Supplementary Methods

Data Acquisition and Preprocessing

All analyses were conducted in R (version 4.4.3) and the analysis code is available at GitHub (https://github.com/Yizhou-Y/Supplementary-File_Code). Author processed series matrix files were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) using the GEOquery package, including the following peripheral blood samples: 944 from CHF patients enrolled in the BIOSTAT-CHF study (GSE181114, Affymetrix Human Transcriptome Array 2.0, GPL30456), 11 from normal individuals (GSE213313, Agilent SurePrint G3 Human GE v3 8×60K, GPL21185), and 197 from CHF patients (GSE77343, Affymetrix Human Gene 1.1 ST, GPL11532), for which complete blood count differentials measured at the time of blood collection are also deposited. This cohort was used in the feature screening step and for the independent cohort validation described below. All three datasets were generated on oligonucleotide microarrays. RNA for GSE77343 was extracted from PAXgene blood, and the collection tube is not reported in the GEO records for GSE181114 or GSE213313. Probe identifiers were mapped to gene symbols using the Ensembl GRCh38.104 annotation for GSE181114 and the platform annotation files for GSE213313 and GSE77343. Where several probes mapped to the same gene symbol the first was retained, and probes without an assigned symbol were discarded. As deposited, GSE181114 had been normalized with the IRON method, GSE213313 had been processed with limma, with control, outlier and below background probes removed and duplicate probes merged by avereps, and GSE77343 had been RMA

normalized and ComBat corrected by the original authors. GSE181114 and GSE213313 were further normalized with limma::normalizeBetweenArrays, and GSE77343 was used as deposited. For GSE213313, normalization was performed across all 94 samples of the series, and the 11 normal samples were extracted afterwards. The GSE181114 and GSE213313 matrices were then merged on shared gene symbols and limma::removeBatchEffect(expr, batch = dataset, design = model.matrix(~tissue)) was applied, where dataset distinguishes the two series and tissue denotes disease group. Because these two factors partition the samples identically, the augmented design has rank two for three columns, the batch coefficient was reported as not estimable for all 11,964 shared genes, and the returned matrix was identical to the input at all 11,425,620 entries. The merged matrix was therefore used uncorrected, and this comparison was used only to generate the candidate gene set. The table below lists the demographic and clinical fields carried by each deposit.

Characteristics and available metadata fields of the three datasets.

Characteristic	**Discovery cohort GSE181114 (n = 944)**	**Normal controls GSE213313 (n = 11)**	**Independent cohort GSE77343 (n = 197)**
Samples, n	944	11	197
Platform	Affymetrix HTA 2.0 (GPL30456)	Agilent SurePrint G3 8x60K (GPL21185)	Affymetrix HuGene 1.1 ST (GPL11532)
Processing as deposited	IRON normalization	limma, controls and below background probes removed, avereps	RMA with ComBat
Material	whole blood, PAXgene	whole blood, tube not reported	whole blood, PAXgene

Condition	chronic heart failure	healthy	chronic heart failure
Age, years, median (IQR)	73 (64 to 79)	not deposited	66 (59 to 74)
Age, range	34 to 92	not deposited	24 to 91
Male, n (%)	711 (75.3%)	not deposited	149 (75.6%)
Female, n (%)	233 (24.7%)	not deposited	48 (24.4%)
Neutrophil proportion, median (IQR)	not deposited	not deposited	0.662 (0.593 to 0.718)
Lymphocyte proportion, median (IQR)	not deposited	not deposited	0.197 (0.153 to 0.265)
Monocyte proportion, median (IQR)	not deposited	not deposited	0.088 (0.076 to 0.102)
Eosinophil proportion, median (IQR)	not deposited	not deposited	0.025 (0.015 to 0.042)
Basophil proportion, median (IQR)	not deposited	not deposited	0.007 (0.005 to 0.010)
Ejection fraction, NYHA class, comorbidity, medication, outcome	not deposited, collected in the parent study	not deposited	not deposited

Within the discovery cohort the median age by cluster is 72, 72, 75 and 73 years and the proportion of men is 74.3%, 81.9%, 74.6% and 69.2% for clusters 1 to 4, differing across clusters at p = 0.040 for age by the Kruskal-Wallis test and at p = 0.021 for sex by the chi-squared test.

Differential Expression Analysis

Differential expression analysis was conducted via the limma package with empirical Bayes moderation. Three sets of comparisons were performed: CHF cohort vs. normal cohort, individual CHF clusters vs. normal cohort, and pairwise comparisons among

CHF clusters. Differentially expressed genes (DEGs) were required to satisfy $p < 0.05$ and adjusted $p < 0.05$ in every comparison, with the fold change threshold set at |log2FC| > 2 for CHF vs. normal and for individual CHF clusters vs. normal and at |log2FC| > 0.5 for inter-cluster comparisons. The Venn diagrams were constructed using the VennDiagram package to illustrate overlaps in DEGs among clusters. Volcano plots were generated with the ggplot2 package, and heatmaps were created through the pheatmap package to visualize expression patterns. Adjusted p values were obtained with the Benjamini-Hochberg procedure throughout.

Functional Enrichment Analysis

Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA) were performed using the clusterProfiler package. GO and KEGG analyses were based on 2,350 of the 3,902 DEGs from the CHF vs. normal comparison selected at mean expression ≥ 5 on the $\log_2$ scale (approximately the median across all genes analyzed), with terms retained at $p < 0.05$, adjusted p < 0.05 and q < 0.05. enrichGO and enrichKEGG were called without a universe argument, so the background is the default annotation space of org.Hs.eg.db and of the human KEGG annotation rather than the genes measured on both platforms. Results were visualized using the circlize package. To characterize functional features of individual CHF clusters, DEGs with |log2FC| > 2 (clusters vs. normal) were analyzed and visualized using the ggplot2 package. GSEA was conducted with the GSEA

function of clusterProfiler on the DEGs from each cluster vs. normal comparison at the more permissive cut of |log2FC| > 1, $p < 0.05$ and adjusted $p < 0.05$, ranked by log2 fold change in descending order, so that the background is this selected list rather than the full ranked transcriptome. Gene sets were taken from the Molecular Signatures Database (MSigDB, version 2024.1.Hs, http://www.gsea-msigdb.org/gsea/index.jsp), using the H hallmark, C2CP canonical pathways, C4 computational and C7 ImmuneSigDB immunologic signature collections with Entrez identifiers. Gene set permutation was applied with nPermSimple = 1,000,000, gene sets were restricted to between 10 and 7,000 genes (minGSSize = 10, maxGSSize = 7,000), and significant terms were defined by |NES| > 1, $p < 0.05$, and adjusted $p < 0.05$ with the Benjamini-Hochberg procedure. Every enriched term reported in the manuscript satisfies an adjusted p value and a q value below 0.05.

Weighted Gene Co-expression Network Analysis (WGCNA)

WGCNA was performed using the WGCNA package, with all genes from the merged expression matrix included for network construction. Hierarchical clustering of all 955 samples was conducted to detect potential outliers, using a dendrogram cut-off height of 271; no samples were excluded. The optimal soft-thresholding power ($\beta = 3$) was determined using the pickSoftThreshold function with an R-squared threshold of 0.85, and the power was selected based on scale-free topology criteria. A minimum module size of 25 was set, and modules were identified using a dynamic tree-cutting algorithm.

Module eigengenes were further clustered with a merging threshold of 0.25 to define final module assignments. Among the resulting modules, the turquoise module, which contained the largest number of genes, was selected for further analysis due to its strong Pearson correlation with CHF status (correlation coefficient = 0.99, $p < 0.001$). A gene co-expression network was constructed for this module using the igraph package, applying an edge weight threshold of 0.575 to retain high-confidence interactions.

Integrative Machine Learning Strategy for Features Screening

Three machine learning algorithms were applied to select feature variables from the candidate genes: the Least Absolute Shrinkage and Selection Operator (LASSO), Random Forest (RF), and Support Vector Machine-Recursive Feature Elimination (SVM-RFE). All models were fitted on the merged matrix (n = 955, of which 98.85% are CHF samples), using an input gene set derived from the intersection of DEGs and genes in the WGCNA turquoise module, restricted to genes also present on the GSE77343 platform. Because the class sizes are extremely unbalanced, the three algorithms were used here only as a heuristic filter to rank and reduce the candidate set, and not as diagnostic classifiers. No class imbalance handling was applied, and preprocessing and feature selection were carried out once on the full matrix rather than repeated inside the cross-validation folds, so the cross-validation figures are not unbiased estimates of classification performance. Sensitivity was recorded for each algorithm, and precision is not informative in this design because GSE77343 contains

CHF samples only. Overlapping genes among models were visualized via a Venn diagram, and genes shared by any two methods were retained for downstream analysis.

LASSO logistic regression was implemented with the glmnet package. Input features were standardized, and 10-fold cross-validation was performed to determine the optimal regularization parameter (λ). Genes with nonzero coefficients at λ.min were selected. Model performance was assessed via coefficient paths and cross-validation error curves.

RF modeling was conducted using the randomForest package (ntree = 500). Feature importance was ranked using the Mean Decrease Accuracy metric. Genes with positive importance scores were retained, and the top 20 features were visualized. Out-of-bag (OOB) error rate and its trend across trees were also evaluated.

SVM-RFE was performed with the caret and e1071 packages. Feature elimination was guided by 10-fold cross-validation. The top 500 genes were selected based on the cross-validation accuracy curve. A support vector machine with a radial basis function (RBF) kernel was trained on this subset using 5-fold cross-validation, with hyperparameter tuning for C and sigma.

Protein-Protein Interaction (PPI) Network Construction and Markov Cluster Algorithm (MCL)-Based Module Detection

Overlapping genes from the three machine learning models were submitted to the

STRING database (https://string-db.org) to construct a PPI network with a minimum interaction score of 0.4. Genes with predicted or known interactions were retained, and MCL clustering was applied (inflation parameter = 2). One module comprising five genes was selected for further analysis, as all five were enriched in WP1772, the only annotated pathway in the enrichment network. Pearson correlation analysis was performed using the corrplot package to assess co-expression among all overlapping genes, as well as among the five genes in the selected module. The WP1772 pathway diagram was downloaded from WikiPathways (https://www.wikipathways.org), and DEGs were mapped onto the pathway.

Partitioning Around Medoids (PAM)-Based Clustering and Principal Component Analysis (PCA) Visualization of the CHF Cohort

The machine-learning-based PAM algorithm was applied to cluster the CHF cohort (n = 944) based on the expression profiles of the five MCL-identified key genes with the ConsensusClusterPlus package. Pearson correlation served as the distance metric, so the distance between two samples is one minus the correlation between their five expression values and the partition follows the profile across the five features rather than their absolute levels. The features were not standardized before this step, and ConsensusClusterPlus applies no centering or scaling of its own. Clustering used 100 resampling iterations (reps = 100, pItem = 0.8, pFeature = 1, seed = 123456). The optimal number of clusters (k = 4) was determined via cumulative distribution function (CDF) curves, delta area plots, and the elbow method.

PCA was performed on the merged gene expression matrix to visualize the sample distribution across PAM-defined clusters. The analysis was carried out using the FactoMineR package, and an interactive 3D plot was generated with the plotly package. Violin plots were created to assess the expression differences of the five MCL-identified key genes among clusters, and statistical significance was evaluated with the Wilcoxon test.

Independent Cohort Validation

The expression matrix of GSE77343 (197 CHF peripheral blood samples, Affymetrix Human Gene 1.1 ST) was obtained from GEO. All five features were present on the platform. Expression values for the five features were standardized within each cohort using the population standard deviation, so that the two platforms were never placed in a common matrix. This standardization was applied at this step, when the centroids and the two directions were computed, and not before consensus clustering.

Cluster centroids were computed as the mean standardized expression of each feature within each of the four discovery clusters. For CASP6, CAPNS1, BAK1, CASP4 and BAD the centroids are 0.0344, −0.0259, −0.4327, 0.0124 and 0.0995 for cluster 1, 0.2275, 0.9214, 0.0757, −1.0695 and −0.1184 for cluster 2, −0.1507, −0.0263, 1.4804, 0.4029 and −0.3788 for cluster 3, and −0.2619, −1.1634, −0.3334, 1.1082 and 0.2407 for cluster 4. Two unit vectors were derived, $\mathbf{w}_1$ from the cluster 2 minus cluster 4 centroid difference and $\mathbf{w}_2$ from the cluster 3 minus cluster 1 difference, and the cosine

between them was 0.002. Scores were obtained by projecting the standardized five feature matrix of each cohort onto these fixed vectors. Assignment of validation patients used the minimum Euclidean distance to the four centroids in the same standardized space. The cluster 2 versus 4 direction weighted *CASP6*, *CAPNS1*, *BAK1*, *CASP4* and *BAD* by 0.16, 0.67, 0.13, −0.70 and −0.12, and the cluster 3 versus 1 direction by −0.09, 0.00, 0.95, 0.19 and −0.24.

Immune composition in GSE181114 was taken from the CIBERSORTx output, with lymphocyte content defined as the sum of the twelve lymphoid populations of the LM22 matrix. In GSE77343 the corresponding quantities are the complete blood count differential proportions recorded in the sample characteristics fields. The two are not on a common scale, since the CIBERSORTx values are relative fractions of 22 leukocyte populations and the differential proportions are fractions of a five-part count; they were therefore standardized within cohort before the cluster medians were compared, and correlations were computed within each cohort only. Partial correlations were obtained from the residuals of linear models.

Immune Microenvironment Characterization

The immune landscapes of CHF and normal samples were characterized through CIBERSORTx (https://cibersortx.stanford.edu/) and single-sample Gene Set Enrichment Analysis (ssGSEA). CIBERSORTx was executed on the web platform with the LM22 leukocyte signature matrix, using the log2 expression matrix as input, 1,000

permutations, quantile normalization disabled and no batch correction mode. Relative abundances of 22 immune cell types were estimated by transcriptome deconvolution. All 955 samples were retained and none was filtered, the deconvolution p value reported by the platform being below 0.05 for every sample, and the values are relative fractions of the 22 populations that sum to one within each sample rather than proportions of a differential count. Repeating the run with the matrix on a linear scale returned the same estimates, with a median Pearson correlation of 0.994 across the 22 populations, a correlation of 0.9998 across the whole matrix, and all 955 samples passing in both runs. ssGSEA was performed with the GSVA package to calculate enrichment scores for predefined immune-related functional gene sets. These scores are on an arbitrary scale and are not cell proportions, and no absolute cell counts were available in either cohort. Comparative analyses between CHF and normal cohorts, as well as among CHF clusters, were visualized using bar plots, heatmaps, and violin plots. Statistical significance was assessed via the Wilcoxon test or Kruskal-Wallis test, as appropriate, and p values were adjusted across measures with the Benjamini-Hochberg procedure.

Immunohistochemistry and Semi-Quantitative Analysis

Immunohistochemistry (IHC) images of the five MCL-identified key genes were retrieved from HPA (https://www.proteinatlas.org/). Protein expression was examined in normal heart, lung, and breast tissues, as well as in breast invasive ductal carcinoma

(IDC), breast invasive lobular carcinoma (ILC), lung adenocarcinoma (LUAD), and lung squamous cell carcinoma (LUSC). Expression levels were qualitatively evaluated based on staining intensity and subcellular localization, as annotated by HPA.

For each gene and each patient, the staining intensity and the proportion of positively stained cells annotated by the Human Protein Atlas were converted to numeric values following the Remmele–Stegner immunoreactive scoring (IRS) system. The categories are the ones the atlas annotates, negative, weak, moderate and strong for intensity and none, under 25%, 25 to 75% and over 75% for proportion, so the proportion is scored on four levels rather than the five levels of the original score and the product ranges from 0 to 9 rather than from 0 to 12. The values were taken from that annotation rather than scored again, so no observer procedure or measure of inter observer agreement applies. The analysis covers 468 annotated images from 94 patients and 10 antibodies across the five genes, forming 70 combinations of gene, antibody and tissue, with one to three patients per combination in the normal tissues and two to nine in the tumors. Every antibody the atlas provides for each of the five genes was used and none was excluded.

The modified IRS was calculated according to the following formula:

$$IRS_j^{(g,o,a)} = \frac{1}{n_j}\sum_{i=1}^{n_j}(I_i^{(g,o,a)} \times Q_i^{(g,o,a)})$$

- $IRS_j^{(g,o,a)}$ is the mean modified IRS for patient j in gene g, tissue or disease o, and antibody a;

- $I_i^{(g,o,a)}$ represents the staining intensity (0 = negative, 1 = weak, 2 = moderate, 3 = strong);
- $Q_i^{(g,o,a)}$ denotes the proportion score of positive cells (0 = none, 1 = <25%, 2 = 25–75%, 3 = >75%);
- n_j is the number of valid IHC images available for that patient within the corresponding gene/tissue/antibody category.

If multiple representative cell types appeared in a single image (for example, alveolar type I and type II cells in lung tissue), the average score across those cell types was used for that image.

Only functional parenchymal or tumor cells were included in the evaluation: breast glandular cells, lung alveolar type I and type II cells, heart cardiomyocytes, and malignant epithelial cells in tumors.

Samples derived from lymph nodes or non-relevant tissues were excluded from the analysis.

Survival analysis in solid malignancies relevant to cardio-oncology (breast and lung cancer)

To explore the potential relevance of the selected molecular features beyond chronic heart failure, Kaplan–Meier survival analyses were performed using the KM Plotter database (https://kmplot.com/analysis/). OS was evaluated in breast cancer (n = 1879), LUSC (n = 780), and LUAD (n = 1161) cohorts. Lung cohorts were restricted by

histology to squamous cell carcinoma and adenocarcinoma, respectively, with no restriction applied for stage, grade, sex, smoking history, or treatment. For all analyses, only the JetSet best probe set was used for each gene, biased arrays were excluded by the platform's array quality control, and the median was computed over the entire database.

For individual gene analyses, patients were stratified into high- and low-expression groups according to the optimal cutoff automatically determined by the platform. Hazard ratios and corresponding 95% confidence intervals were estimated using univariate Cox proportional hazards models, and statistical significance was assessed using the log-rank test.

For combined analyses, the signature consisting of *CASP6, CAPNS1, BAK1, CASP4,* and *BAD* was generated using the mean expression values of these genes as provided by the platform, and patients were divided into high- and low-expression groups using the same automatically selected cutoff approach. This mean is unweighted and is therefore distinct from the weighted projections w_1 and w_2 used in the CHF cohort, whose weights derive from contrasts between CHF clusters and have no counterpart in these tumor cohorts. The log-rank p values from the 18 analyses, six in each of the three cohorts, were subsequently adjusted across the whole set with the Bonferroni correction.