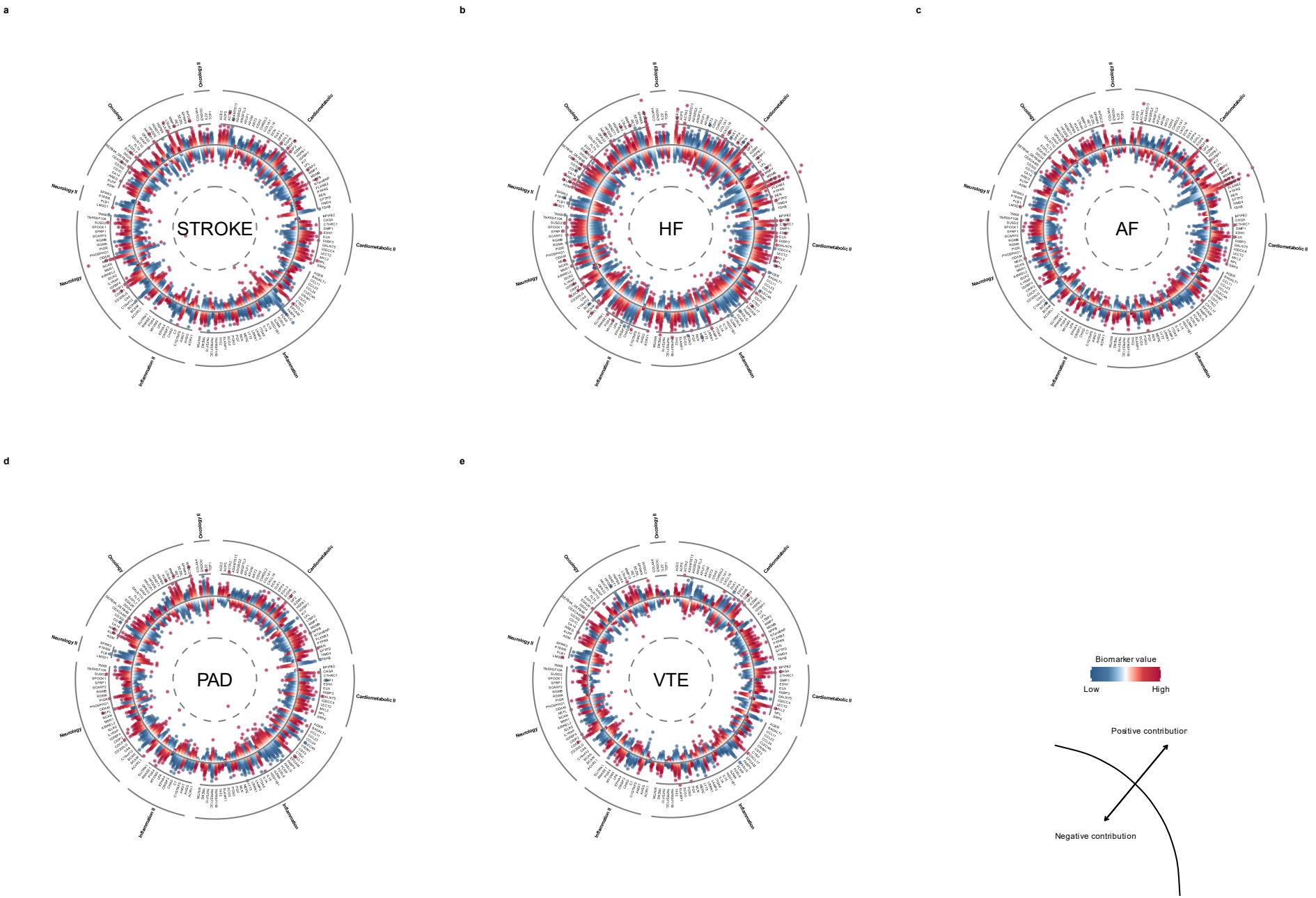
Supplementary Figure 8: Model (AgeSex) calibration and net benefit curves for cardiovascular diseases. **a.** Calibration curves for Cox proportional hazard models based on the clinical predictor set (AgeSex) and its combination with multiomics information. **b.** Net benefit curves for Cox proportional hazard models based on the clinical predictor set (AgeSex) and its combination with multiomics information. Horizontal solid gray lines indicate “treat none” and vertical solid gray lines indicate “treat all”. Differences in the area under the net benefit curve (A-NBC) represent improvements in net benefit compared with the AgeSex-based models.



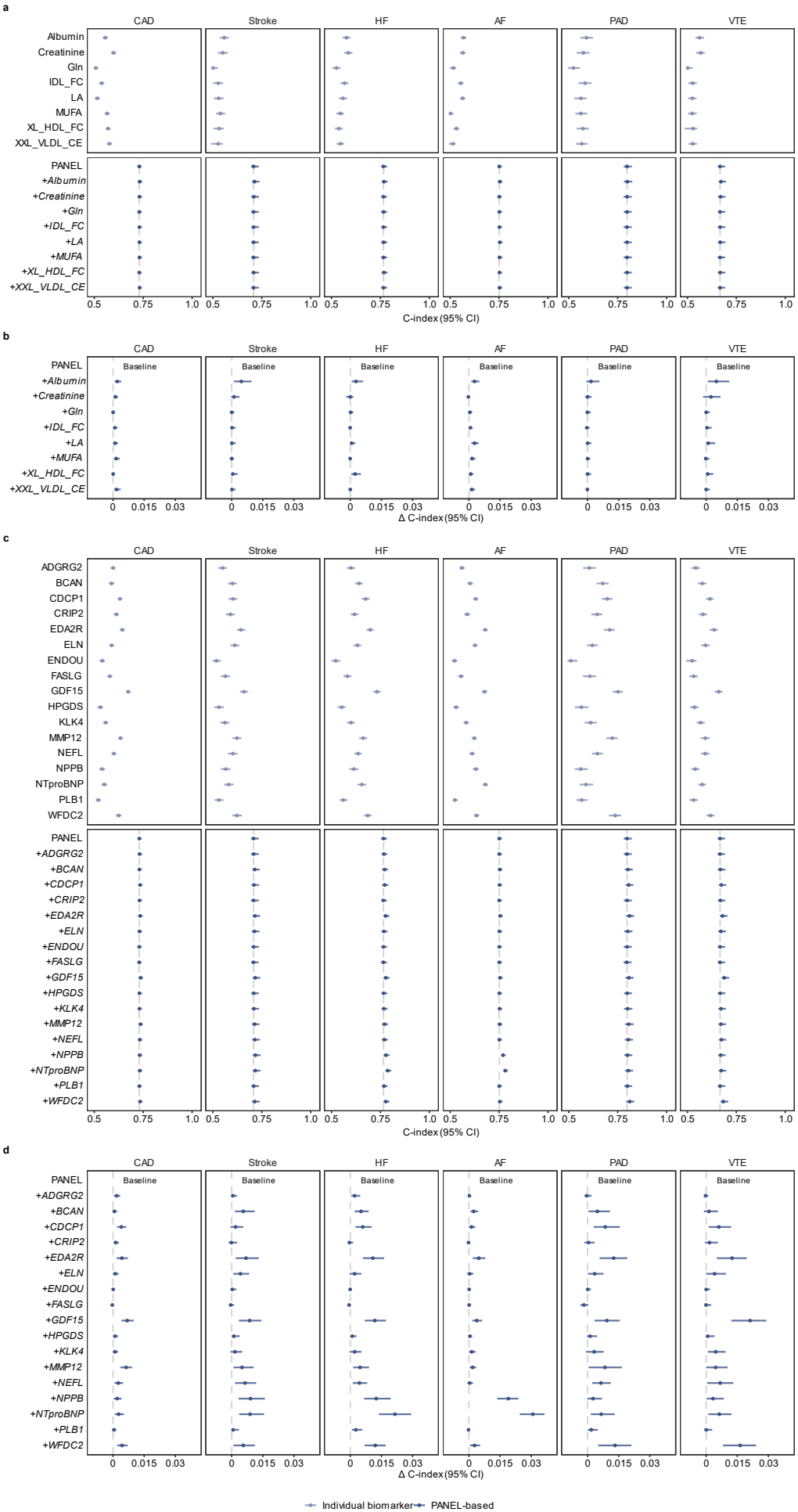
Supplementary Figure 9: Model (Clin) calibration and net benefit curves for cardiovascular diseases. **a.** Calibration curves for Cox proportional hazard models based on the clinical predictor set (Clin) and its combination with multiomics information. **b.** Net benefit curves for Cox proportional hazard models based on the clinical predictor set (Clin) and its combination with multiomics information. Horizontal solid gray lines indicate "treat none" and vertical solid gray lines indicate "treat all". Differences in the area under the net benefit curve (A-NBC) represent improvements in net benefit compared with the Clin-based models.



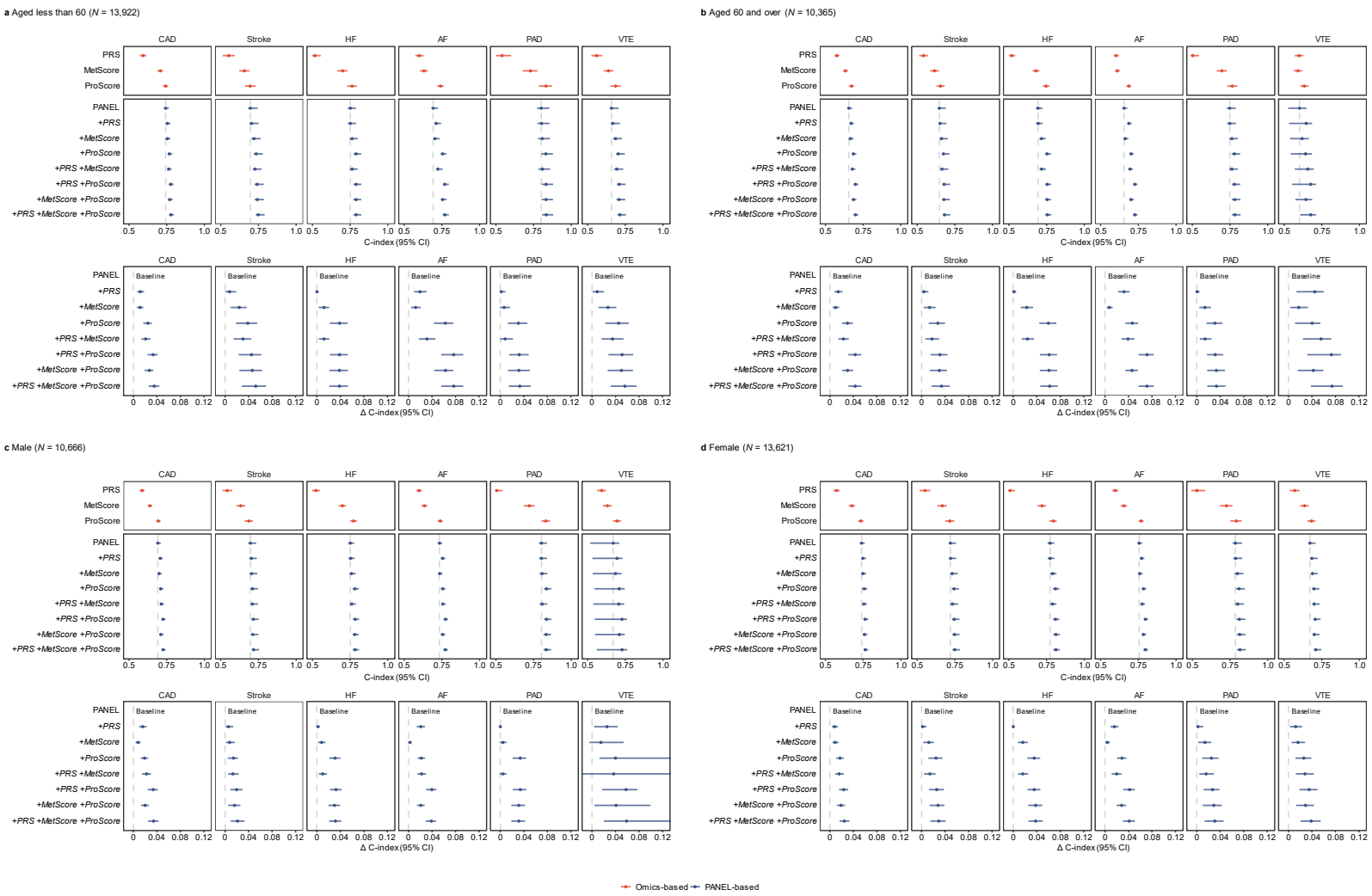
Supplementary Figure 10: SHAP beeswarm plots for the MetScore of stroke (**a**), HF (**b**), AF (**c**), PAD (**d**), and VTE (**e**).



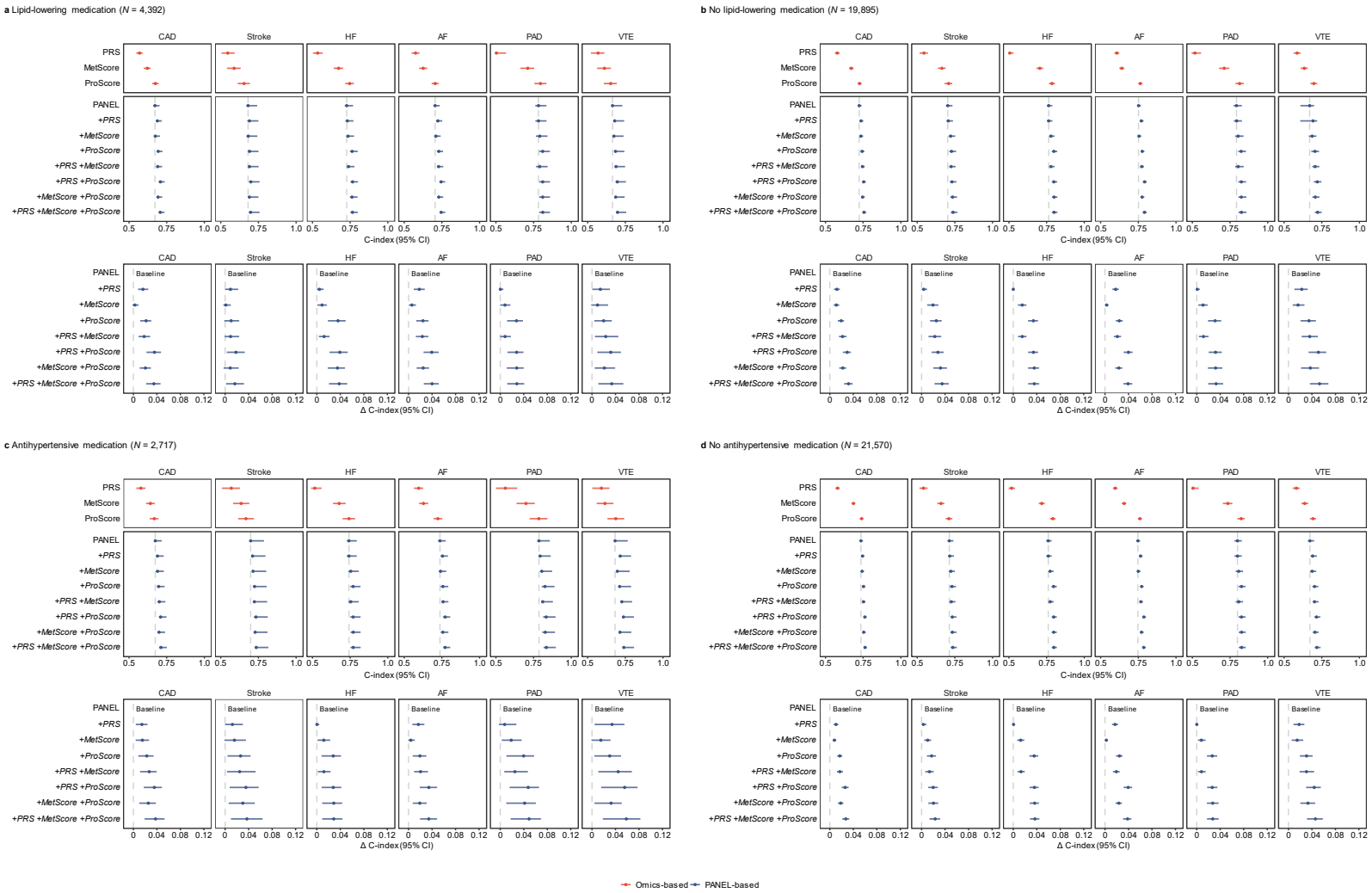
Supplementary Figure 11: SHAP beeswarm plots for the ProScore of stroke (**a**), HF (**b**), AF (**c**), PAD (**d**), and VTE (**e**).


a
CAD
Stroke
HF
AF
PAD
VTE
Albumin
Creatinine
Gln
IDL_FC
LA
MUFA
XL_HDL_FC
XXL_VLDL_CE
PANEL
+Albumin
+Creatinine
+Gln
+IDL_FC
+LA
+MUFA
+XL_HDL_FC
+XXL_VLDL_CE
0.5
0.75
1.0
C-index (95% CI)
b
Baseline
0
0.015
0.03
Δ C-index (95% CI)
c
ADGRG2
BCAN
CDCP1
CRIP2
EDA2R
ELN
ENDOU
FASLG
GDF15
HPGDS
KLK4
MMP12
NEFL
NPPB
NTproBNP
PLB1
WFDC2
+ADGRG2
+BCAN
+CDCP1
+CRIP2
+EDA2R
+ELN
+ENDOU
+FASLG
+GDF15
+HPGDS
+KLK4
+MMP12
+NEFL
+NPPB
+NTproBNP
+PLB1
+WFDC2
d
Individual biomarker
PANEL-based

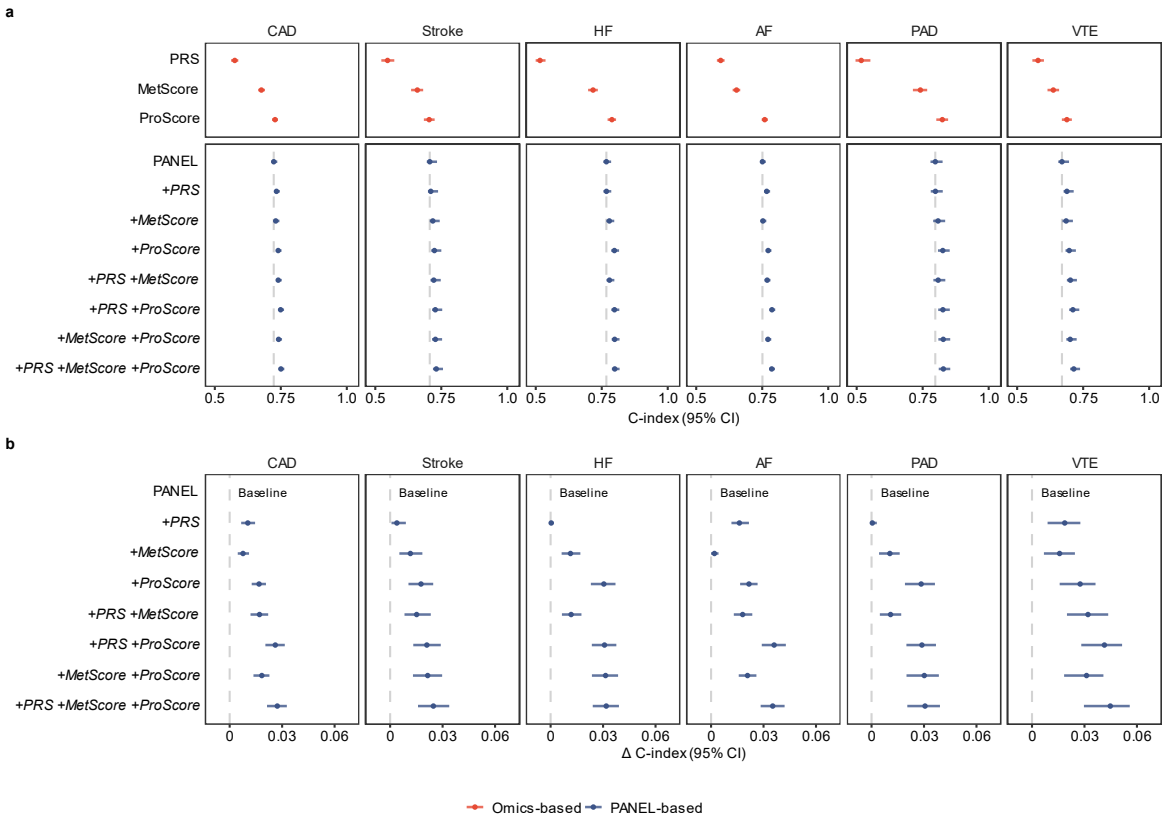
Supplementary Figure 12: Predictive performance of individual biomarkers for cardiovascular diseases (*N* = 24,287). **a, c.** Discriminative performance for key metabolites (**a**, *n* = 8) and proteins (**c**, *n* = 17). Vertical dashed lines indicate the performance of the PANEL set. Discriminative performance was evaluated by the Harrell's C-index. **b, d.** Differences in discriminative performance (ΔC-index) by adding each individual metabolite (**b**) or protein (**d**) to the PANEL set. In all panels, data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



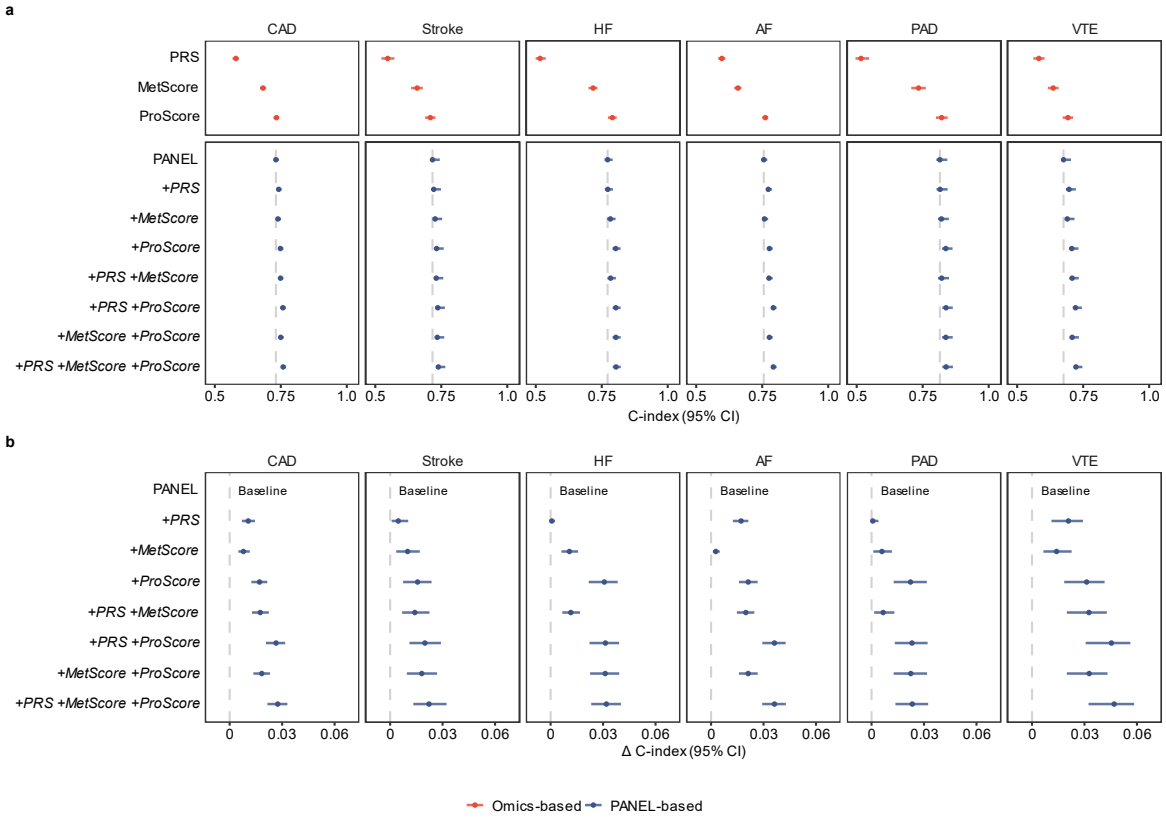
Supplementary Figure 13: Predictive performance of multiomics for cardiovascular diseases across subgroups, including participants aged under 60 (**a**), those aged 60 and over (**b**), males (**c**), and females (**d**). **a.** Discriminative performance of models trained on various predictor sets. Vertical dashed lines indicate the performance of the PANEL set. Discriminative performance was evaluated by the Harrell's C-index. **b.** Differences in discriminative performance (ΔC-index) by adding omics information to the PANEL set. In all panels, data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



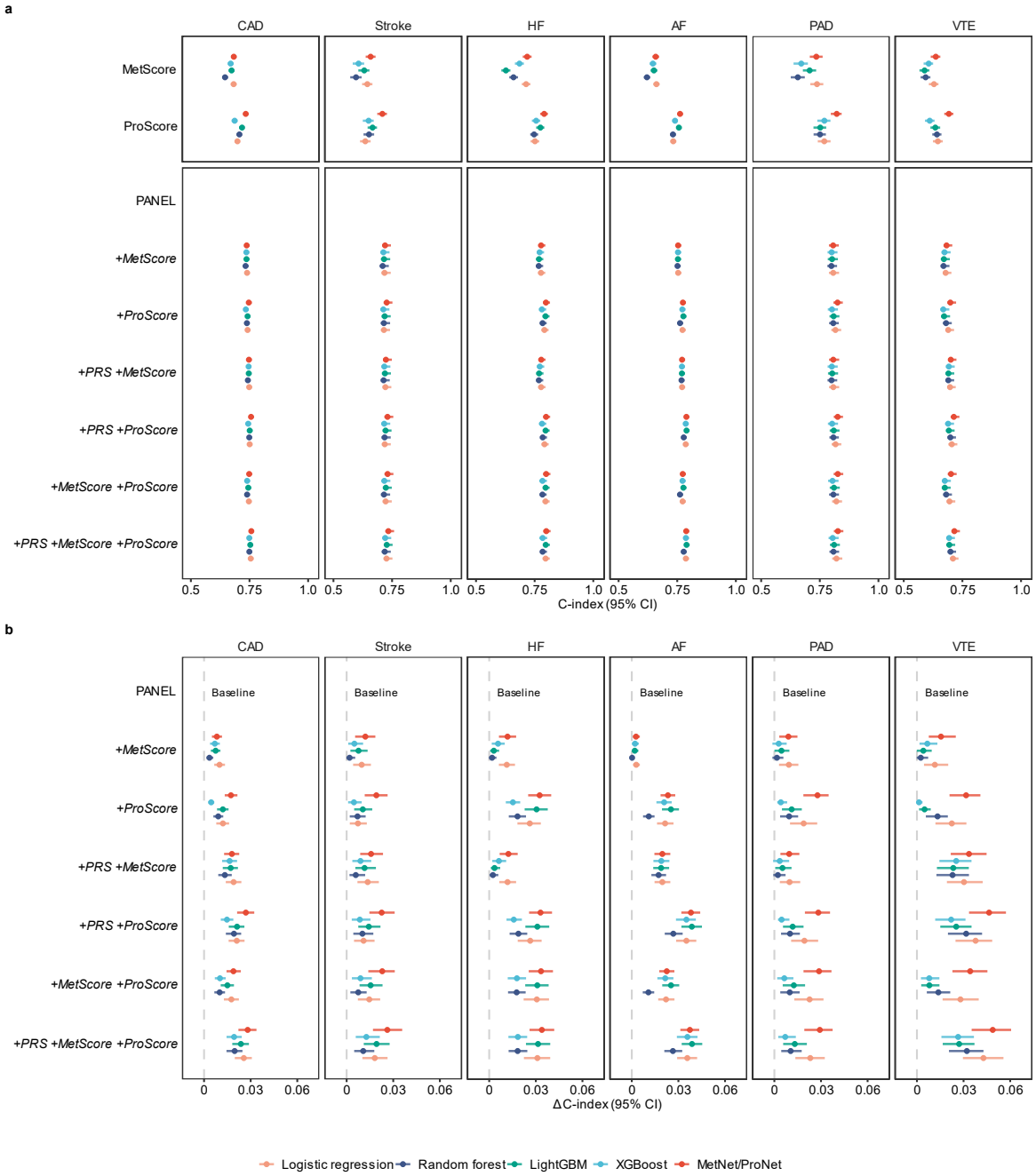
Supplementary Figure 14: Predictive performance of multiomics for cardiovascular diseases across subgroups, including participants receiving lipid-lowering medication (**a**), not receiving lipid-lowering medication (**b**), receiving antihypertensive medication (**c**), and not receiving antihypertensive medication (**d**). **a.** Discriminative performance of models trained on various predictor sets. Vertical dashed lines indicate the performance of the PANEL set. Discriminative performance was evaluated by the Harrell's C-index. **b.** Differences in discriminative performance (ΔC-index) by adding omics information to the PANEL set. In all panels, data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



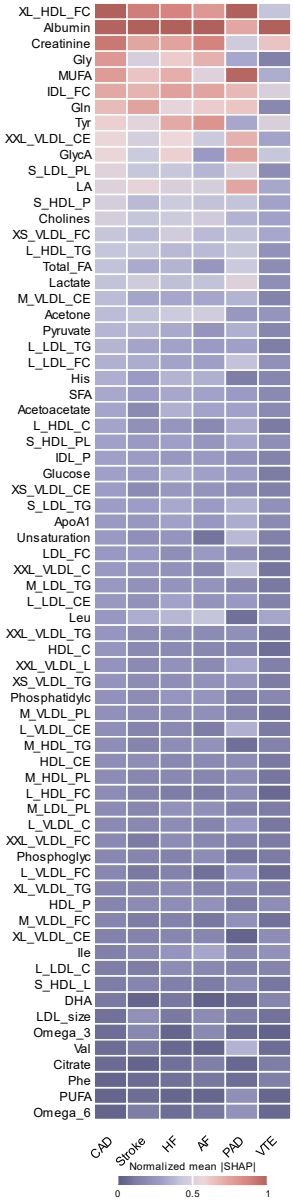
Supplementary Figure 15: Predictive performance of multiomics for cardiovascular diseases after excluding participants who experienced the event within the first two years of follow-up. **a.** Discriminative performance trained on various predictor sets. Vertical dashed lines indicate the performance of the PANEL set. Discriminative performance was evaluated by the Harrell's C-index. **b.** Differences in discriminative performance (ΔC-index) by adding omics information to the PANEL set. The number of participants remaining for each analysis was: 24,104 for CAD, 24,251 for stroke, 24,243 for HF, 24,163 for AF, 24,269 for PAD, and 24,241 for VTE. In both panels, data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



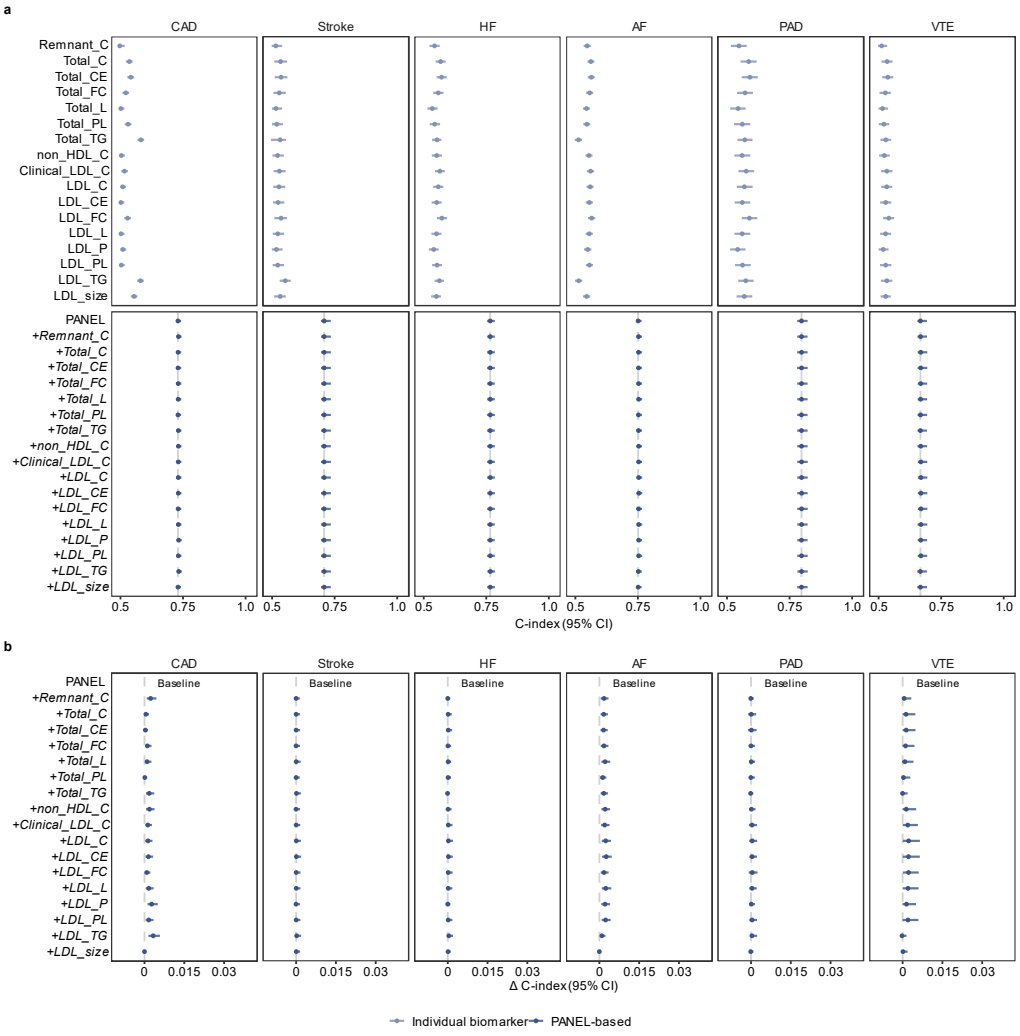
Supplementary Figure 16: Predictive performance of multiomics for cardiovascular diseases after considering competing risk of death (*N* = 24,287). **a.** Discriminative performance trained on various predictor sets. Vertical dashed lines indicate the performance of the PANEL set. Discriminative performance was evaluated by the Harrell's C-index. **b.** Differences in discriminative performance (ΔC-index) by adding omics information to the PANEL set. All-cause death was treated as a competing event. The number of primary (CVD-specific) and competing (death) events were as follows: CAD (1,788 primary, 1,592 competing), stroke (584 primary, 1,791 competing), HF (798 primary, 1,694 competing), AF (1,524 primary, 1,627 competing), PAD (324 primary, 1,879 competing), and VTE (679 primary, 1,727 competing). In both panels, data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



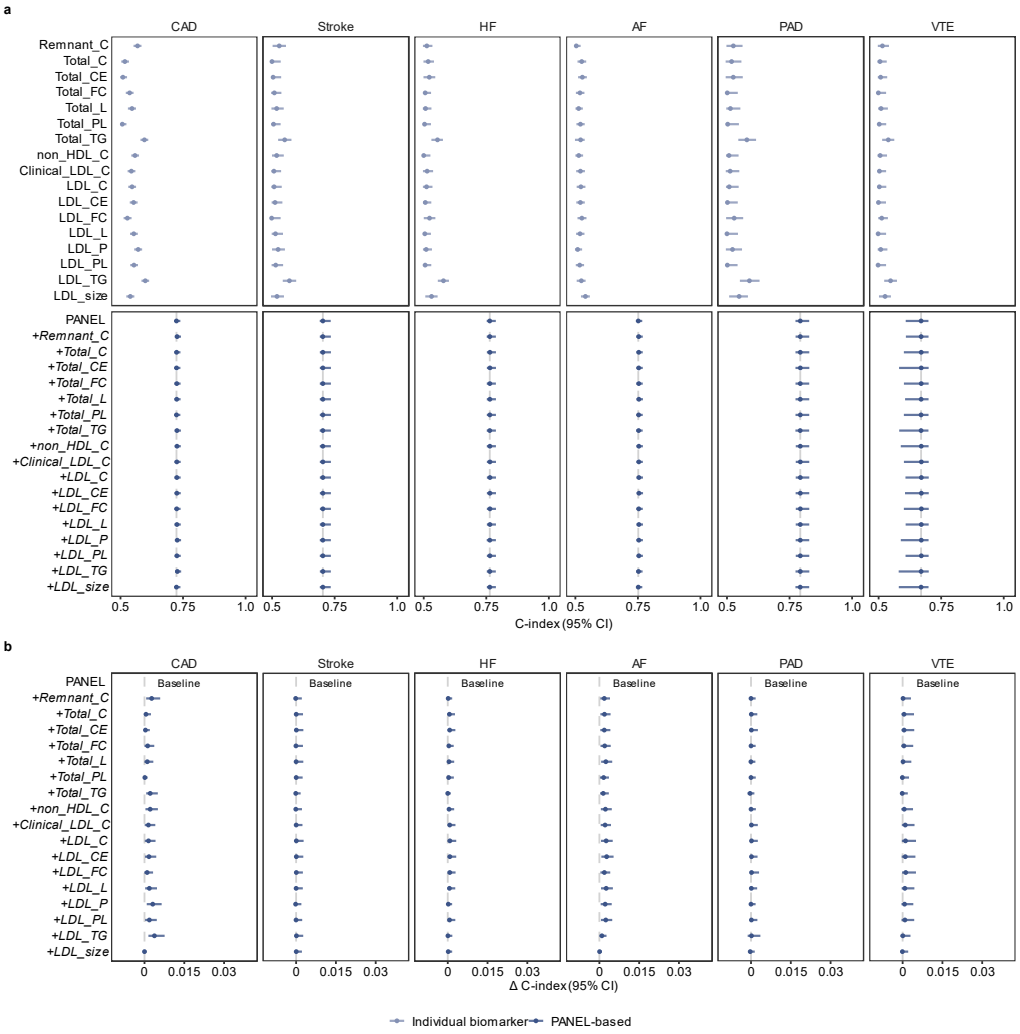
Supplementary Figure 17: Performance comparison of CardiOmicScore framework against other statistical machine learning models (N = 24,287). **a.** Discriminative performance trained on various predictor sets. MetScore and ProScore were estimated using different algorithms, including logistic regression, random forest, LightGBM, XGBoost, and MetNet/ProNet. Discriminative performance was evaluated by the Harrell's C-index. **b.** Differences in discriminative performance (ΔC-index) by adding omics information to clinical predictor sets. In both panels, data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



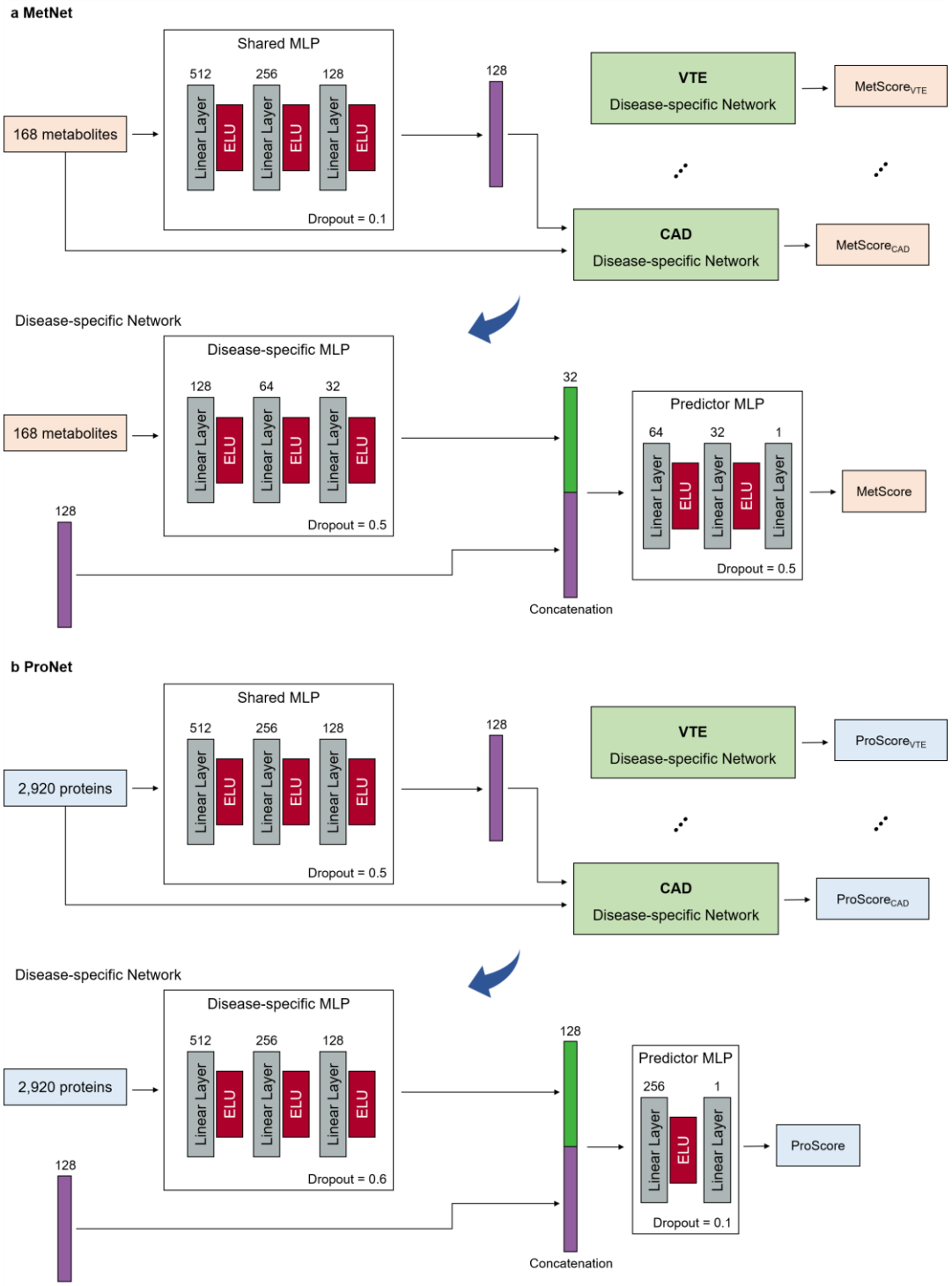
Supplementary Figure 18: Feature importance from MetNet refitted after excluding after excluding participants on lipid-lowering medication. The MetNet was refitted using an identical architecture on a subgroup of the Metabolomics-only cohort that excluded individuals on baseline lipid-lowering medication (N = 185,086). Feature importance was measured by calculating mean absolute SHAP values for the participants in the Multiomics cohort who were also not receiving lipid-lowering medication (N = 19,895). The color represents normalized mean absolute SHAP values, ranging from blue (low) to red (high). For each disease, the 50 most important metabolites were identified. The union of these features across all diseases yielded a total of 70 metabolites.



Supplementary Figure 19: Predictive performance of traditional lipid-related biomarkers for cardiovascular diseases (*N* = 24,287). **a.** Discriminative performance for 17 lipid-related biomarkers. Vertical dashed lines indicate the performance of the PANEL set. Discriminative performance was evaluated by the Harrell's C-index. **b.** Differences in discriminative performance (ΔC-index) by adding each individual biomarker to the PANEL set. In both panels, data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



Supplementary Figure 20: Predictive performance of traditional lipid-related biomarkers for cardiovascular diseases among participants not receiving lipid-lowering medication (*N* = 19,895). **a.** Discriminative performance for 17 lipid-related biomarkers. Vertical dashed lines indicate the performance of the PANEL set. Discriminative performance was evaluated by the Harrell's C-index. **b.** Differences in discriminative performance (ΔC-index) by adding each individual biomarker to the PANEL set. In both panels, data are presented as point estimates (measure of centre) with error bars representing 95% confidence intervals estimated from 1,000 bootstrap resamples.



Supplementary Figure 21: The architecture of MetNet (**a**) and ProNet (**b**). MLP = Multilayer perceptron, CAD = Coronary artery disease, VTE = Venous thromboembolism.

References

1. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, *et al*. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. *Circulation* 2014;6;24 S49–73 doi: https://doi.org/10.1161/01.cir.0000437741.48606.98

2. Yadlowsky S, Hayward RA, Sussman JB, McClelland RL, Min YI, Basu S. Clinical Implications of Revised Pooled Cohort Equations for Estimating Atherosclerotic Cardiovascular Disease Risk. *Ann Intern Med* 2018;7;3 20–9 doi: https://doi.org/10.7326/M17-3011

3. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J* 2021;7;1 2439–54 doi: https://doi.org/10.1093/eurheartj/ehab309