Cardiovascular Diabetology

RESEARCH

Open Access

Distinct immune-metabolic phenotypes underlie poor coronary collateral circulation

Zi-Tong Guo[1], Hong-Mei Lai[2], Run-Xuan Hu[1], Ya-Jing Qiu[1], Jing Tao[2], Xiao-Lin Yu[2] and Yi-Ning Yang[1,2,3*]

Abstract

Background Coronary collateral circulation (CCC) significantly impacts myocardial perfusion and clinical outcomes in coronary artery disease patients, yet the underlying molecular heterogeneity remains inadequately characterized.

Objective To identify distinct molecular phenotypes in patients with poor CCC, validate these phenotypes using clinical parameters, and evaluate their prognostic implications.

Methods This study enrolled 149 patients (80 with good CCC and 69 with poor CCC) for high-throughput proteomic profiling. Unsupervised consensus clustering identified molecular subtypes within poor CCC patients, followed by differential expression analysis and KEGG pathway enrichment. Boruta feature selection was implemented, and multiple machine learning algorithms were tested on clinical data, with XGBoost optimization (accuracy 80.0%, F1-score 80.31%) and SHAP value interpretation. External validation was performed using the MIMIC database. Kaplan–Meier analysis and Cox regression models assessed major adverse cardiovascular events (MACE).

Results Two distinct phenotypes emerged among poor CCC patients: Cluster 1 (n = 39, Complement-Driven Vascular Remodeling [CDVR]) and Cluster 2 (n = 30, Immuno-Thrombotic Myocardial Dysfunction [ITMD]). An XGBoost model incorporating fasting glucose, eosinophil percentage, and HbA1c achieved excellent discrimination (AUC > 0.91). External validation confirmed the phenotype-specific clinical patterns. Notably, Cluster 2 demonstrated significantly higher MACE incidence compared to Cluster 1 (Log-rank $p < 0.05$), with KEGG analysis revealing significant upregulation of platelet activation, diabetic cardiomyopathy, and metabolic pathways in the ITMD phenotype.

Conclusion Poor CCC encompasses distinct immune-metabolic phenotypes that can be accurately classified using integrated proteomic-clinical modeling. This classification enables more precise risk stratification and may guide personalized therapeutic strategies for coronary artery disease patients with inadequate collateralization.

Keywords Coronary collateral circulation, Proteomics, Molecular phenotypes, Machine learning, Immune-metabolic profiles, Prognosis

*Correspondence:
Yi-Ning Yang
yangyn5126@163.com
[1]First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang, China
[2]People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi 830001, Xinjiang, China
[3]Xinjiang Key Laboratory of Cardiovascular Homeostasis and Regeneration Research, Urumqi, Xinjiang, China

Introduction

Coronary chronic total occlusion (CTO) is among the most challenging manifestations of advanced coronary artery disease, defined as complete luminal blockage of a coronary segment for at least three months, most commonly as a consequence of progressive atherosclerotic plaque evolution [1]. In this setting, the myocardium distal to the occlusion becomes reliant on alternative perfusion routes, most notably, the coronary collateral circulation (CCC), which represents a natural bypass system [2]. Development of a well-functioning CCC can preserve myocardial viability, reduce infarct size, and improve long-term survival as demonstrated by studies examining the impact of collaterals on myocardial perfusion [3, 4]. However, despite similar degrees of epicardial stenosis, a substantial proportion of CTO patients exhibit poor collateralization and experience worse clinical outcomes, including higher rates of major adverse cardiovascular events (MACE) [5].

The formation of CCC Is a complex, multifactorial adaptive process Integrating hemodynamic, cellular, and molecular signals. Multiple mechanisms contribute to the pathophysiology of coronary collateral circulation, with metabolic and inflammatory factors playing crucial roles [6, 7]. Diabetes mellitus is particularly associated with poor collateral vessel formation [8], with impaired arteriogenesis resulting from the interplay of inflammation and endothelial dysfunction [9]. Hyperglycemia can induce capillary rarefaction [10], while oxidative stress and advanced glycation end products collectively hamper effective collateral development [11, 12]. Comparative studies consistently demonstrate that diabetic patients with CTO have less developed CCC than their non-diabetic counterparts, independent of coronary disease severity [13]. Inflammatory processes further modulate this relationship, as inflammatory markers like the neutrophil–lymphocyte ratio correlate with collateral development status [14, 15]. These observations highlight the integral role of immune-metabolic interactions in determining CCC adequacy.

Despite advances in understanding the clinical and physiological determinants of CCC, the molecular heterogeneity underlying poor collateralization remains poorly characterized. Current approaches to CCC assessment, while clinically valuable, provide limited insight into the biological mechanisms that differentiate patients with similar clinical presentations but divergent collateral development [16–18]. This knowledge gap hampers the development of targeted interventions to enhance collateral growth in high-risk populations. Furthermore, the molecular subtypes within the poor CCC population remain undefined, potentially obscuring distinct pathophysiological mechanisms that could serve as targets for personalized therapeutic approaches.

Against this background, we applied quantitative plasma proteomics and comprehensive clinical profiling to 149 CTO patients to delineate biologically distinct subgroups among those with poor CCC. Using unsupervised consensus clustering, we identified two reproducible and mechanistically plausible phenotypes: Cluster 1, Complement-Driven Vascular Remodeling (CDVR), defined by hyperglycemia-driven chronic inflammation, persistent complement activation, and neutrophilia; and Cluster 2, Immuno-Thrombotic Myocardial Dysfunction (ITMD), characterized by platelet receptor signaling, increased eosinophil activity, and chemokine-mediated leukocyte recruitment. From these insights, we developed a parsimonious diagnostic model capable of differentiating controls from both phenotypes with high accuracy and validated it externally. This classification highlights distinct immunologic and metabolic axes that may serve as therapeutic targets, offering a potential roadmap for phenotype-guided intervention and personalized vascular medicine in CTO patients with compromised collateral circulation.

Method

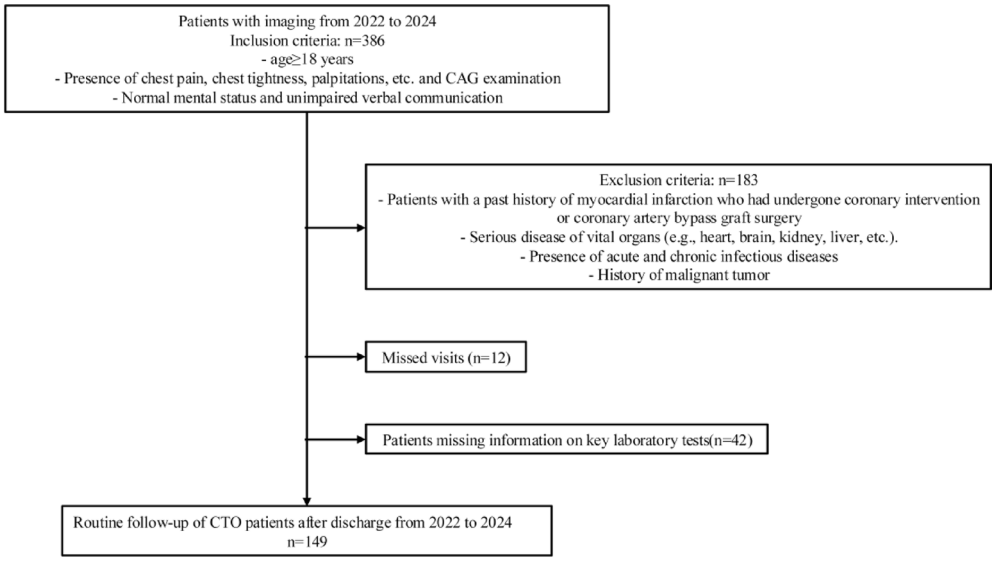
Study population

This study enrolled 149 patients with chronic total occlusion (CTO) of the coronary arteries who underwent coronary angiography at People's Hospital of Xinjiang Uygur Autonomous Region between January 2022 and June 2024. Coronary collateral circulation (CCC) was evaluated using the Rentrop collateral grading system (grades 0–3) [19]. Patients with Rentrop grades 0–1 were defined as having poor CCC, while those with grades 2–3 were considered to have good CCC, in accordance with previous studies. Of all enrolled patients, 69 had poor CCC (Rentrop grades 0–1) and underwent proteomic analysis, while 80 patients with good CCC (Rentrop grades 2–3) served as controls. Inclusion and exclusion criteria are described in Table 1. Written informed consent was obtained from all participants. The study protocol conformed to the principles of the Declaration of Helsinki and was approved by the Institutional Review Board of People's Hospital of Xinjiang Uygur Autonomous Region (approval number: KY20220311108).

Sample preparation and peptide digestion

100 μL of each sample was used to enrich low-abundance proteins using magnetic beads (Promega, Madison, WI, USA). 100 μL of plasma/serum was added to 4 μL of magnetic beads (washed twice with wash buffer, reagents from Sigma-Aldrich, St. Louis, MO, USA), mixed, and incubated for 2 h. The beads were then placed on a magnetic rack, the plasma/serum removed, and then washed with wash reagent, with each wash cycle repeated five times with inversion for 5 min. All samples

Table 1 Inclusion and exclusion flow chart



CAG: Coronary Angiography
CTO: Chronic Total Occlusion

were digested with trypsin (Thermo Scientific, Waltham, MA, USA) using an in-solution method. Peptides from all samples were desalted using a C18 cartridge (Waters, Milford, MA, USA). After lyophilization, the peptides were reconstituted in 20 µL of 0.1% formic acid solution (Sigma-Aldrich, St. Louis, MO, USA), and the peptide concentrations were determined by OD280. An appropriate amount of iRT standard peptides (Biognosys, Schlieren, Switzerland) was added to the digested peptides from each sample for mass spectrometry analysis.

Mass spectrometry methods

Instrument calibration was performed using Pierce™ Positive Ion Calibration Solution (Thermo Scientific, Waltham, MA, USA). Chromatographic separation was performed using a nanoflow Vanquish Neo system (Thermo Fisher Scientific, Waltham, MA, USA). Samples were injected onto a C18 analytical column (Thermo Scientific, ES906, 1.9 µm, 150 µm × 15 cm, Waltham, MA, USA) for linear gradient separation (0.1% formic acid in acetonitrile (80% acetonitrile), Sigma-Aldrich, St. Louis, MO, USA). The gradient started with 2% B for 5 min, increased to 8% B in 5 min, then to 25% B in 80 min, followed by an increase to 35% B in 10 min, 80% B in 5 min, and held at 80% B for 5 min, before re-equilibration at 2% B for 10 min. The total run time was 120 min, with a flow rate of 300 nL/min. After nano-HPLC separation, samples were analyzed by data-independent (DIA) mass spectrometry using an Astral high-resolution mass spectrometer (Thermo Scientific, Waltham, MA, USA). Detection mode: positive ion, parent ion scan range: 380–980 m/z, primary mass spectrometry resolution: 240,000 at 200 m/z, Normalized AGC Target: 500%, Maximum IT: 5 ms, MS2 adopts DIA data acquisition mode, sets 299 scan windows, Isolation Window: 2 m/z, HCD Collision Energy: 25 eV, Normalized AGC Target: 500%, Maximum IT: 3 ms.

Proteomic data quality control and preprocessing

Strict quality control and preprocessing were performed on the proteomics data from 69 samples of patients with poor collateral circulation. Proteins not detected in all samples (i.e., expression values of all zero or missing) were removed. Specific filtering criteria retained only proteins with non-zero expression values detected in at least one sample. All protein expression values were log-transformed to stabilize variance and normalize data distribution. To identify proteins with significant expression differences, the variance of the log-transformed expression values for each protein across the 69 samples was calculated. The top 1,000 proteins with the highest degree of variation were selected as the feature gene set. These highly variable proteins were normalized using median centering, where each protein's expression value was subtracted from its median expression value across all samples.

Principal component analysis

To validate the effectiveness of molecular typing, principal component analysis (PCA) was used to reduce and visualize the high-dimensional proteomic data. Based on the standardized protein expression matrix, the

singular value decomposition (SVD) algorithm was used to extract principal components. An orthogonal transformation converted original correlated variables into linearly uncorrelated principal components. The number of principal components to retain was determined by creating a scree plot and applying the elbow rule. PC1 and PC2 explained 26.4% and 4.9% of the variance, respectively. The first two principal components (PC1 and PC2) were used to construct a two-dimensional scatter plot, with 95% confidence ellipses illustrating the clustering and boundary demarcation of each molecular subtype.

Machine learning model construction and evaluation

Multiple machine learning classifiers were constructed for molecular subtype prediction. Stratified random sampling partitioned the dataset into training and test sets (8:2 ratio) to ensure consistent sample distribution across molecular subtypes in both datasets. The constructed machine learning models encompassed eight classic algorithms: ensemble learning, linear classifiers, instance-based, and tree-based algorithms, including XGBoost, random forest, logistic regression, support vector machine, k-nearest neighbor, AdaBoost, decision tree, and gradient boosting. Model hyperparameters were optimized using a grid search strategy combined with stratified cross-validation (folds = 4–10). Using accuracy as the primary evaluation metric, each model was evaluated across hundreds to thousands of parameter combinations. To comprehensively assess model performance, we established a multi-dimensional performance evaluation system, calculating and recording key metrics such as average cross-validation accuracy, training accuracy, test accuracy, weighted precision, weighted recall, and weighted F1 score. Receiver operating characteristic (ROC) curves were used to assess classification performance for each model, and the area under the curve (AUC) for each molecular subtype was calculated to quantify discriminative ability. The optimal parameter combination was selected based on test set accuracy, and a complete record of the hyperparameter search and corresponding performance metrics was saved for reproducibility. Finally, by comparing the classification performance and generalization ability of different models (including accuracy under 4–tenfold cross-validation), the optimal classification model for clinical molecular subtype prediction was determined.

The SHapley Additive exPlanations (SHAP) framework (version 0.41.0) was used to analyze the interpretability of the optimal classification model. The contribution of each clinical feature to molecular subtype prediction was quantified based on the game-theoretic Shapley value theory, providing an interpretable biological basis for clinical decision-making. A confusion matrix was used to evaluate the predictive performance of the optimal classification model. A 3×3 matrix visually displayed the model's classification accuracy for each molecular subtype and identified potential classification confusion patterns.

To verify the model's generalizability and clinical applicability, external validation was performed using the MIMIC-IV database. Eligible patient populations were selected from the MIMIC-IV database based on the study's inclusion and exclusion criteria, and clinical feature variables consistent with those in the training set were extracted. The preprocessed MIMIC-IV validation set was then fed into the trained optimal classification model to evaluate its classification performance on an independent external dataset, verifying the model's robustness and clinical translational potential.

Integrated clinical-proteomic correlation analysis

Based on the biological characteristics of molecular subtypes, we focused on analyzing the expression patterns of complement system-related proteins, glucose metabolism-related proteins, eosinophil marker proteins, and platelet function-related proteins. Protein expression distribution was visualized using boxplots combined with scatterplots, and statistically significant differences between marker groups were identified. Correlation analysis was performed based on the characteristic clinical indicators of each molecular subtype and protein expression profiles. The Spearman correlation coefficient assessed the strength and direction of association. A correlation network diagram was constructed using the linkET R package (version 0.0.1) to visually demonstrate the association patterns between clinical characteristics and protein expression.

Statistical analysis

All statistical analyses were performed using R (version 4.3.1) and Python (version 3.10). A fixed random seed (seed = 42) was used throughout all computational analyses to ensure reproducibility. All P values were two-tailed, with significance set at $P < 0.05$. For differential protein analyses and multiple comparisons of clinical features, False Discovery Rate (FDR) adjusted P values were calculated using the Benjamini–Hochberg method, with an adjusted $P < 0.05$ considered statistically significant.

Effect-size (Cohen's d) calculation [20]

For each protein, we computed the standardized mean difference:

$$d = \frac{\overline{X_1} - \overline{X_2}}{s_p},$$

where $\overline{X_1}$ and $\overline{X_2}$ represent the means of the two groups, and s_p is the pooled standard deviation, calculated as:

$$s_p = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}}.$$

Here, s_1^2 and s_2^2 are the respective variances, and n_1 and n_2 are the sample sizes of the two groups.

Power calculation [21]

Statistical power was computed using the noncentral t distribution for a two-sample, two-sided test with a significance level α = 0.05. The degrees of freedom (df) and noncentrality parameter (ncp) were calculated as:

Degrees of freedom: $\mathrm{df} = n_1 + n_2 - 2$

Noncentrality parameter: $\mathrm{ncp} = d\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$

The statistical power was then determined by:

$$\begin{aligned}\mathrm{Power} = 1 - \big[&F_{\mathrm{nct}}\left(t_{1-\alpha/2}; \mathrm{df}, \mathrm{ncp}\right) \\ &- F_{\mathrm{nct}}\left(-t_{1-\alpha/2}; \mathrm{df}, \mathrm{ncp}\right)\big],\end{aligned}$$

where F_{nct} denotes the CDF of the noncentral t distribution.

Back-calculation of minimum sample size

Given a desired effect size d, significance level α, target power, and an allocation ratio $r = n_2/n_1$, an iterative algorithm was employed to determine the minimum required sample size:

1. An initial estimate of n1 was obtained using a normal approximation method;
2. An integer search was performed around this initial estimate;
3. The calculated sample size was verified using the noncentral t-based power formula;
4. The minimum total sample size $n_{total} = n_1 + \lceil r \cdot n_1 \rceil$ that achieves the specified target power was returned.

Consensus clustering analysis

Molecular subtyping was performed using the ConsensusClusterPlus R package (version 1.64.0). This method employs an unsupervised learning strategy that improves classification stability through multiple resampling and cluster integration. We implemented the following parameter settings: Partitioning Around Medoids (PAM) was selected as the clustering algorithm with k values ranging from 2 to 10. The Spearman correlation coefficient was used as the distance metric due to its robustness to non-normally distributed data and its ability to capture monotonic relationships, which are common in biological datasets and less sensitive to outliers compared to Euclidean distance. We performed 1000 bootstrap iterations, randomly selecting 80% of samples in each iteration. Ward.D hierarchical clustering was employed for sample merging, chosen for its ability to minimize total within-cluster variance and provide stable, interpretable cluster solutions. This choice was further informed by its implementation as the default linkage method within the ConsensusClusterPlus R package that we utilized for our consensus clustering analysis. Crucially, while Ward.D2 typically uses squared Euclidean distances, Ward.D was specifically deemed appropriate for our dataset given the Spearman distance metric, which focuses on rank-based correlations. Our primary goal was to obtain stable and biologically interpretable cluster solutions, which this combination effectively achieved. The optimal number of clusters was determined by examining three key metrics: (1) the cumulative distribution function (CDF) curve, (2) the relative change in area under the CDF curve (delta area plot), and (3) consensus matrices. We applied the "elbow method" to identify the point where the rate of increase in the area under the CDF curve significantly decreased, which indicated 2 as the optimal number of clusters, successfully dividing the study cohort into two subpopulations with distinct molecular characteristics. Cluster robustness was assessed using silhouette width analysis (using the cluster package, version 2.1.4) and bootstrap stability measures.

Differential protein analysis

Differential protein expression analysis was performed using the limma R package (version 3.56.2), which implements linear models with empirical Bayesian moderation of standard errors. This approach is particularly suited for proteomics data as it addresses challenges associated with small sample sizes and high dimensionality. Three comparison strategies were constructed: (1) each molecular subtype versus the control group; (2) direct comparison between two molecular subtypes; and (3) poor collateral circulation versus good collateral circulation.

For each comparison, a design matrix was constructed to represent the experimental groups, and linear models were fitted using the lmFit function. Statistical inference was performed using the empirical Bayesian method via the eBayes function, which shrinks protein-wise sample variances toward a common value, increasing statistical power. The Benjamini–Hochberg method was applied for multiple testing correction to control the false discovery rate. Differential proteins were identified using criteria of absolute log2FoldChange > 0.5 and FDR-adjusted *P* value < 0.05. According to the direction of expression change, differential proteins were classified into three types: up-regulated, down-regulated, and no significant change. To comprehensively analyze the overlapping relationship between differential proteins in different comparison groups, Venn diagrams and UpSet diagrams

were constructed for visual display using the UpSetR package (version 1.4.0). Set operations were used to identify specific differential proteins and common differential proteins in each comparison group, and the number of proteins contained in each intersection and union was counted. The ggvolcano package (version 0.1.1) was used to draw volcano plots to intuitively display the statistical significance and biological significance of differential proteins between molecular subtypes.

Functional enrichment analysis

Based on the results of differential protein analysis, functional enrichment analysis was performed using the gseapy Python package (version 1.0.5) to elucidate the biological characteristics of each molecular subtype. Differential proteins were categorized as upregulated or downregulated based on the direction of expression change, and gene lists were constructed for enrichment analysis. The KEGG pathway database was used as a reference gene set, and the hypergeometric test was used to assess the statistical significance of gene set enrichment. Pathways with an adjusted P-value < 0.05 were considered significantly enriched.

Clinical data analysis and feature selection

Fifty-six baseline clinical characteristics were collected. Variables with more than 50% missing values were removed to ensure data quality and analytical robustness. For remaining missing values, a stratified imputation strategy was used: continuous variables were imputed using the k-nearest neighbor algorithm (KNN, k = 5) within each molecular subtype, and categorical variables were imputed using the mode. To address class imbalance between molecular subtypes, a SMOTENC hybrid resampling strategy was employed, oversampling the minority class (50 samples each for CDVR and ITMD) and randomly undersampling the control group (n = 50) to achieve balanced representation. All preprocessing steps were implemented using the Python scikit-learn (version 1.2.2) and imbalanced-learn (version 0.11.0) libraries.

To assess the strength of associations between clinical characteristics, we used the Spearman rank correlation coefficient for correlation analysis. To control for false-positive results due to multiple comparisons, we corrected the p-values for multiple testing for all pairwise comparisons using the Benjamini–Hochberg method (also known as False Discovery Rate correction). To assess distributional symmetry of continuous variables, skewness analysis was performed using the skew function in the SciPy library (version 1.10.1). A skewness value of 0 indicates a completely symmetrical data distribution; a positive skewness indicates a right-skewed distribution with many maxima; a negative skewness indicates a left-skewed distribution with many minima. Appropriate statistical methods compared clinical characteristics between groups. Continuous variables were tested for normality (Shapiro–Wilk test) and homogeneity of variance (Levene test). For normally distributed data with homogeneous variance, one-way analysis of variance (ANOVA) was used for overall comparisons between groups, and independent sample t-tests were used for pairwise comparisons between groups. For normally distributed data with unequal variances, Welch analysis of variance was used, and unequal variance t-tests were used for pairwise comparisons between groups. For data with non-normal distribution, the Kruskal–Wallis nonparametric test was used for overall comparisons between groups, and the Mann–Whitney U test was used for pairwise comparisons between groups. Normal distribution data were presented as mean ± standard deviation (SD), while non-normal distribution data were presented as median interquartile range (IQR). Categorical variables were described using frequencies and percentages, and group comparisons used the chi-square test or Fisher's exact test. The statistical significance level was set at $P < 0.05$, with $P < 0.05$, $P < 0.01$, and $P < 0.001$ marked as *, **, and *** respectively. Feature importance was assessed using the BorutaPy algorithm combined with the XGBoost classifier. This algorithm constructs random features as a baseline and iteratively compares the importance of the original features with the random features to identify key features with true predictive value. Based on feature importance ranking, the top 10 important features were selected for subsequent modeling and analysis to ensure the biological significance and clinical predictive value of the selected features.

Survival and prognostic analysis

Major adverse cardiovascular events (MACE) were the primary endpoint, with survival expressed in days. Patients lost to follow-up were censored. Survival functions for each molecular subtype were estimated using the Kaplan–Meier method, and survival curves were plotted to compare prognostic differences. The log-rank test was used to assess the statistical significance of survival differences between groups, and pairwise comparisons and multivariate analyses were performed. All survival analyses were performed using the survival package in R (version 3.7-0) and the lifelines library in Python (version 0.27.0).

Cox proportional hazards regression analysis

To identify potential risk factors associated with major adverse cardiovascular events (MACE), we performed univariate and multivariate Cox proportional hazards regression analyses using the ezcox (version 1.0.0) and survival (version 3.4-0) R packages. In univariate analysis,

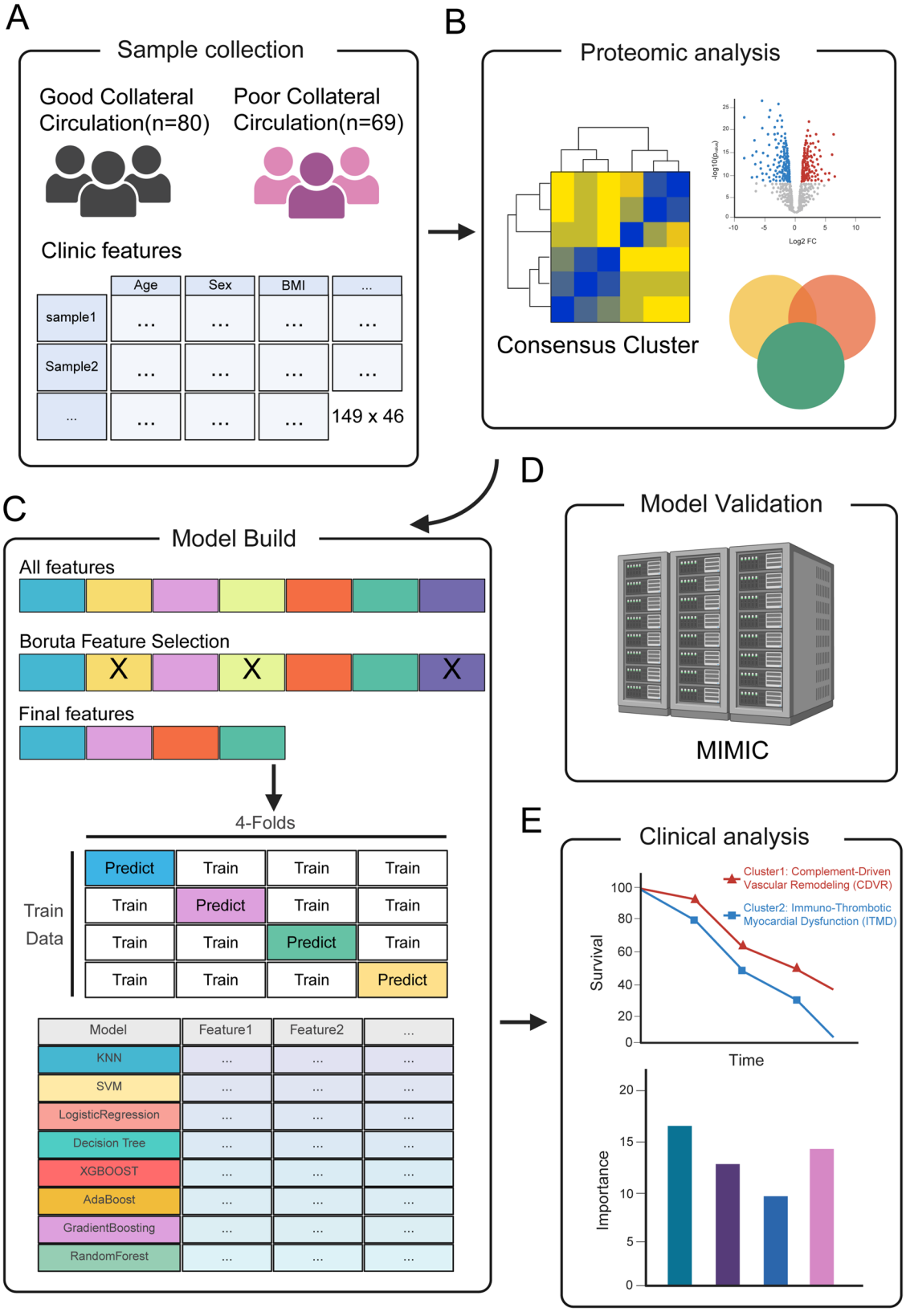


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Workflow for identification and validation of molecular phenotypes associated with poor coronary collateral circulation (CCC). **A** Patient grouping and clinical data collection. The discovery cohort included 80 patients with good CCC and 69 with poor CCC, with 46 clinical variables recorded for each case. **B** Proteomic analysis of patients with poor CCC identified two molecular clusters using consensus clustering. Volcano plots show differentially expressed proteins (DEPs) between clusters; Venn diagrams depict cluster-specific and shared DEPs. **C** Machine learning workflow. All clinical features were subjected to Boruta feature selection to identify the most informative variables. Models were trained and validated using fourfold cross-validation across eight algorithms (KNN, SVM, logistic regression, decision tree, XGBoost, AdaBoost, gradient boosting, and random forest). **D** External validation of molecular phenotypes using the MIMIC database. **E** Clinical outcome analysis. Kaplan–Meier curves show worse survival for Immuno-Thrombotic Myocardial Dysfunction (ITMD, Cluster 2) compared with Complement-Driven Vascular Remodeling (CDVR, Cluster 1); bar plots display the importance of selected features

each clinical and molecular variable was individually evaluated. Variables showing significance ($P < 0.1$) in univariate analysis, along with clinically important factors based on prior knowledge, were included in multivariate models. The proportional hazards assumption was assessed using Schoenfeld residuals (cox.zph function in the survival package) for each variable and globally for the entire model. No significant violations of this assumption were detected (all $P > 0.05$). Cox regression models quantified associations using hazard ratio (HR) and its 95% confidence interval (CI). An HR > 1 indicated a risk factor, while an HR < 1 indicated a protective factor. The statistical significance of the model was assessed using the Wald test, with $P < 0.05$ considered significant. The model's predictive performance was evaluated using Harrell's concordance index (C-index), and internal validation was performed using 1000 bootstrap resamples to assess potential overfitting.

Software and statistical packages

All analyses were performed using R (version 4.3.1) and Python (version 3.10). Key R packages included: ConsensusClusterPlus (v1.64.0) for molecular subtyping, limma (v3.56.2) for differential analysis, ggplot2 (v3.4.2) for visualization, UpSetR (v1.4.0) for set visualization, ggvolcano (v0.1.1) for volcano plots, cluster (v2.1.4) for silhouette analysis, linkET (v0.0.1) for correlation networks, ezcox (v1.0.0) and survival (v3.4-0) for survival analysis, and survminer (v0.4.9) for survival visualization. Key Python packages included: scikit-learn (v1.2.2) for machine learning algorithms and data preprocessing, imbalanced-learn (v0.11.0) for handling class imbalance, SciPy (v1.10.1) for statistical computations, forestploter (v1.1.2) for forest plots, gseapy (v1.0.5) for functional enrichment analysis, SHAP (v0.41.0) for model interpretation, and lifelines (v0.27.0) for survival analysis.

Results

Study workflow

We implemented a systematic study design to identify and validate molecular phenotypes associated with poor coronary collateral circulation (CCC). The discovery cohort comprised 149 patients, including 80 with good CCC and 69 with poor CCC. Detailed clinical characteristics were recorded, resulting in a 46-clinical-variable × 149-sample data matrix, covering demographic variables, anthropometric indices, and laboratory measurements (Fig. 1A). High-throughput proteomic profiling was performed in the poor CCC group (n = 69). Consensus clustering analysis revealed two distinct molecular subtypes. Differential expression analysis visualized by volcano plots demonstrated specific molecular signatures between clusters, and Venn diagrams illustrated both cluster-specific and shared differentially expressed proteins (Fig. 1B). To build a clinically applicable prediction model, all available clinical features were initially included. The Boruta feature selection algorithm was applied to identify the most informative variables and reduce dimensionality. The final features were used to construct models. To determine the optimal cross-validation strategy, models were evaluated using stratified cross-validation across a range of folds [4–10]. The cross-validation accuracy trend is visualized in Fig. S1, with detailed performance metrics provided in Table S1. Based on these evaluations, a fourfold cross-validation scheme, with 75% of data for training and 25% for validation in each fold, was selected for the final model construction. Eight machine learning algorithms—k-nearest neighbors (KNN), support vector machine (SVM), logistic regression, decision tree, XGBoost, AdaBoost, gradient boosting, and random forest were compared to determine the optimal classifier (Fig. 1C). External validation of the identified molecular phenotypes was conducted using the Medical Information Mart for Intensive Care (MIMIC) database (Fig. 1D). Lastly, survival analysis assessed long-term outcomes between the two clusters. Patients in ITMD exhibited significantly worse survival compared with CDVR, supporting the prognostic relevance of the identified molecular phenotypes (Fig. 1E). Quantitative effect-size and power analyses indicated that the observed sample size (n = 69) was adequate for the subtype comparison (median Cohen's d ≈ 0.97; achieved power ≈ 97.6%; median minimum total n for 80% power = 38; Fig. S2, Table S2).

Proteomic clustering reveals two distinct molecular subtypes in poor coronary collateral circulation

Initially, a differential expression analysis comparing poor CCC patients (n = 69) to good CCC patients (n = 80) was performed. This analysis identified a total

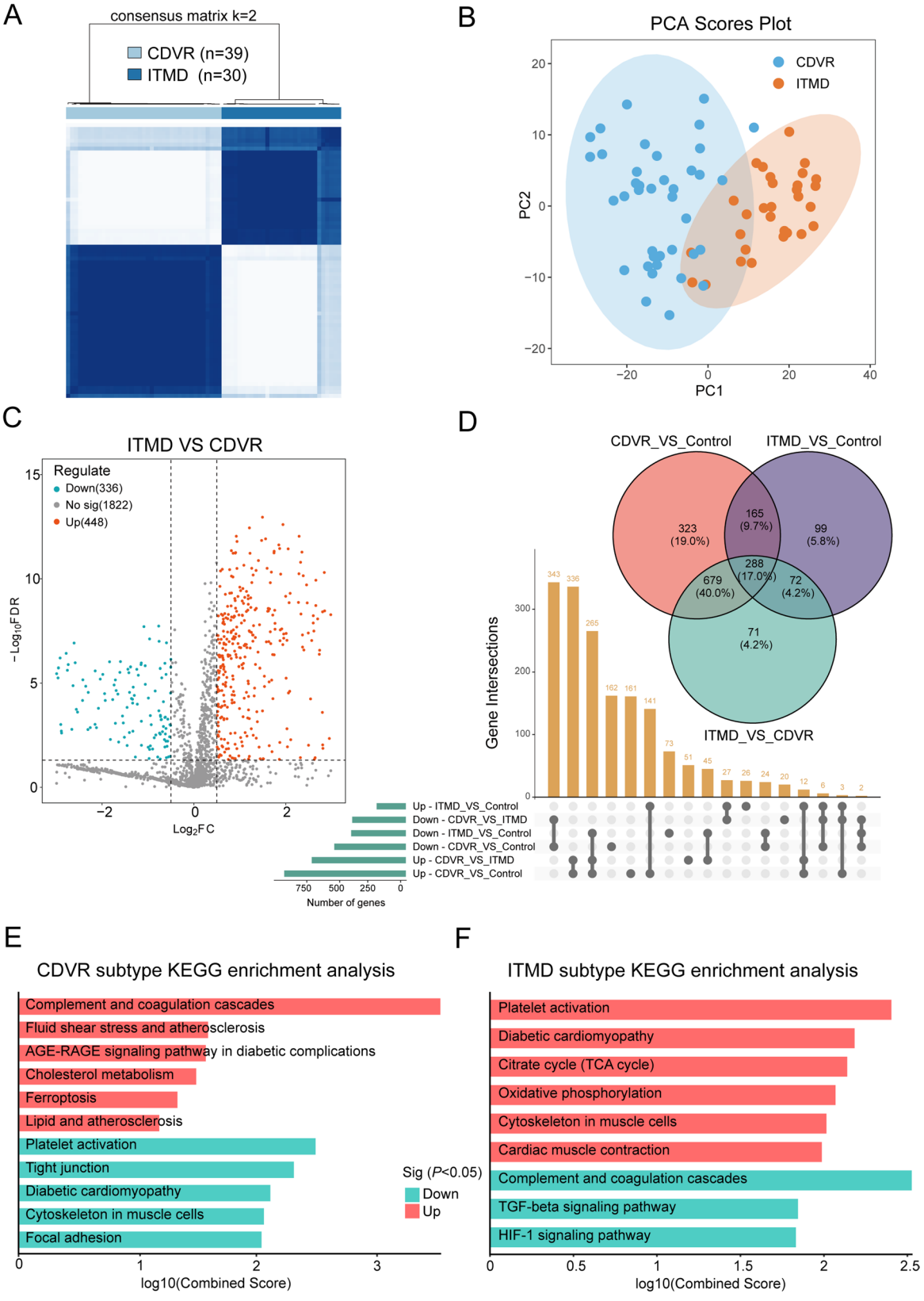


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Consensus clustering and proteomic characterization of two phenotypes in poor coronary collateral circulation. **A** Consensus clustering heatmap showing the optimal cluster number (k = 2) with 39 patients in Complement-Driven Vascular Remodeling (CDVR) and 30 patients in Immuno-Thrombotic Myocardial Dysfunction (ITMD). **B** PCA plot demonstrating distinct separation between CDVR and ITMD based on proteomic profiles. **C** Volcano plot showing 448 upregulated (red) and 336 downregulated (green) proteins in ITMD compared to CDVR. **D** Venn diagram and UpSet plot illustrating the overlap and specificity of DEGs in CDVR vs. Control, ITMD vs. Control, and CDVR vs. ITMD comparisons. **E** KEGG pathway enrichment analysis for CDVR, highlighting upregulated pathways (red) and downregulated pathways (green). **F** KEGG pathway enrichment analysis for ITMD, highlighting enriched and suppressed pathways. Control: patients with good CCC (Rentrop grade 2–3); Cluster 1 (Complement-driven vascular remodeling subtype) and Cluster 2 (Immuno-thrombotic myocardial dysfunction): two phenotypes within poor CCC patients (Rentrop grade 0–1)

of 532 differentially expressed proteins (DEPs), with 198 upregulated and 334 downregulated in poor CCC (Fig. S3A, Table S3). To explore potential molecular subtypes in patients with poor CCC, we performed unsupervised consensus clustering on proteomic profiles from 69 patients. The optimal cluster number was determined to be k = 2, resulting in two distinct groups: Cluster 1 Complement-Driven Vascular Remodeling (CDVR) (n = 39) and Cluster 2 Immuno-Thrombotic Myocardial Dysfunction (ITMD) (n = 30) (Fig. 2A). Cluster stability analysis indicated high internal consistency and clear separation between the two clusters. Principal component analysis (PCA) further confirmed the molecular distinction, with PC1 and PC2 explaining 26.43% and 14.95% of the total variance, respectively. The two clusters exhibited distinct spatial separation on the PCA plot, with CDVR mainly positioned on the left and ITMD on the right, demonstrating robust clustering performance (Figs. 2B, S3B). Differential expression analysis between the clusters identified 448 upregulated and 336 downregulated proteins, with most differentially expressed proteins upregulated in ITMD (Fig. 2C, Table S4). UpSet and Venn diagram analyses showed the overlap and specificity of differentially expressed proteins (DEPs) across three comparisons: CDVR vs. Control, ITMD vs. Control, and CDVR vs. ITMD. These analyses revealed 323 DEPs unique to CDVR vs. Control, 99 unique toITMD vs. Control, and 71 unique to CDVR vs. ITMD, with 288 DEPs shared across all three comparisons (Fig. 2D, Table S5). KEGG pathway enrichment analysis revealed distinct biological signatures in each cluster. For Cluster 1 CDVR, enriched pathways included complement and coagulation cascades, fluid shear stress and atherosclerosis, AGE-RAGE signaling pathway in diabetic complications, cholesterol metabolism, ferroptosis, and lipid and atherosclerosis pathways. Meanwhile, CDVR showed downregulation in platelet activation, tight junction, diabetic cardiomyopathy, cytoskeleton in muscle cells, and focal adhesion pathways. In contrast, Cluster 2 ITMD was enriched in platelet activation, diabetic cardiomyopathy, citrate cycle (TCA cycle), oxidative phosphorylation, cytoskeleton in muscle cells, and cardiac muscle contraction, while showing downregulation in complement and coagulation cascades, TGF-β signaling pathway, and HIF-1 signaling pathway (Fig. 2E, F). These findings reveal the molecular heterogeneity underlying poor coronary collateral circulation and suggest distinct pathophysiological mechanisms driving impaired collateralization in different patient subgroups. The CDVR subtype appears to be primarily characterized by dysregulated inflammatory responses and vascular remodeling deficiencies, with prominent involvement of complement activation and lipid metabolism pathways. Conversely, the ITMD subtype demonstrates features consistent with altered myocardial energetics, platelet hyperactivity, and cardiac structural abnormalities. The identification of these distinct molecular subtypes may explain the variable clinical outcomes and treatment responses observed in poor CCC patients.

Eosinophil parameters and glycemic indices distinguish molecular subtypes in poor CCC

To identify the clinical parameters most relevant to the two molecular phenotypes of poor coronary collateral circulation (CCC), we first assessed the skewness of all recorded variables. Several parameters exhibited high skewness (skewness > 2.5), including red cell distribution width–coefficient of variation (RDW–CV), creatinine, triglyceride, and eosinophil percentage, suggesting potential differences between the two clusters. Moderate skewness (1.0–2.5) was observed for total cholesterol, blood urea nitrogen (BUN), eosinophil count, and basophil count, while most other parameters were approximately normally distributed (Fig. 3A). XGBoost-based feature importance ranking identified the most discriminative clinical markers between the two phenotypes. Eosinophil percentage emerged as the top predictor, followed by fasting plasma glucose and glycated haemoglobin. Eosinophil count also ranked highly. Additional contributing variables included BMI, monocyte count, total white blood cell count, basophil percentage, neutrophil count, and mean red cell haemoglobin concentration, with importance scores ranging from 1.0 to 3.0 (Fig. 3B). Spearman correlation analysis revealed distinct association patterns among clinical features. Eosinophil-related parameters (count and percentage) displayed specific correlations with other inflammatory markers. Metabolic parameters, including fasting plasma glucose, HbA1c, and lipid profile, formed a relatively independent cluster, suggesting a separation between inflammatory and metabolic axes. Red cell indices such as haemoglobin, red blood cell count, and mean corpuscular volume

exhibited strong internal correlations (Fig. 3C). Overall, eosinophil-related measures, glycaemic indices (fasting plasma glucose and HbA1c), and selected haematological markers formed the core biomarker panel distinguishing the two CCC phenotypes.

Complement/Metabolic and Eosinophil/Platelet signatures define poor CCC subtypes

To elucidate the molecular characteristics of the two phenotypes, we first constructed protein–clinical parameter correlation matrices for each cluster. In CDVR subtype, protein expression patterns were strongly and positively correlated with neutrophil count, neutrophil percentage, fasting plasma glucose, and glycated haemoglobin (HbA1c) (Fig. 4A). In contrast, ITMD exhibited a distinct correlation pattern, with protein expression mainly associated with eosinophil count, eosinophil percentage, and platelet count (Fig. 4B). Proteomic analysis revealed that complement activation was a prominent feature of CDVR, with key components C3, C5, C6, and C9 significantly upregulated (fold change 1.5–3, all $P < 0.0001$) (Fig. 4C). This cluster also exhibited metabolic reprogramming, with significantly increased expression of genes such as ALDOB, RBP4, CST3, and ADIPOQ (Fig. 4D). In contrast, ITMD was characterized by platelet and eosinophil activation. Key platelet receptors (GP1BA, ITGA2B, ITGB3) and eosinophil-derived proteins (EPX, RNASE2, and RNASE3) were significantly upregulated (all $P < 0.0001$) (Fig. 4E, F), suggesting a prothrombotic and inflammatory phenotype. STRING protein–protein interaction analyses further depicted the functional organization of these molecular features. In CDVR, ALB occupied a central position in the glucose-associated protein network, closely linking CST3, RBP4, HBA2, ADIPOQ, ALDOB, and HBB (Fig. 4G), and in ITMD, EPX was the central hub in the eosinophil-associated protein network, interacting with PRG3, FCER1G, RNASE3, and RNASE2 (Fig. 4H).

Machine learning model optimization and performance validation for phenotype classification

We evaluated eight supervised machine learning algorithms for their ability to discriminate between controls, CDVR and ITMD phenotypes. Among the tested models, XGBoost achieved the highest overall performance on the test dataset (accuracy = 80.0%, precision = 80.91%, recall = 80.0%, F1 = 80.31%), followed closely by Gradient Boosting (accuracy = 73.33%) and Decision Tree (accuracy = 76.67%). Logistic Regression, Random Forest, KNN, SVM, and AdaBoost exhibited lower test accuracies (≤ 69.17%) (Fig. 5A). SHAP (SHapley Additive exPlanations) analysis of the XGBoost model identified fasting plasma glucose, eosinophil percentage, and HbA1c as the top three contributors to classification performance, followed by white blood cell count, BMI, and eosinophil count (Fig. 5B). Radar plot comparison confirmed that XGBoost and Gradient Boosting consistently outperformed other models across accuracy, precision, recall, and F1 score metrics (Fig. 5C). Furthermore, model calibration was assessed using calibration curves (Fig. 5D). XGBoost demonstrated excellent calibration, with its predicted probabilities closely aligning with the ideal 45-degree line, particularly in the low to moderate probability ranges, indicating reliable probability estimates. Gradient Boosting also showed good calibration, while models like Decision Tree, KNN, and AdaBoost exhibited greater deviations from the ideal curve, suggesting poorer calibration of their probability outputs. ROC curve analysis demonstrated excellent discriminative ability for all three classes using XGBoost (AUC = 0.912 for controls, 0.965 for CDVR, and 0.963 for ITMD) (Fig. 5E). The confusion matrix revealed that the model achieved high classification accuracy, with minimal misclassification between CDVR and ITMD, and the majority of errors occurring between control and disease phenotypes (Fig. 5F).

To provide a more granular understanding of the distinction between the two poor CCC subtypes, we conducted a dedicated binary classification analysis focusing solely on discriminating Cluster 1 CDVR from Cluster 2 ITMD. Similar to the 3-class classification, a comprehensive model selection for this binary task was performed (Fig. S4A). Among the tested algorithms, XGBoost once again emerged as the best-performing model, achieving superior test performance metrics including an accuracy of 95.0%, precision of 96.97%, recall of 95.0%, and F1 score of 95.97% (Fig. S4A). SHAP analysis of this binary model revealed key features contributing to the distinction between Cluster 1 CDVR and Cluster 2 ITMD, with HbA1c and fasting plasma glucose being prominent contributors, alongside other variables (Fig. S4B). The ROC curve for this binary XGBoost classifier further confirmed its excellent discriminative ability with an AUC of 0.981 (Fig. S4C). The confusion matrix for this binary classification demonstrated a very high degree of separation between the two subtypes, with only minimal misclassifications (Fig. S4D).

Collectively, these analyses demonstrate the robust capability of machine learning, particularly the XGBoost model, to accurately classify complex phenotypes, not only distinguishing between controls and disease but also effectively differentiating between closely related disease subtypes. This strong performance, supported by interpretable feature importance, highlights the potential of such models for precise phenotypic characterization and clinical decision support.

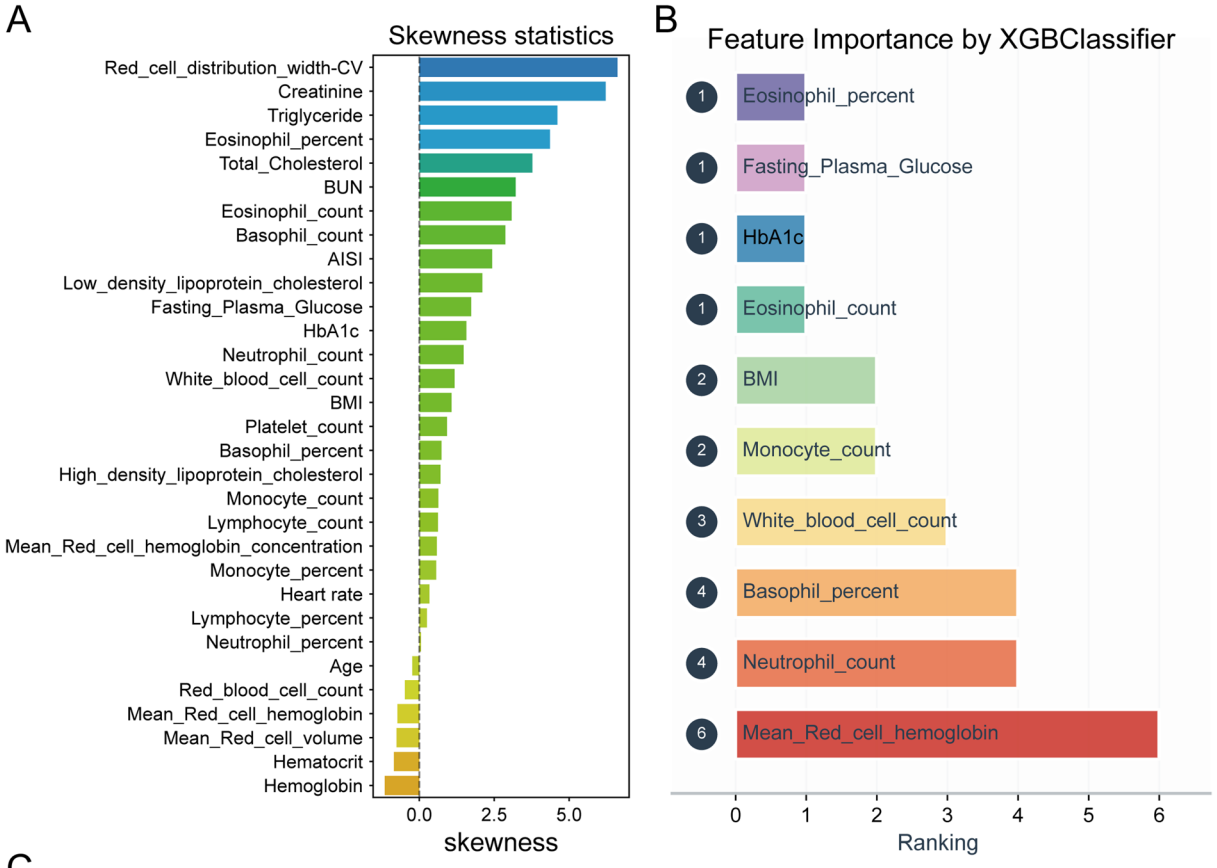


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Machine learning based feature selection and clinical correlation analysis for poor CCC phenotypes. **A** Skewness statistics for all clinical variables in patients with poor CCC, highlighting parameters with high skewness (> 2.5). **B** Feature importance ranking derived from XGBoost classification, showing the most discriminative clinical parameters between CDVR and ITMD. **C** Spearman correlation heatmap of clinical parameters, illustrating distinct clustering of eosinophil-related indicators, metabolic variables, and red cell indices. False Discovery Rate (FDR) correction was applied using the Benjamini–Hochberg method. Asterisks indicate significance levels for FDR-adjusted p-values: *$p < 0.05$, **$p < 0.01$, *$p < 0.001$. CDVR: Driven Vascular Remodeling CDVR subtype, ITMD: Immuno-Thrombotic Myocardial Dysfunction subtype

External validation confirms distinct prognostic profiles of poor CCC phenotypes.

Clinical feature comparison between the two poor CCC phenotypes and controls revealed distinct metabolic and inflammatory profiles. In the discovery cohort, fasting plasma glucose and HbA1c levels were significantly higher in CDVR compared with ITMD and controls ($P < 0.001$ and $P < 0.05$, respectively), consistent with a hyperglycaemic phenotype. Other haematological and biochemical parameters, including complete blood count, lipid profile, and renal function, demonstrated no significant inter-cluster differences in this cohort (Table 2). These findings were further validated using the MIMIC database. In the external validation cohort, BMI, white blood cell count, monocyte count, basophil percentage, mean red cell haemoglobin, fasting plasma glucose, and HbA1c all differed significantly between phenotypes and controls (overall $P < 0.001$ for most comparisons). Notably, eosinophil percentage and basophil percentage were markedly higher in ITMD, supporting the eosinophil–platelet axis observed in molecular analyses, whereas fasting plasma glucose and HbA1c were elevated in CDVR, aligning with the Complement-Driven Vascular Remodeling phenotype (Fig. 6A). Kaplan–Meier survival analysis demonstrated significant differences in major adverse cardiovascular event (MACE) rates across groups (log-rank $P < 0.05$). ITMD patients exhibited the highest cumulative event rate, followed by CDVR, while controls had the lowest risk (Fig. 6B). Univariate Cox regression analysis identified fasting plasma glucose as significantly associated with increased MACE risk in CDVR ($P < 0.05$) and eosinophil count as a significant predictor in ITMD ($P < 0.05$) (Fig. S5). Importantly, in multivariable Cox models fitted within each cluster and adjusted for age, sex, hypertension, and dyslipidemia, HbA1c and fasting plasma glucose in CDVR and eosinophil percentage/count in ITMD remained significant predictors of MACE (Fig. 6C). These results link key cluster-specific clinical features to adverse cardiovascular outcomes, reinforcing their potential utility for risk stratification.

Discussion

In this integrative proteomic and clinical study, we identified two reproducible immuno-metabolic phenotypes associated with poor collateral circulation, each underpinned by distinct biological pathways and prognostic implications. Whereas traditional evaluations of collateral status have largely relied on angiographic grading or individual inflammatory markers [2, 22], our work combines high-dimensional proteomics with machine learning to delineate mechanistically distinct subgroups: Complement-Driven Vascular Remodeling (CDVR, Cluster 1) and Immuno-Thrombotic Myocardial Dysfunction (ITMD, Cluster 2). This dual-axis framework extends existing knowledge of collateral biology by anchoring phenotypic classification to immune–metabolic mechanisms, a perspective particularly relevant to the diabetic vasculature [17, 23, 24]. The replication of these phenotypes in the heterogeneous MIMIC cohort strengthens the robustness of our findings and supports the feasibility of translating them into bedside risk stratification.

CDVR (Cluster 1) was characterized by enrichment of complement and coagulation cascades, elevated neutrophil counts, and higher HbA1c. Specifically, our KEGG enrichment analysis for CDVR clearly showed significant upregulation of "Complement and coagulation cascades," "Fluid shear stress and atherosclerosis," "AGE-RAGE signaling pathway in diabetic complications," "Cholesterol metabolism," "Ferroptosis," and "Lipid and atherosclerosis." Concurrently, pathways related to vascular integrity and cellular function, such as "Platelet activation," "Tight junction," "Diabetic cardiomyopathy," "Cytoskeleton in muscle cells," and "Focal adhesion" were significantly downregulated. This pathway activation is further supported by our clinical data showing significantly higher fasting plasma glucose and HbA1c in this cluster, and the SHAP analysis identifying fasting plasma glucose and HbA1c as top contributors to classification. The correlation network also highlights strong positive correlations between these clinical markers and neutrophil counts. Moreover, gene expression analysis confirmed the upregulation of key complement-related genes (C3, C4A, C5, C7, C9, CD14, CD44) and glucose metabolism-related genes (HBB, HBA2, ALB). This aligns with the well-established interplay between chronic hyperglycemia and complement activation in diabetic microvascular injury [25–27]. Sustained hyperglycemia can lead to protein glycation, which in turn modifies complement regulatory proteins like factor H, thereby impairing their inhibitory function and exacerbating vascular inflammation [26]. Elevated HbA1c levels in this cluster reinforce the relevance of cumulative glucose exposure in driving these processes, consistent with observations in diabetic retinopathy and nephropathy [28, 29]. Our proteomic data

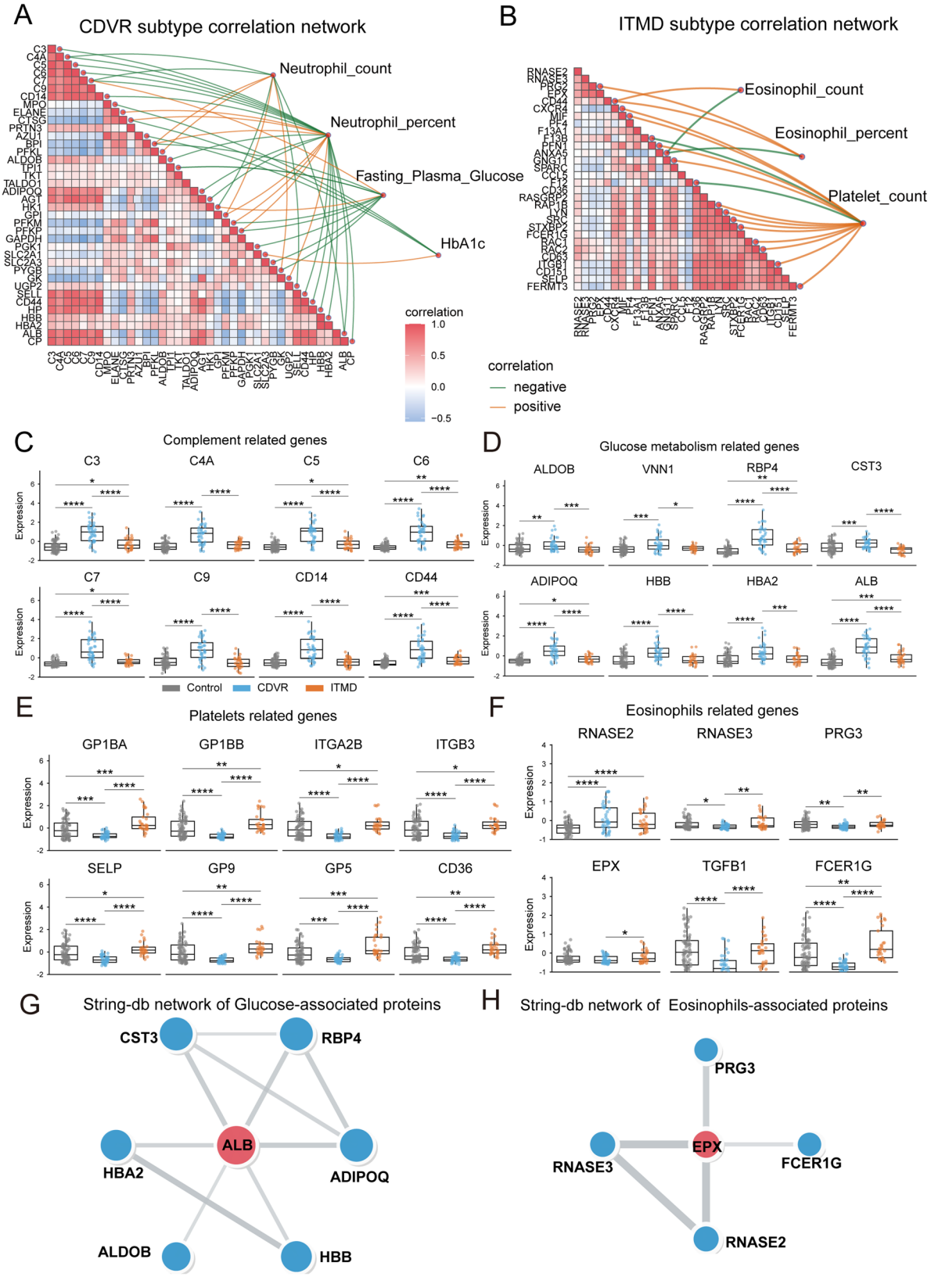


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Correlation patterns and specific molecular pathways characterising poor CCC phenotypes. **A** Protein–clinical correlation heatmap and network for CDVR, showing strong associations with neutrophil count, neutrophil percentage, fasting plasma glucose, and HbA1c. **B** Protein–clinical correlation heatmap and network for ITMD, highlighting associations with eosinophil count, eosinophil percentage, and platelet count. **C** Expression profiles of complement-related genes, markedly elevated in CDVR but not in ITMD. **D** Expression of glucose metabolism–related genes enriched in CDVR. **E** Platelet-related gene expression elevated in ITMD. **F** Eosinophil-related gene expression enriched in ITMD. **G** STRING network of glucose-associated proteins in CDVR, with ALB as the central hub. **H** STRING network of eosinophil-associated proteins in ITMD, with EPX as the central hub. CDVR: Driven Vascular Remodeling CDVR subtype, ITMD: Immuno-Thrombotic Myocardial Dysfunction subtype

also revealed an over-representation of neutrophil extracellular traps (NETs) formation pathways in CDVR. NETs have been implicated in endothelial damage, microvascular occlusion, and impaired collateral remodeling in ischemic tissues [30–32]. Hyperglycemia can enhance NETs release through oxidative stress and advanced glycation end-product (AGE) pathways, providing a mechanistic link between metabolic dysregulation and innate immune over-activation [33, 34]. The constellation of complement activation, neutrophilia, and hyperglycemia observed here thus delineates a metabolic-inflammatory phenotype in which chronic glucose toxicity amplifies innate immunity, ultimately hampering arteriogenesis. Given these insights, targeted therapeutic approaches for CDVR patients could involve intensive glycemic control to mitigate hyperglycemia-driven damage, coupled with complement inhibition (e.g., C5 blockade) to dampen the inflammatory cascade. Strategies addressing cholesterol and lipid metabolism, as indicated by the upregulated KEGG pathways, could also be beneficial in managing the vascular remodeling component and reducing atherosclerotic burden.

However, our comprehensive analysis of the ITMD subtype revealed two interconnected pathophysiological mechanisms driving myocardial dysfunction, which together explain the unique phenotypic characteristics observed in this patient population. ITMD displayed a distinctive inflammatory signature characterized by significant upregulation of eosinophil-associated granule proteins (EPX, PRG3, RNASE2) and platelet receptor signaling molecules (GP1BA, GP1BB, ITGA2B, ITGB3, SELP, GP9, GP5, CD36), with KEGG enrichment analysis showing prominent activation of pathways directly related to cardiac function and metabolism, including "Cardiac muscle contraction," "Diabetic cardiomyopathy," and "Oxidative phosphorylation" [35, 36]. This transcriptomic profile strongly suggests direct myocardial involvement and altered cardiac energy metabolism in ITMD, distinguishing it from other cardiovascular disease subtypes. Clinically, ITMD patients exhibited markedly higher eosinophil and basophil percentages, with SHAP analysis identifying eosinophil parameters as key distinguishing features of this subtype. Eosinophils, traditionally viewed primarily as effectors in allergic diseases and parasitic infections, have emerged as critical mediators in cardiac pathophysiology through complex immunomodulatory functions [37]. Eosinophils exert direct cardiotoxic effects through multiple mechanisms, with eosinophil granule proteins such as Major Basic Protein (MBP) and Eosinophil Cationic Protein (ECP) disrupting cardiomyocyte membrane integrity, inducing cytotoxicity, and interfering with cardiac ion channels [38]. DeMello et al. demonstrated cardiac localization of eosinophil granule major basic protein in acute necrotizing myocarditis [39], while Tai et al. confirmed deposits of eosinophil granule proteins in cardiac tissues of patients with eosinophilic endomyocardial disease, providing direct evidence for eosinophil-mediated cardiac pathology [40]. These highly cationic proteins can directly compromise cardiomyocyte function and initiate inflammatory cascades that perpetuate tissue damage and adverse remodeling. The severity of cardiac dysfunction in ITMD patients likely reflects this direct toxicity mechanism, which distinguishes it from other inflammatory cardiac conditions. The cytokine profile secreted by eosinophils, including IL-4, IL-13, and TGF-β, profoundly influences cardiac fibroblast phenotype and function, orchestrating a type-2 immune response that critically affects cardiac remodeling. Diny et al. demonstrated that eosinophil-derived IL-4 drives progression of myocarditis to dilated cardiomyopathy through effects on cardiac fibroblasts and alternative macrophage activation [41]. In experimental models, Toor et al. showed that eosinophil deficiency promotes aberrant repair and adverse remodeling following acute myocardial infarction by altering fibroblast activation and extracellular matrix organization [36]. Liu et al. further established that eosinophils significantly influence cardiac pathology through IL-4-mediated activation of alternatively activated macrophages, creating an immunological milieu that critically impacts myocardial repair processes [42]. The unique pattern of cytokine production and immune cell recruitment orchestrated by eosinophils creates a tissue environment that can either promote adaptive repair or pathological remodeling, depending on the context and duration of the response. Metabolic dysregulation represents another crucial aspect of eosinophil-mediated cardiac dysfunction in ITMD. Our pathway analysis revealed significant upregulation of key metabolic pathways (citrate cycle, oxidative phosphorylation) in ITMD patients.

Yang et al. demonstrated that eosinophil-derived mediators directly alter cardiomyocyte metabolism, with eosinophil expression of IL-4, IL-13, and mEar1 being essential to control cardiomyocyte hypertrophy

A **Model Selection**

Model	CV_accuracy	Train_Accuracy	Test_Accuracy	Test_Precision	Test_Recall	Test_F1
GradientBoosting	73.33%	91.67%	73.33%	76.07%	73.33%	73.87%
XGBoost	70.83%	99.17%	80.00%	80.91%	80.00%	80.31%
DecisionTree	70.00%	87.50%	76.67%	77.21%	76.67%	76.80%
LogisticRegression	66.67%	71.67%	60.00%	64.81%	60.00%	60.93%
RandomForest	65.83%	69.17%	66.67%	72.22%	66.67%	67.95%
KNN	64.17%	100.00%	73.33%	74.68%	73.33%	73.29%
SVM	62.50%	70.00%	63.33%	68.52%	63.33%	64.44%
AdaBoost	55.00%	57.50%	56.67%	47.83%	56.67%	47.65%

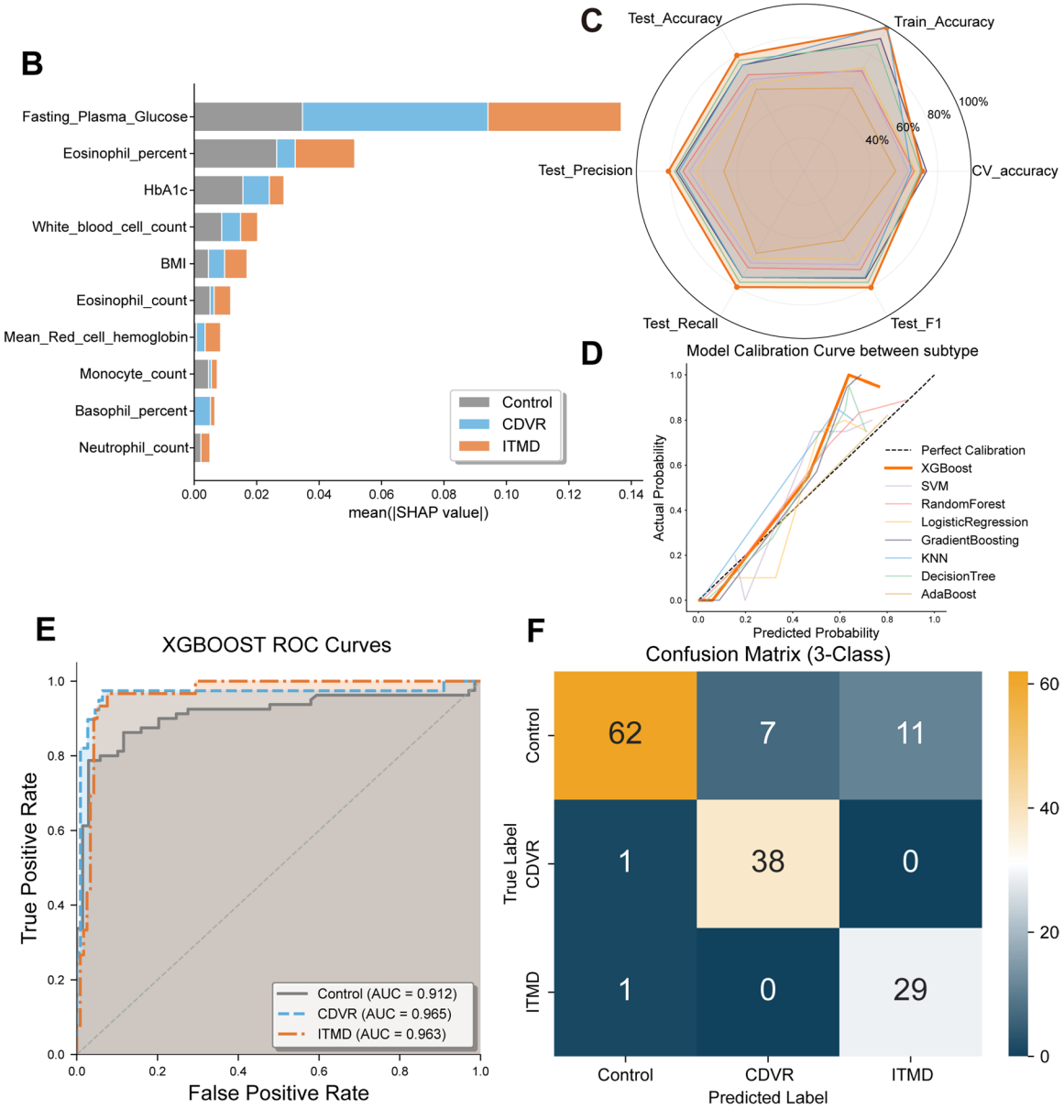


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Comparative performance of machine learning models in classifying poor CCC phenotypes. **A** Performance metrics for eight supervised learning models, including cross-validation accuracy (CV_accuracy), training accuracy, test accuracy, precision, recall, and F1 score. **B** SHAP analysis showing the mean contribution of top clinical features to the XGBoost classification model. **C** Radar chart comparing the performance metrics of all models. **D** Calibration curves assess the agreement between predicted probabilities (x-axis) and observed frequencies (y-axis) for eight machine learning algorithms. **E** ROC curves for XGBoost classification across the control, CDVR, and ITMD groups, with corresponding AUC values. **F** Confusion matrix of XGBoost predictions in the three-class classification task. CDVR: Driven Vascular Remodeling CDVR subtype, ITMD: Immuno-Thrombotic Myocardial Dysfunction subtype. control: good coronary collateral circulation

[43]. This finding suggests that eosinophils play a previously unappreciated role in regulating cardiac energy utilization and metabolic adaptation to stress. The precise interplay between eosinophilic inflammation and metabolic reprogramming in cardiomyocytes remains incompletely understood, but our data suggest it may be a crucial determinant of myocardial function in ITMD patients. The metabolic signature we observed indicates a shift toward increased oxidative metabolism, which may initially represent a compensatory response but could ultimately contribute to increased oxidative stress and cellular damage. The platelet-eosinophil axis represents another critical mechanism in ITMD pathophysiology. These interactions, mediated through adhesion molecules like P-selectin and chemokine release, function as potent amplifiers of microvascular thrombosis [44]. Marx et al. demonstrated that eosinophil-platelet interactions promote atherosclerosis and stabilize thrombosis through eosinophil extracellular traps [45], while Mackman highlighted the importance of eosinophils in atherosclerosis and thrombosis [46]. Jiang et al. confirmed significant eosinophil accumulation in thrombi and their interaction with platelets in acute coronary syndrome, providing direct clinical evidence for this mechanism [47]. Eosinophil granule proteins can activate coagulation cascades and promote platelet aggregation, potentially leading to microvascular obstruction, myocardial ischemia, and subsequent dysfunction [48]. This synergistic relationship between eosinophils and platelets creates a feed-forward cycle of inflammation and thrombosis that may be particularly relevant in ITMD. Patella et al. further demonstrated that eosinophil cationic proteins selectively activate human heart mast cells, which can affect vascular smooth muscle cell function through release of vasoactive mediators, potentially representing a crucial link between inflammation and microcirculatory dysfunction in ITMD patients [49].

Remarkably, our analysis revealed a paradoxical downregulation of HIF-1 and TGF-β signaling pathways in ITMD despite the chronic hypoxic environment. While these pathways are typically crucial for adaptive hypoxia-induced vascular remodeling and fibrosis, their suppressed activity in ITMD suggests a failure to mount an effective reparative response, potentially contributing to the observed myocardial dysfunction and impaired collateralization, rather than driving pathological remodeling. This counterintuitive finding can be explained by several potential mechanisms that warrant further investigation. Chronic or severe ischemic conditions may lead to a phenomenon of HIF-1 exhaustion or desensitization. Although HIF-1α protein may be stabilized initially, prolonged hypoxic stress can impair its transcriptional activity or downstream signaling cascade, resulting in a net downregulation of its functional output [50]. Ginouvès et al. demonstrated that chronic hypoxia can lead to prolyl hydroxylase domain protein (PHD) overactivation and subsequent "desensitization" of the HIF pathway [51]. Hölscher et al. reported that chronic cardiac HIF-1α stabilization can have unfavorable consequences, including metabolic alterations and impaired cardiac function [52]. This phenomenon may represent a maladaptive response that contributes to disease progression in ITMD patients. The unique inflammatory milieu of the ITMD subtype, characterized by prominent platelet activation and eosinophil interactions, may actively interfere with normal HIF-1 and TGF-β functions. Specifically, platelet-derived mediators can modulate hypoxia signaling pathways, as demonstrated by Chaurasia et al., who showed that platelets express HIF-2α that mediates prothrombotic responses under hypoxic conditions [53]. Eosinophil granule proteins and cytokines can directly or indirectly inhibit HIF-1 stabilization or alter TGF-β receptor signaling. Zhu et al. found that TGF-β signaling promotes eosinophil activation in inflammatory conditions [54], while Baek et al. demonstrated that hypoxia potentiates allergen-induced HIF-1α expression, which can increase eotaxin-1 and TGF-β1 expression, potentially creating a feedback loop that may become dysregulated in chronic conditions [55]. This interference aligns with our concept of "molecular mimicry," where vascular injury erroneously triggers immune programs designed for pathogen defense, leading to a dysfunctional rather than adaptive response [56, 57]. The parasitic response signature observed in ITMD likely reflects this molecular mimicry phenomenon, where the immune system misinterprets tissue damage signals as indicative of parasitic infection, deploying eosinophil-mediated responses that are inappropriate for the context of vascular injury. Excessive oxidative stress, prominent in chronic ischemia, inflammation, and metabolic dysregulation (as evidenced by the upregulation of oxidative phosphorylation in ITMD) can paradoxically lead to HIF-1α degradation despite hypoxic conditions. Pialoux et al. demonstrated that increased oxidative stress can affect HIF-1α mRNA levels during

hypoxic exposure [58], while Bae et al. showed the complex interplay between oxidative stress and HIF signaling [59]. This creates a vicious cycle where hypoxia-induced reactive oxygen species further impair the adaptive HIF-1 response. Zheng et al. found that impaired HIF-1 signaling contributes to excess production of mitochondrial reactive oxygen species [60], potentially exacerbating cellular damage and dysfunction. The chronic pathological state in ITMD may induce a cellular phenotype that is inherently less responsive to HIF-1 or TGF-β signaling, or where the downstream effects are diverted towards non-functional outcomes. This phenomenon has been observed in other chronic cardiovascular diseases where adaptive pathways become dysregulated over time, as reviewed by Semenza in the context of HIF-1 and cardiovascular disease [61] and revealed by Hrabalova et al., who demonstrated dysregulation of HIF-1α in cardiomyopathy [62].

The identification of these dual mechanisms-eosinophil-mediated cardiac dysfunction and paradoxical downregulation of adaptive signaling pathways has significant therapeutic implications for ITMD patients. Traditional approaches to cardiovascular disease may be insufficient for this distinct subtype, necessitating targeted interventions that address the specific pathophysiological mechanisms at play. Therapeutic strategies could focus on integrating anti-platelet therapy to counteract the heightened platelet activation with eosinophil-targeted agents (such as anti-IL-5 monoclonal antibodies like mepolizumab or reslizumab) to disrupt the inflammatory-thrombotic cycle affecting the myocardium. These biologics, which have shown efficacy in eosinophilic asthma and other eosinophilic disorders, might prove beneficial in reducing eosinophil-mediated cardiac damage in ITMD patients. Additionally, interventions aimed at restoring HIF-1 and TGF-β homeostatic signaling could potentially reverse the adaptive response failure observed in this subtype. Given the significant metabolic alterations and direct involvement of cardiac muscle contraction pathways, metabolic modulators (such as trimetazidine or ranolazine) and therapies specifically targeting myocardial function might prove particularly beneficial. The recognition of this eosinophil-dominated phenotype opens new avenues for personalized therapeutic approaches targeting the specific immune-cardiac interactions underlying myocardial dysfunction in these patients.

Early intervention with targeted anti-inflammatory therapy might prevent the development of irreversible cardiac structural changes, potentially improving long-term outcomes in this distinct patient population. Furthermore, biomarkers of eosinophil activation and HIF-1 signaling could serve as valuable tools for monitoring treatment response and disease progression. Future research should focus on validating these potential therapeutic targets in appropriate animal models and clinical studies, with the goal of developing personalized treatment strategies for ITMD patients based on their unique pathophysiological profile. The complex interplay between eosinophilic inflammation, thrombotic processes, and dysregulated adaptive responses in ITMD highlights the importance of integrated approaches to understanding and treating cardiovascular disease subtypes.

The clinical significance of this phenotype is highlighted by higher eosinophil percentages and worse MACE outcomes in ITMD patients, consistent with emerging evidence linking eosinophilic inflammation to adverse post-ischemic recovery [63]. This suggests that inappropriate activation of anti-parasite immune programs contributes to both poor collateralization and worse clinical outcomes. These distinct phenotypes offer targeted opportunities for precision interventions. For CDVR, combining intensive glycemic control with complement inhibition (e.g., C5 blockade) may attenuate vascular injury and enhance collateral development. For ITMD, integrating anti-platelet therapy with eosinophil-targeted agents (such as anti-IL-5 monoclonal antibodies) could disrupt the platelet-eosinophil inflammatory loop and modulate aberrant parasite-response pathways. These approaches move beyond one-size-fits-all strategies toward immunometabolic phenotype-guided interventions. Importantly, clinical translation may not require comprehensive proteomic profiling. In the MIMIC validation cohort, a parsimonious biomarker panel including fasting plasma glucose, eosinophil percentage, and HbA1c achieved high discrimination between phenotypes, suggesting feasibility for point-of-care implementation. This panel requires minimal additional testing beyond routine laboratory workups and could be integrated into existing clinical decision support systems to guide therapy selection, particularly for patients with metabolic comorbidities.

Strengths and limitations

Our study boasts several key strengths. Foremost is our proteomics-driven classification strategy, adopted after recognizing fundamental limitations in traditional approaches. Current clinical CCC Rentrop grading offers limited resolution for molecular heterogeneity within poor CCC patients, who may exhibit distinct pathophysiological mechanisms despite identical angiographic scores [17, 64, 65]. Similarly, conventional clinical parameters demonstrate limited discriminative power for poor CCC subtyping due to substantial feature overlap [66, 67]. This integrative approach, combining high-dimensional proteomics with comprehensive clinical data and advanced machine learning, robustly identified two

Table 2. Clinical comparison of poor CCC phenotypes in the discovery cohort.

Feature	Control (N=80)	CDVR (N=39)	ITMD (N=30)	CDVR vs ITMD	overall test
Demographics					
Age	58.91±10.51	61.79±9.84	58.77±10.21	0.22	0.31
BMI	27.68[24.82-30.68]	27.76[24.82-29.88]	27.32[25.26-30.53]	1.00	0.91
Complete Blood Count					
White_blood_cell_count	7.43[6.00-8.62]	7.22[6.40-8.97]	6.73[5.86-7.83]	0.15	0.38
Red_blood_cell_count	4.68[4.43-5.03]	4.78[4.41-5.05]	4.74[4.46-5.14]	0.89	0.79
Platelet_count	244.00[209.25-285.75]	246.00[189.00-298.50]	239.50[206.75-276.75]	0.61	0.81
Hemoglobin	141.50[133.00-151.00]	144.00[133.50-154.00]	141.50[131.50-148.00]	0.51	0.70
Hematocrit	0.43[0.40-0.46]	0.43[0.41-0.47]	0.42[0.40-0.46]	0.66	0.87
White Blood Cell Differential					
Neutrophil_count	4.59[3.74-5.82]	4.74[4.02-5.52]	4.23[3.53-4.95]	*	0.16
Neutrophil_percent	65.12±8.27	65.04±7.19	62.96±8.33	0.27	0.43
Lymphocyte_count	1.79[1.53-2.32]	1.90[1.42-2.30]	1.77[1.40-2.28]	0.76	0.86
Lymphocyte_percent	25.55±6.98	25.03±6.37	27.07±8.25	0.25	0.47
Monocyte_count	0.49[0.36-0.62]	0.54[0.41-0.64]	0.46[0.36-0.59]	0.22	0.46
Monocyte_percent	6.70[5.38-8.00]	6.90[5.80-8.00]	6.90[5.93-7.67]	0.98	0.92
Eosinophil_count	0.14[0.07-0.22]	0.14[0.08-0.22]	0.20[0.12-0.22]	0.15	0.27
Eosinophil_percent	1.90[1.10-2.82]	1.90[1.00-2.65]	2.85[1.55-3.30]	*	0.06
Basophil_count	0.03[0.02-0.04]	0.03[0.02-0.04]	0.03[0.02-0.04]	0.79	0.66
Basophil_percent	0.40[0.30-0.60]	0.40[0.25-0.60]	0.40[0.30-0.70]	0.32	0.48
Red Blood Cell Morphology					
Red_cell_distribution_width-CV	13.20[12.90-13.60]	13.10[12.85-13.80]	13.15[12.75-13.88]	0.78	0.90
Mean_Red_cell_volume	92.00[88.70-94.70]	92.60[89.85-93.95]	89.70[88.45-93.50]	0.19	0.39
Mean_Red_cell_hemoglobin	30.10[29.40-31.30]	30.40[29.30-31.55]	29.65[29.10-30.68]	0.14	0.29
Mean_Red_cell_hemoglobin_concentration	327.00[321.00-334.00]	330.00[323.00-339.50]	327.50[322.75-333.00]	0.38	0.49
Lipid Profile					
Total_Cholesterol	3.96[3.42-4.64]	3.97[3.35-4.50]	3.75[3.58-4.22]	0.83	0.74
Triglyceride	1.66[1.26-2.42]	1.52[0.84-1.92]	1.69[1.34-2.06]	0.19	0.22
High_density_lipoprotein_cholesterol	0.99[0.86-1.12]	1.05[0.94-1.17]	0.94[0.84-1.12]	0.15	0.31
Low_density_lipoprotein_cholesterol	2.38[1.81-2.78]	2.20[1.76-2.79]	2.16[1.86-2.77]	1.00	0.89
Glucose Metabolism					
Fasting_Plasma_Glucose	5.56[4.91-5.56]	7.82[7.69-7.82]	5.44[4.56-5.44]	***	***
HbA1c	7.05[6.17-7.05]	7.46[6.15-7.46]	6.76[5.90-7.09]	*	*
Renal Function					
Creatinine	78.35[66.83-87.67]	76.20[66.05-87.30]	73.80[67.55-81.97]	0.89	0.87
BUN	6.37[5.24-7.53]	6.71[5.47-7.78]	6.36[5.69-7.48]	0.50	0.67
Inflammatory & Cardiovascular Risk					
AISI	293.88[193.71-490.56]	328.00[234.48-455.40]	243.76[162.29-404.32]	0.16	0.33
Heart rate	78.82±10.24	79.44±10.17	78.43±11.05	0.70	0.92

Table 2 (continued)

Comparison of demographic, haematological, biochemical, and cardiovascular risk parameters among controls, CDVR, and ITMD subtypes. Data are presented as median (interquartile range) or mean ± SD. Given the non-normal distribution of the data, nonparametric tests were employed: the Kruskal-Wallis H test was used for overall differences among the three groups, and the Mann-Whitney U test was used for pairwise comparisons between CDVR and ITMD. P-values are reported for both overall and pairwise comparisons. Statistical significance is denoted as: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Key findings include significantly higher HbA1c in the CDVR group compared to ITMD ($p < 0.05$), and significantly elevated fasting plasma glucose in the CDVR group compared to both Control and ITMD ($p < 0.001$). The ITMD group also showed a numerically higher eosinophil percentage (2.85) compared to the other two groups, although this difference did not reach statistical significance ($p = 0.06$). CDVR: Driven Vascular Remodeling subtype; ITMD: Immuno-Thrombotic Myocardial Dysfunction subtype; Control: patients with good coronary collateral circulation. Data are presented as median (interquartile range) or mean ± SD, with P values indicated for intergroup and pairwise comparisons in the figure legend of Table 2

distinct immuno-metabolic phenotypes. The dimensional advantage of plasma proteomics (> 3000 quantified proteins versus 46 clinical variables) provides significantly higher molecular information, enabling unbiased discovery of biologically meaningful subtypes concealed by traditional methods. This "omics-driven classification" strategy has been successfully applied in cardiovascular research [68, 69], effectively capturing the integrated output of genetic, environmental, and pathophysiological factors [70].

The phenotypes' reproducibility was rigorously confirmed through external validation using the heterogeneous MIMIC cohort, enhancing generalizability and robustness. Crucially, a thorough comparison of medication usage between CDVR and ITMD clusters revealed no statistically significant differences across standard cardiovascular medications (Table S6). This suggests the observed proteomic and clinical distinctions are not confounded by pharmaceutical interventions, reinforcing their biological distinctiveness.

Our dual-axis framework provides mechanistic clarity through detailed KEGG enrichment analysis, gene expression validation, and correlation networks, offering deep insights into the underlying biological pathways. Specifically, distinct patterns of upregulated and downregulated pathways in CDVR (e.g., complement activation, AGE-RAGE) versus ITMD (e.g., platelet activation, cardiac dysfunction) provide a granular understanding of their unique pathomechanisms.

Furthermore, the prognostic implications of these phenotypes, demonstrated through differential MACE outcomes, underscore their clinical relevance for risk stratification. SHAP analysis identified key contributing features, informing the development of a parsimonious biomarker panel (fasting plasma glucose, eosinophil percentage, HbA1c) with high discrimination, promising feasible and cost-effective clinical translation for personalized medicine. Multivariate Cox regression analysis further elucidated the independent prognostic value of specific features within each cluster, highlighting, for instance, eosinophil percentage as a significant risk factor for MACE (HR > 1) in ITMD, and neutrophil count as a risk factor (HR > 1) in CDVR. This robust statistical validation strengthens the clinical utility of our identified phenotypes and completes our translational loop from molecular mechanisms to clinical practice for understanding CCC heterogeneity.

Our analytical strategy follows a comprehensive translational framework: (1) using unbiased high-dimensional proteomics for molecular subtype discovery; (2) validating these subtypes clinically and prognostically; and (3) developing a clinically feasible machine learning model using readily available clinical parameters. This approach balances biological discovery with clinical applicability, representing a significant strength.

Despite these strengths, several limitations of the current study warrant consideration. Firstly, the relatively modest size of our discovery cohort might have limited the statistical power to detect less abundant proteomic features. Secondly, our external validation relied on clinical surrogates rather than direct proteomic measurements, which could potentially overlook specific phenotype-defining proteins. Thirdly, as an observational study, it cannot definitively establish causality. Future research should include targeted functional assays to confirm the proposed mechanisms, such as evaluating complement inhibition in diabetic ischemia models or platelet-eosinophil blockade in hypoxic vascular models. Fourthly, the relatively short follow-up period of less than one year might lead to an underestimation of long-term MACE incidence and limit the assessment of sustained therapeutic effects, potentially introducing attrition bias. Additionally, the temporal stability of these identified phenotypes and their response to therapeutic interventions remain unexplored, representing an important area for future investigation. Finally, while the MIMIC cohort provided robust external validation, further validation in diverse populations and ethnicities would strengthen the generalizability of these findings.

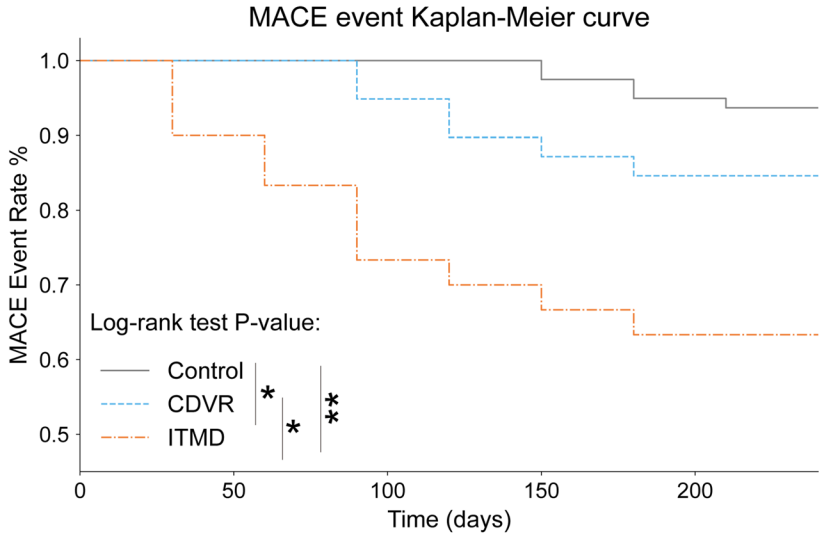
In conclusion, we propose a mechanistic classification for poor collateral circulation that identifies two robust and biologically plausible phenotypes: Complement-Driven Vascular Remodeling (CDVR) and Immuno-Thrombotic Myocardial Dysfunction (ITMD). While both converge on maladaptive immune-metabolic interactions, they differ fundamentally in their immunologic effectors and therapeutic targets. This framework provides a foundation for personalized vascular medicine, particularly in the context of diabetes and metabolic dysregulation, and offers new insights into the

A

Mimic Validation

feature	Control (N=300)	CDVR (N=1354)	ITMD (N=354)	CDVR vs ITMD	overall test
Demographics					
BMI	29.18[25.27-31.54]	29.46[29.46-29.46]	29.46[26.27-29.46]	***	***
Complete Blood Count					
White_blood_cell_count	7.40[5.67-10.03]	8.20[6.60-10.30]	6.85[5.70-8.47]	***	***
White Blood Cell Differential					
Neutrophil_count	6.77[5.15-7.96]	6.77[5.00-6.77]	6.77[4.46-6.77]	0.07	*
Monocyte_count	0.71[0.59-0.75]	0.71[0.59-0.73]	0.71[0.56-0.71]	*	*
Eosinophil_count	0.17[0.08-0.17]	0.17[0.10-0.17]	0.17[0.12-0.17]	0.21	*
Eosinophil_percent	1.50[0.80-2.18]	1.60[0.70-2.50]	3.10[1.80-4.20]	***	***
Basophil_percent	0.40[0.20-0.60]	0.50[0.30-0.60]	0.60[0.30-0.90]	***	***
Red Blood Cell Morphology					
Mean_Red_cell_hemoglobin	30.45[29.10-31.70]	30.30[29.00-31.50]	30.75[29.30-32.08]	***	***
Glucose Metabolism					
Fasting_Plasma_Glucose	6.06[5.61-6.44]	7.96[7.96-8.17]	5.56[5.28-7.96]	***	***
HbA1c	6.20[5.70-6.63]	6.63[5.90-6.90]	5.90[5.60-6.63]	***	***

B



C

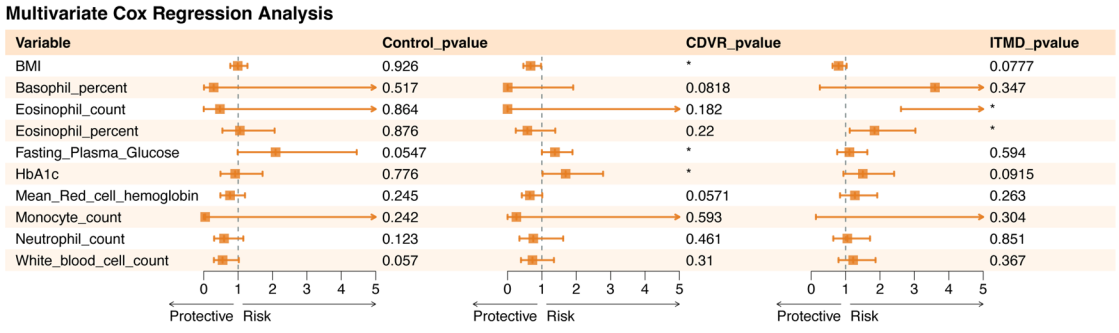


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 External validation (MIMIC database) and prognostic implications of poor CCC phenotypes. **A** Clinical feature comparison in the MIMIC validation cohort, confirming phenotype-specific differences in BMI, inflammatory cell counts, eosinophil and basophil percentages, mean red cell haemoglobin, fasting plasma glucose, and HbA1c. **B** Kaplan–Meier curves showing MACE incidence across groups; ITMD displayed the highest risk. **(C)** Multivariate Cox regression forest plots showing hazard ratios and 95% confidence intervals for clinical variables in control, CDVR, and ITMD subtypes. CDVR: Driven Vascular Remodeling CDVR subtype, ITMD: Immuno-Thrombotic Myocardial Dysfunction subtype. control: good coronary collateral circulation

heterogeneous nature of collateral vessel development. By targeting the specific immune-metabolic pathways driving poor collateralization in individual patients, this approach may ultimately improve outcomes in ischemic vascular disease.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12933-025-02988-y.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
AcknowledgmentsWe sincerely thank Dr. Dong Cong (Guangzhou, China) for the precious suggestion for data analysis.

Author contributions
The study was conceptualized by Y.N.Y. The methodology was developed and designed by Z.T.G. Investigation and experimental procedures were conducted by Z.T.G., H.M.L., R.X.H., and Y.J.Q. Formal analysis and statistical interpretation of data were performed by Z.T.G., H.M.L., R.X.H., Y.J.Q., and J.T. Data visualization and figure preparation were carried out by J.T. The original draft of the manuscript was written by J.T. Critical review and editing of the manuscript were provided by X.L.Y. The study was supervised by Z.T.G., who also handled project administration. Funding for this research was acquired by Y.N.Y., who serves as the guarantor of this work. All authors reviewed the manuscript and approved the final version for publication.

Funding
The study was supported in part by grants from the National Natural Science. Foundation of China (82260073);Tianshan Talent Cultivation Program Project of Xinjiang Uygur Autonomous Region (2022TSYCLJ0028)and "Tianshan Talents" medical and health high-level personnel training program (TSYC202401B021).

Data availability
The proteomics data matrix included in the study are available in the supplementary materials, raw data can be achieved from the corresponding author upon reasonable request.

Declarations

Competing interests
The authors declare no competing interests.

Received: 18 August 2025 / Accepted: 18 October 2025
Published online: 18 November 2025

References

1. Brilakis ES, Mashayekhi K, Tsuchikane E, Abi Rafeh N, Alaswad K, Araya M, et al. Guiding principles for chronic total occlusion percutaneous coronary intervention. Circulation. 2019;140(5):420–33.
2. Seiler C. The human coronary collateral circulation. Eur J Clin Invest. 2010;40(5):465–76.
3. Meier P, Hemingway H, Lansky AJ, Knapp G, Pitt B, Seiler C. The impact of the coronary collateral circulation on mortality: a meta-analysis. Eur Heart J. 2012;33(5):614–21.
4. Habib GB, Heibig J, Forman SA, Brown BG, Roberts R, Terrin ML, et al. Influence of coronary collateral vessels on myocardial infarct size in humans. Results of phase I thrombolysis in myocardial infarction (TIMI) trial. The TIMI investigators. Circulation. 1991;83(3):739–46.
5. Jones DA, Weerackody R, Rathod K, Behar J, Gallagher S, Knight CJ, et al. Successful recanalization of chronic total occlusions is associated with improved long-term survival. JACC Cardiovasc Interv. 2012;5(4):380–8.
6. Schaper W, Scholz D. Factors regulating arteriogenesis. Arterioscler Thromb Vasc Biol. 2003;23(7):1143–51.
7. van Royen N, Piek JJ, Buschmann I, Hoefer I, Voskuil M, Schaper W. Stimulation of arteriogenesis; a new concept for the treatment of arterial occlusive disease. Cardiovasc Res. 2001;49(3):543–53.
8. Abaci A, Oğuzhan A, Kahraman S, Eryol NK, Unal S, Arinç H, et al. Effect of diabetes mellitus on formation of coronary collateral vessels. Circulation. 1999;99(17):2239–42.
9. Xu S, Ilyas I, Little PJ, Li H, Kamato D, Zheng X, et al. Endothelial dysfunction in atherosclerotic cardiovascular diseases and beyond: from mechanism to pharmacotherapies. Pharmacol Rev. 2021;73(3):924–67.
10. Hinkel R, Howe A, Renner S, Ng J, Lee S, Klett K, et al. Diabetes mellitus-induced microvascular destabilization in the myocardium. J Am Coll Cardiol. 2017;69(2):131–43.
11. Waltenberger J. Impaired collateral vessel development in diabetes: potential cellular mechanisms and therapeutic implications. Cardiovasc Res. 2001;49(3):554–60.
12. Chaudhuri J, Bains Y, Guha S, Kahn A, Hall D, Bose N, et al. The role of advanced glycation end products in aging and metabolic diseases: bridging association and causality. Cell Metab. 2018;28(3):337–52.
13. Mao LS, Geng L, Wang YX, Qi Y, Wang MH, Ding FH, et al. Clinical risk score to predict poor coronary collateralization in type 2 diabetic patients with chronic total occlusion. BMC Cardiovasc Disord. 2025;25(1):250.
14. Demir K, Avci A, Altunkeser BB, Yilmaz A, Keles F, Ersecgin A. The relation between neutrophil-to-lymphocyte ratio and coronary chronic total occlusions. BMC Cardiovasc Disord. 2014;14:130.
15. Li C, Zhang F, Shen Y, Xu R, Chen Z, Dai Y, et al. Impact of neutrophil to lymphocyte ratio (NLR) index and its periprocedural change (NLR(Δ)) for percutaneous coronary intervention in patients with chronic total occlusion. Angiology. 2017;68(7):640–6.
16. Zimarino M, D'Andreamatteo M, Waksman R, Epstein SE, De Caterina R. The dynamics of the coronary collateral circulation. Nat Rev Cardiol. 2014;11(4):191–7.
17. Faber JE, Chilian WM, Deindl E, van Royen N, Simons M. A brief etymology of the collateral circulation. Arterioscler Thromb Vasc Biol. 2014;34(9):1854–9.
18. Ong SB, Hernández-Reséndiz S, Crespo-Avilan GE, Mukhametshina RT, Kwek XY, Cabrera-Fuentes HA, et al. Inflammation following acute myocardial infarction: Multiple players, dynamic roles, and novel therapeutic opportunities. Pharmacol Ther. 2018;186:73–87.
19. Rentrop KP, Cohen M, Blanke H, Phillips RA. Changes in collateral channel filling immediately after controlled coronary artery occlusion by an angioplasty balloon in human subjects. J Am Coll Cardiol. 1985;5(3):587–92.
20. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Rev. ed. New York: Lawrence Erlbaum Associates; 1988
21. Bacchetti P, Leung JM. Sample size calculations in clinical research. Anesthesiology. 2002;97(4):1028–9; author reply 9–32.
22. Oğuz D, Atmaca Y, Ozdöl C, Ozdemir AO, Kaya CT, Erol C. The relationship between coronary collateral artery development and inflammatory markers. Anadolu Kardiyol Derg. 2014;14(4):336–41.
23. Myszko M, Bychowski J, Skrzydlewska E, Łuczaj W. The dual role of oxidative stress in atherosclerosis and coronary artery disease: pathological mechanisms and diagnostic potential. Antioxidants. 2025;14(3):275.
24. Cochain C, Zernecke A. Protective and pathogenic roles of CD8(+) T cells in atherosclerosis. Basic Res Cardiol. 2016;111(6):71.

25. Ghosh P, Sahoo R, Vaidya A, Chorev M, Halperin JA. Role of complement and complement regulatory proteins in the complications of diabetes. Endocr Rev. 2015;36(3):272–88.
26. Qin X, Goldfine A, Krumrei N, Grubissich L, Acosta J, Chorev M, et al. Glycation inactivation of the complement regulatory protein CD59: a possible role in the pathogenesis of the vascular complications of human diabetes. Diabetes. 2004;53(10):2653–61.
27. Flyvbjerg A. The role of the complement system in diabetic nephropathy. Nat Rev Nephrol. 2017;13(5):311–8.
28. Zoungas S, Arima H, Gerstein HC, Holman RR, Woodward M, Reaven P, et al. Effects of intensive glucose control on microvascular outcomes in patients with type 2 diabetes: a meta-analysis of individual participant data from randomised controlled trials. Lancet Diabetes Endocrinol. 2017;5(6):431–7.
29. Stitt AW, Curtis TM, Chen M, Medina RJ, McKay GJ, Jenkins A, et al. The progress in understanding and treatment of diabetic retinopathy. Prog Retin Eye Res. 2016;51:156–86.
30. Zeineddine HA, Hong SH, Peesh P, Dienel A, Torres K, Thankamani Pandit P, et al. Neutrophils and neutrophil extracellular traps cause vascular occlusion and delayed cerebral ischemia after subarachnoid hemorrhage in mice. Arterioscler Thromb Vasc Biol. 2024;44(3):635–52.
31. Kang L, Yu H, Yang X, Zhu Y, Bai X, Wang R, et al. Neutrophil extracellular traps released by neutrophils impair revascularization and vascular remodeling after stroke. Nat Commun. 2020;11(1):2488.
32. Yu S, Liu J, Yan N. Endothelial dysfunction induced by extracellular neutrophil traps plays important role in the occurrence and treatment of extracellular neutrophil traps-related disease. Int J Mol Sci. 2022;23(10):5626.
33. Rowan S, Bejarano E, Taylor A. Mechanistic targeting of advanced glycation end-products in age-related diseases. Biochim Biophys Acta Mol Basis Dis. 2018;1864(12):3631–43.
34. Zhou M, Zhang Y, Shi L, Li L, Zhang D, Gong Z, et al. Activation and modulation of the AGEs-RAGE axis: implications for inflammatory pathologies and therapeutic interventions - a review. Pharmacol Res. 2024;206:107282.
35. Kologrivova I, Shtatolkina M, Suslova T, Ryabov V. Cells of the immune system in cardiac remodeling: main players in resolution of inflammation and repair after myocardial infarction. Front Immunol. 2021;12:664457.
36. Toor IS, Rückerl D, Mair I, Ainsworth R, Meloni M, Spiroski AM, et al. Eosinophil deficiency promotes aberrant repair and adverse remodeling following acute myocardial infarction. JACC Basic Transl Sci. 2020;5(7):665–81.
37. Rothenberg ME, Hogan SP. The eosinophil. Annu Rev Immunol. 2006;24:147–74.
38. Spry CJ, Tai PC, Davies J. The cardiotoxicity of eosinophils. Postgrad Med J. 1983;59(689):147–53.
39. deMello DE, Liapis H, Jureidini S, Nouri S, Kephart GM, Gleich GJ. Cardiac localization of eosinophil-granule major basic protein in acute necrotizing myocarditis. N Engl J Med. 1990;323(22):1542–5.
40. Tai PC, Ackerman SJ, Spry CJ, Dunnette S, Olsen EG, Gleich GJ. Deposits of eosinophil granule proteins in cardiac tissues of patients with eosinophilic endomyocardial disease. Lancet. 1987;1(8534):643–7.
41. Diny NL, Baldeviano GC, Talor MV, Barin JG, Ong S, Bedja D, et al. Eosinophil-derived IL-4 drives progression of myocarditis to inflammatory dilated cardiomyopathy. J Exp Med. 2017;214(4):943–57.
42. Liu J, Yang C, Liu T, Deng Z, Fang W, Zhang X, et al. Eosinophils improve cardiac function after myocardial infarction. Nat Commun. 2020;11(1):6396.
43. Yang C, Li J, Deng Z, Luo S, Liu J, Fang W, et al. Eosinophils protect pressure overload- and β-adrenoreceptor agonist-induced cardiac hypertrophy. Cardiovasc Res. 2023;119(1):195–212.
44. Ramirez GA, Yacoub MR, Ripa M, Mannina D, Cariddi A, Saporiti N, et al. Eosinophils from physiology to disease: a comprehensive review. BioMed Res Int. 2018;2018:9095275.
45. Marx C, Novotny J, Salbeck D, Zellner KR, Nicolai L, Pekayvaz K, et al. Eosinophil-platelet interactions promote atherosclerosis and stabilize thrombosis with eosinophil extracellular traps. Blood. 2019;134(21):1859–72.
46. Mackman N. Eosinophils, atherosclerosis, and thrombosis. Blood. 2019;134(21):1781–2.
47. Jiang P, Wang DZ, Ren YL, Cai JP, Chen BX. Significance of eosinophil accumulation in the thrombus and decrease in peripheral blood in patients with acute coronary syndrome. Coron Artery Dis. 2015;26(2):101–6.
48. Shah SA, Page CP, Pitchford SC. Platelet-eosinophil interactions as a potential therapeutic target in allergic inflammation and asthma. Front Med. 2017;4:129.
49. Patella V, de Crescenzo G, Marino I, Genovese A, Adt M, Gleich GJ, et al. Eosinophil granule proteins are selective activators of human heart mast cells. Int Arch Allergy Immunol. 1997;113(1–3):200–2.
50. Fong GH, Takeda K. Role and regulation of prolyl hydroxylase domain proteins. Cell Death Differ. 2008;15(4):635–41.
51. Ginouvès A, Ilc K, Macías N, Pouysségur J, Berra E. PHDs overactivation during chronic hypoxia "desensitizes" HIFalpha and protects cells from necrosis. Proc Natl Acad Sci USA. 2008;105(12):4745–50.
52. Hölscher M, Silter M, Krull S, von Ahlen M, Hesse A, Schwartz P, et al. Cardiomyocyte-specific prolyl-4-hydroxylase domain 2 knock out protects from acute myocardial ischemic injury. J Biol Chem. 2011;286(13):11185–94.
53. Chaurasia SN, Kushwaha G, Kulkarni PP, Mallick RL, Latheef NA, Mishra JK, et al. Platelet HIF-2α promotes thrombogenicity through PAI-1 synthesis and extracellular vesicle release. Haematologica. 2019;104(12):2482–92.
54. Zhu Z, Homer RJ, Wang Z, Chen Q, Geba GP, Wang J, et al. Pulmonary expression of interleukin-13 causes inflammation, mucus hypersecretion, subepithelial fibrosis, physiologic abnormalities, and eotaxin production. J Clin Invest. 1999;103(6):779–88.
55. Baek KJ, Cho JY, Rosenthal P, Alexander LEC, Nizet V, Broide DH. Hypoxia potentiates allergen induction of HIF-1α, chemokines, airway inflammation, TGF-β1, and airway remodeling in a mouse model. Clin Immunol. 2013;147(1):27–37.
56. Rojas M, Restrepo-Jiménez P, Monsalve DM, Pacheco Y, Acosta-Ampudia Y, Ramírez-Santana C, et al. Molecular mimicry and autoimmunity. J Autoimmun. 2018;95:100–23.
57. Fujinami RS, von Herrath MG, Christen U, Whitton JL. Molecular mimicry, bystander activation, or viral persistence: infections and autoimmune disease. Clin Microbiol Rev. 2006;19(1):80–94.
58. Pialoux V, Mounier R, Brown AD, Steinback CD, Rawling JM, Poulin MJ. Relationship between oxidative stress and HIF-1 alpha mRNA during sustained hypoxia in humans. Free Radic Biol Med. 2009;46(2):321–6.
59. Bae YS, Kang SW, Seo MS, Baines IC, Tekle E, Chock PB, et al. Epidermal growth factor (EGF)-induced generation of hydrogen peroxide. Role in EGF receptor-mediated tyrosine phosphorylation. J Biol Chem. 1997;272(1):217–21.
60. Zheng X, Ruas JL, Cao R, Salomons FA, Cao Y, Poellinger L, et al. Cell-type-specific regulation of degradation of hypoxia-inducible factor 1 alpha: role of subcellular compartmentalization. Mol Cell Biol. 2006;26(12):4628–41.
61. Semenza GL. Hypoxia-inducible factor 1 and cardiovascular disease. Annu Rev Physiol. 2014;76:39–56.
62. Hrabalova P, Bohuslavova R, Matejkova K, Papousek F, Sedmera D, Abaffy P, et al. Dysregulation of hypoxia-inducible factor 1α in the sympathetic nervous system accelerates diabetic cardiomyopathy. Cardiovasc Diabetol. 2023;22(1):88.
63. Xu J, Guo J, Liu T, Yang C, Meng Z, Libby P, et al. Differential roles of eosinophils in cardiovascular disease. Nat Rev Cardiol. 2025;22(3):165–82.
64. Chillo O, Kleinert EC, Lautz T, Lasch M, Pagel JI, Heun Y, et al. Perivascular mast cells govern shear stress-induced arteriogenesis by orchestrating leukocyte function. Cell Rep. 2016;16(8):2197–207.
65. Eltzschig HK, Carmeliet P. Hypoxia and inflammation. N Engl J Med. 2011;364(7):656–65.
66. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet. 2019;394(10201):861–7.
67. Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920–30.
68. Langley SR, Willeit K, Didangelos A, Matic LP, Skroblin P, Barallobre-Barreiro J, et al. Extracellular matrix proteomics identifies molecular signature of symptomatic carotid plaques. J Clin Invest. 2017;127(4):1546–60.
69. Ekhzaimy AA, Masood A, Benabdelkamel H, Elhassan T, Musambil M, Alfadda AA. Plasma proteomics reveals an improved cardio-metabolic profile in patients with type 2 diabetes post-liraglutide treatment. Diab Vasc Dis Res. 2022;19(3):14791641221094322.
70. Lindsey ML, Mayr M, Gomes AV, Delles C, Arrell DK, Murphy AM, et al. Transformative impact of proteomics on cardiovascular health and disease: a scientific statement from the American Heart Association. Circulation. 2015;132(9):852–72.

Publisher's note