

Multiple Plasma Biomarkers for Risk Stratification in Patients With Heart Failure and Preserved Ejection Fraction



Julio A. Chirinos, MD, PhD,^{a,b,*} Alena Orlenko, PhD,^{b,*} Lei Zhao, MD, PhD,^c Michael D. Basso, MS,^c Mary Ellen Cvijic, PhD,^c Zhuyin Li, PhD,^c Thomas E. Spires, MS,^c Melissa Yarde, MS,^c Zhaoqing Wang, MS,^c Dietmar A. Seiffert, MD,^c Stuart Prenner, MD,^{a,b} Payman Zamani, MD, MTR,^{a,b} Priyanka Bhattacharya, MD,^{a,b} Anupam Kumar, MD,^d Kenneth B. Margulies, MD,^{a,b} Bruce D. Car, PhD,^c David A. Gordon, PhD,^c Jason H. Moore, PhD,^{a,b} Thomas P. Cappola, MD, ScM^{a,b}

ABSTRACT

BACKGROUND Better risk stratification strategies are needed to enhance clinical care and trial design in heart failure with preserved ejection fraction (HFpEF).

OBJECTIVES The purpose of this study was to assess the value of a targeted plasma multi-marker approach to enhance our phenotypic characterization and risk prediction in HFpEF.

METHODS In this study, the authors measured 49 plasma biomarkers from TOPCAT (Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist) trial participants (n = 379) using a Multiplex assay. The relationship between biomarkers and the risk of all-cause death or heart failure-related hospital admission (DHFA) was assessed. A tree-based pipeline optimizer platform was used to generate a multimarker predictive model for DHFA. We validated the model in an independent cohort of HFpEF patients enrolled in the PHFS (Penn Heart Failure Study) (n = 156).

RESULTS Two large, tightly related dominant biomarker clusters were found, which included biomarkers of fibrosis/tissue remodeling, inflammation, renal injury/dysfunction, and liver fibrosis. Other clusters were composed of neuro-hormonal regulators of mineral metabolism, intermediary metabolism, and biomarkers of myocardial injury. Multiple biomarkers predicted incident DHFA, including 2 biomarkers related to mineral metabolism/calcification (fibroblast growth factor-23 and OPG [osteoprotegerin]), 3 inflammatory biomarkers (tumor necrosis factor- α , sTNFRI [soluble tumor necrosis factor-receptor I], and interleukin-6), YKL-40 (related to liver injury and inflammation), 2 biomarkers related to intermediary metabolism and adipocyte biology (fatty acid binding protein-4 and growth differentiation factor-15), angiopoietin-2 (related to angiogenesis), matrix metalloproteinase-7 (related to extracellular matrix turnover), ST-2, and N-terminal pro-B-type natriuretic peptide. A machine-learning-derived model using a combination of biomarkers was strongly predictive of the risk of DHFA (standardized hazard ratio: 2.85; 95% confidence interval: 2.03 to 4.02; $p < 0.0001$) and markedly improved the risk prediction when added to the MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure Risk Score) risk score. In an independent cohort (PHFS), the model strongly predicted the risk of DHFA (standardized hazard ratio: 2.74; 95% confidence interval: 1.93 to 3.90; $p < 0.0001$), which was also independent of the MAGGIC risk score.

CONCLUSIONS Various novel circulating biomarkers in key pathophysiological domains are predictive of outcomes in HFpEF, and a multimarker approach coupled with machine-learning represents a promising strategy for enhancing risk stratification in HFpEF. (J Am Coll Cardiol 2020;75:1281-95) © 2020 by the American College of Cardiology Foundation.



Listen to this manuscript's
audio summary by
Editor-in-Chief
Dr. Valentin Fuster on
JACC.org.

From ^aDivision of Cardiovascular Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania; ^bUniversity of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania; ^cBristol-Myers Squibb Company, Lawrenceville, New Jersey; and the ^dVanderbilt University Medical Center, Nashville, Tennessee. *Drs. Chirinos and Olenko have contributed equally to this work. This work was founded by an Investigator-Initiated research grant from Bristol-Myers Squibb (to Dr. Chirinos) and National Institutes of Health (NIH) grant R01HL088577 (to Dr. Cappola). Dr. Chirinos is supported by NIH grants R01-HL 121510-01A1, R61-HL-146390, R01-AG058969, 1R01-HL104106, P01-HL094307, R03-HL146874-01, and R56-HL136730; has received consulting honoraria from Sanifit, Microsoft, Fukuda-Denshi, Bristol-Myers Squibb, OPKO Healthcare, Ironwood Pharmaceuticals,

ABBREVIATIONS AND ACRONYMS

CI	= confidence interval
DHFA	= death or heart failure-related hospital admission
FABP	= fatty acid binding protein
FGF	= fibroblast growth factor
GDF	= growth differentiation factor
HF	= heart failure
hFABP	= heart-type fatty acid binding protein
HFpEF	= heart failure with preserved ejection fraction
IL	= interleukin
ML	= machine learning
MMP	= matrix metalloproteinase
MPO	= myeloperoxidase
NGAL	= neutrophil gelatinase-associated lipocalin
OPG	= osteoprotegerin
OPN	= osteopontin
PFI	= permutation feature importance
sFLT	= soluble fms-like tyrosine kinase
sTNF	= soluble tumor necrosis factor-receptor
TIMP	= tissue inhibitor of metalloproteinases
TNF	= tumor necrosis factor
TPOT	= tree-based pipeline optimizer
VEGF	= vascular endothelial growth factor

The prevalence of heart failure (HF) has markedly increased and now represents an enormous clinical and public health problem. Heart failure with a preserved ejection fraction (HFpEF) accounts for approximately one-half of all HF cases, a proportion that will likely increase as the population ages. To date, no pharmacological interventions have clearly proven to improve outcomes in randomized trials in HFpEF.

HFpEF is a heterogeneous condition, and accordingly, patients with HFpEF exhibit a variable clinical course and prognosis. At present, more accurate risk-stratification strategies are required, which need to be incremental and independent of clinical prediction scores. Advances in peripheral blood analytical techniques provide an opportunity to measure multiple biomarkers using small volumes of plasma, an approach that could be readily implemented in clinical practice and in participant selection for clinical trials. Targeted multimarker approaches may offer the capability to enhance risk prediction and to better understand underlying biologic abnormalities in HFpEF.

SEE PAGE 1296

In this study, we primarily aimed to assess the value of a targeted multimarker approach in plasma, coupled with machine learning (ML) methods, to enhance the prediction of outcomes in HFpEF. We also assessed the patterns of covariation in plasma biomarkers in this population.

METHODS

STUDY POPULATION. We primarily utilized data and biosamples from TOPCAT (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone

Antagonist Trial). We also utilized data and samples from the PHFS (Penn Heart Failure Study) for external validation of an ML-based outcome prediction model derived from the TOPCAT cohort.

TOPCAT data. TOPCAT data and samples were obtained from the National Heart, Lung, and Blood Institute. The parent trial data are available to other researchers through the National Institutes of Health Biolincc website. TOPCAT was a multicenter, international, randomized, double-blind, placebo-controlled trial of spironolactone that enrolled 3,445 adults with HFpEF across 6 countries from 2006 to 2012. The primary goal of the trial was to determine if spironolactone was associated with a reduction in the composite outcome of cardiovascular mortality, aborted cardiac arrest, or HF hospitalization. The design, general characteristics of the study population, and primary results of the trial have been previously published (1-3). Key inclusion and exclusion criteria for TOPCAT are listed in the [Supplemental Appendix](#). All study participants provided written informed consent.

In this analysis, we examined the relationship between biomarkers and a composite endpoint of death or heart failure-related hospitalization (DHFA), which is increasingly utilized in HFpEF studies (4), and was also prospectively adjudicated in our validation cohort (PHFS). Of the 3,445 participants, 379 patients provided baseline (pre-randomization) plasma samples for analysis.

PHFS data. The PHFS is a prospective cohort study of ambulatory HF patients recruited between 2003 and 2011 at the University of Pennsylvania (Philadelphia, Pennsylvania), Case Western Reserve University (Cleveland, Ohio), and the University of Wisconsin (Madison, Wisconsin) (5,6). Patients with a clinical diagnosis of HF as determined by an HF specialist were enrolled. At the time of enrollment, standardized questionnaires were administered to participants and their physicians to obtain detailed clinical data as described previously (5,6). Participants with expected mortality of 6 months or less from a

Pfizer, Akros Pharma, Merck and Bayer; has received research grants from the National Institutes of Health, American College of Radiology Network, Fukuda-Denshi, Bristol-Myers Squibb, and Microsoft; and he is named as inventor in a University of Pennsylvania patent for the use of inorganic nitrates/nitrites for the treatment of HFpEF and a patent application for the use of novel neopeptide biomarkers of tissue fibrosis in HFpEF. Drs. Zhao, Basso, Cvijic, Spires, Wang, and Seiffert are employees of and own stock in Bristol-Myers Squibb. Drs. Li, Yarde, and Gordon are employees of Bristol-Myers Squibb. Dr. Zamani is supported by grant K23-HL-130551; and has been a consultant for Vyaire. Dr. Margulies has received research grants from Merck, Sanofi, GlaxoSmithKline, AstraZeneca, and Luitpold; and has received consulting honoraria from Merck, GlaxoSmithKline, and Luitpold. Dr. Car is an employee and stockholder of Bristol-Myers Squibb; and is a future employee and stock holder of Agios Pharmaceuticals. Dr. Moore is supported by NIH grant LM010098. Dr. Cappola has received research funding from Bristol-Myers Squibb. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received July 18, 2019; revised manuscript received December 22, 2019, accepted December 23, 2019.

noncardiac condition, as judged by their treating physician; who were on mechanical circulatory support; or who were unable to provide informed consent were excluded. Participants provided written informed consent. Venous blood samples were obtained at the time of enrollment and stored at -80°C . An institutional review board from each of the participating centers approved the protocol. For the present analyses, we only included PHFS participants with HFpEF (left ventricular ejection fraction $\geq 50\%$), rather than HF with reduced or recovered ejection fraction. The PHFS sample with available plasma samples was composed of 156 subjects, of whom 125 had data available for computation of the MAGGIC risk score.

BIOMARKER ANALYSES. We measured 48 protein analytes using a Luminex Bead-Based multiplexed assay (Bristol-Myers Squibb, Ewing Township, New Jersey). Analytes were chosen to represent a diverse number of physiological processes related to cardiovascular disease, and downstream effects including angiogenesis, atherothrombosis, cardiomyocyte injury, extracellular matrix turnover, cell-matrix interactions, tissue remodeling, inflammation, adipocyte signaling, intermediary metabolism, kidney function/injury, calcification/mineral metabolism, neurohormonal regulation, and myocyte stretch (**Table 1, Central Illustration**). The assay range per analyte is shown in **Supplemental Table 1**. We note that this assay is different from that previously utilized by Small et al. (7). All biomarkers were measured from the same aliquot of each patient's baseline sample.

ML METHODS. Biomarker clustering and network analysis. To examine the clustering and covariance of biomarkers, we generated a biomarker correlation matrix and represented it as a heatmap. We also performed formal variable cluster analyses to identify biomarker clusters that exhibit shared variability. We used the variable cluster module from Jmp-Pro version 13 for Mac (SAS Institute, Cary, North Carolina), which is based on the SAS VARCLUS procedure. This procedure is an iterative unsupervised ML technique, which divides a set of numeric variables into either disjoint or hierarchical clusters based on similarity. The clusters are created in a way that variables from the same cluster are correlated with each other but have a low correlation with any other cluster. It starts with all variables in a single cluster, and proceeds by iteratively splitting and assigning variables to new clusters until no new splits or assignments are possible. We also applied network analysis, with the nodes representing individual

TABLE 1 Biomarkers Included in the Luminex Panel (Selected A Priori)

Pathophysiological Domain	Biomarker	Other Common Names
Angiogenesis	Angiopoietin Endoglin Vascular endothelial growth factor (VEGF) A Soluble fms-like tyrosine kinase-1 (sFLT1) Endostatin	Soluble VEGF receptor-1
Atherothrombosis	Soluble P-selectin (sP-selectin) Plasminogen activator inhibitor (PAI)-1	
Cardiomyocyte injury	Troponin T Heart-type fatty acid binding protein (hFABP)	Fatty acid binding protein-3
Extracellular matrix turnover	Matrix metalloproteinase (MMP)-2 MMP-3 MMP-7 MMP-8 MMP-9 MMP-12 Tissue inhibitor of metalloproteinases (TIMP)-1 TIMP-4 Tenascin-C	
Cell-matrix interactions	Syndecan-1 Syndecan-4	
Tissue remodeling, inflammation, and fibrosis	sST-2 Galectin-3 Liver fibrosis: YKL-40/chitinase 3-like 1 (CHI3L1)	Soluble interleukin 1 receptor-like 1 Chitinase-3-like protein-1 (CH3L-1)
Inflammation	C-reactive protein (CRP) Tumor necrosis factor (TNF)- α Soluble TNF-receptor 1 (sTNF-RI) Soluble TNF-receptor 2 (sTNF-RII) Interleukin (IL)-1 β IL-6 IL-8 IL-10 Fas sICAM Pentraxin (PTX)-3 Myeloperoxidase (MPO)	Chemokine (CXC) ligand 8 Apoptosis antigen-1, CD95
Adipocyte biology	Adiponectin Fatty acid binding protein-4 (FABP-4)	
Intermediary metabolism	Fibroblast growth factor (FGF)-21 Growth differentiation factor (GDF)-15	
Kidney function or injury	Neutrophil gelatinase-associated lipocalin (NGAL) Cystatin-C Kim-1	Lipocalin-2 T-cell immunoglobulin and mucin domain 1 (TIM-1), Hepatitis A virus cellular receptor 1 (HAVcr-1)
Mineral metabolism/calcification	Fibroblast growth factor (FGF)-21 Osteopontin Osteoprotegerin	
Neurohormonal regulation and myocyte stretch	Endothelin-1 Renin NT-proBNP NT-proANP	

NT-proANP = N-terminal pro-A-type natriuretic peptide; NT-proBNP = N-terminal pro-B-type natriuretic peptide.

biomarkers and the edges (connections) between nodes representing the correlation coefficient between a given biomarker (node) pair. To better visualize structural patterns within this connection matrix, we extracted the connectivity backbone, which reveals dominant connections and clusters of dense connectivity, as previously described (8).

Development of predictive models for outcomes. We utilized the model selection with tree-based pipeline optimizer (TPOT) platform to generate a classification predictive model for our dataset of interest. TPOT is an automated ML tool that employs “genetic programming” to build pipelines of ML methods for classification or regression along with pre-processing operators, such as data transformers and feature selectors (9). This technique was inspired by biological genetic mutational processes and the subsequent selection of “fit” genes during evolution. “Genetic” programming in the setting of ML algorithm optimization refers to the fact that ML pipelines are subjected to rounds (generations) of modifications in the form of mutation and recombination. At the end of each generation after modifications are implemented, the performance of each individual ML pipeline is evaluated and the best-fitted ones are selected for the next round. This technique allows the selections of the best-performing ML model for a given problem in a completely agnostic manner. TPOT classification pipelines are generated from the subset of ML methods and data pre-processing operators that are extracted from the Scikit-learn Python library, and contains 11 classification methods, 14 data transforming operators, and 5 feature selection methods. During the optimization process, various combinations of transformers are combined with ML methods into a pipeline in a tree-based manner. The methodology for model selection with TPOT for clinical datasets was described previously (10).

We applied z-score normalization for all biomarkers, followed by TPOT optimization for a maximum of 1,000 generations or 24 h, whichever occurred first, using 10-fold cross-validation and balanced accuracy as classification performance estimate. Due to the stochastic nature of the algorithm, we ran TPOT 30 times per dataset and selected the best performing model pipeline for further investigation. We used the permutation feature importance (PFI) approach to calculate the predictive ability of the particular variables. In this approach, we first calculated a pipeline performance on the original dataset, then permuted the values within a variable and calculated the performance of the pipeline on the modified dataset. The resulting difference in performances is the PFI coefficient. This procedure was replicated 100 times for each variable, and the mean of the replicates was taken as a final PFI value. The PFI is the gain of model balanced accuracy (11) introduced by a biomarker, on top of the prediction provided by all other biomarkers in the model, thus measuring the nonredundant prediction provided by that particular biomarker.

Of note, the development of predictive models for outcomes is an independent analysis and does not depend on the results of cluster analyses described in the previous text.

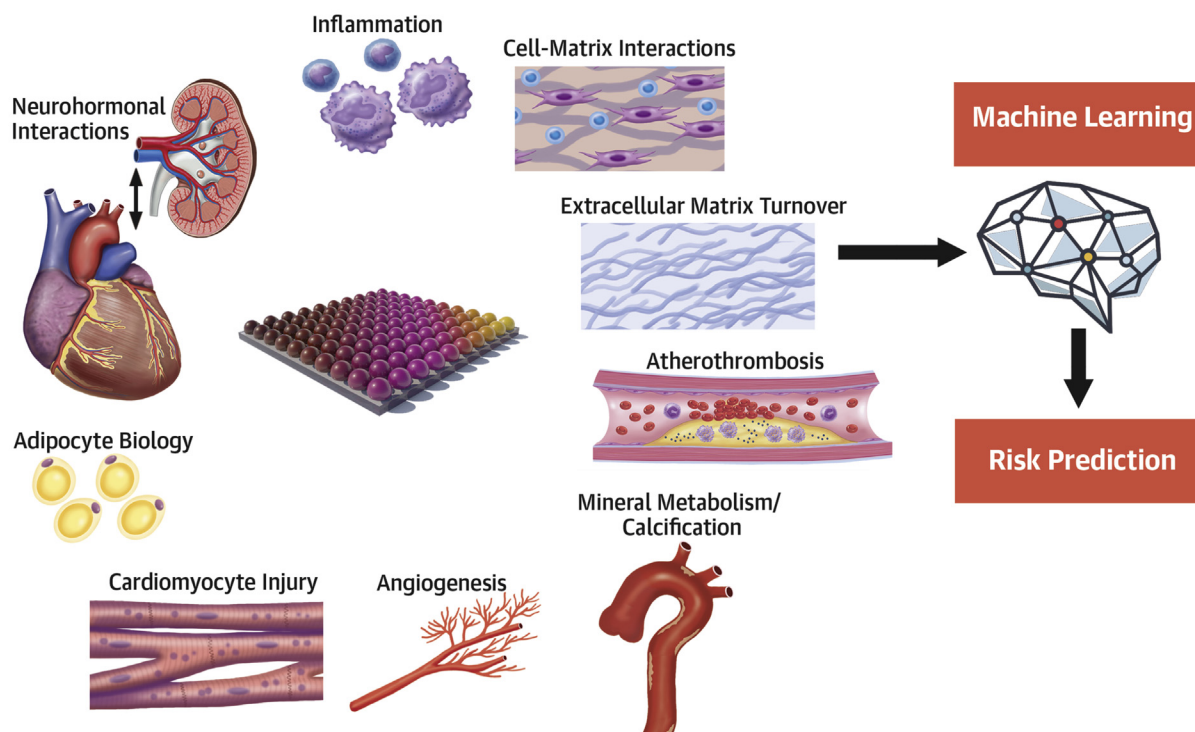
STATISTICAL ANALYSES. Participant characteristics were summarized using mean \pm SD for normally distributed variables and median (interquartile range) for non-normally distributed continuous variables. Categorical variables are expressed as counts (percentages). Because not all TOPCAT participants had available samples, we compared subjects who had available samples for measurement of the biomarkers of interest versus those who did not. We used the nonpaired Student's *t*-test for normally distributed variables, the Kruskal-Wallis test for non-normally distributed variables, and the chi-square test or Fisher exact test, as appropriate, for categorical variables.

We assessed the relationship between individual biomarkers and the risk of DHFA using Cox regression. To provide an intuitive unit-independent comparison between the biomarkers, hazard ratios (HRs) for all biomarkers are standardized (expressed per SD increase, or 1-point increased in the *z* score, after boxcox transformation to improve normality of the distribution as needed). To visualize the prediction of each biomarker relative to each other, we plotted the standardized hazard ratio against the log-10 *p* value (i.e., in order of statistical significance) on a volcano plot, in which the Bonferroni-corrected significance level was also displayed (corrected for 49 individual tests, 1 per biomarker). We also tested interactions between the pre-randomization level of each biomarker and randomized treatment with spironolactone, as predictors of DHFA.

The incremental prediction provided by the multimarker ML model was tested by adding the ML risk score to a baseline Cox model that contains the MAGGIC risk score, which is a well-validated prediction score that incorporates multiple demographic, clinical, and laboratory variables (12). All of these variables were available for TOPCAT participants, except the time of HF diagnosis, for which no points were given to any subject. The model was independently validated in the PHFS. For each validation cohort participant, an ML-model risk score was computed and was analyzed as a predictor of DHFA, in models with and without the MAGGIC risk score. We assessed Schoenfeld and Martingale residuals to test the proportionality and linearity assumptions in Cox models. The Harrel's *c* index, which is analogous to the receiver-operator characteristic curve, was computed to compare various models.

CENTRAL ILLUSTRATION Multimarker-Based Machine Learning Approach for Risk Prediction in Heart Failure With Preserved Ejection Fraction

Multiplex Plasma Protein Measurements



Chirinos, J.A. et al. J Am Coll Cardiol. 2020;75(11):1281-95.

We performed multiplex-based measurements of 49 proteins related to key biological pathways in the TOPCAT (Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist) trial. We then derived a predictive model for outcomes using machine learning. We then validated the prognostic score in a separate cohort (Penn Heart Failure Study).

Statistical significance was defined as a 2-tailed p value <0.05 . All probability values presented are 2-tailed. Statistical analyses were performed using the Matlab statistics and ML toolbox (Matlab 2019a, the Mathworks, Natick, Massachusetts) and SPSS for Mac version 22 (SPSS Inc., Chicago, Illinois).

RESULTS

TOPCAT POPULATION. A comparison of trial participants who had versus those who did not have available frozen plasma samples for biomarker measurements is shown in Table 2. Subjects with available samples were slightly older, with a slightly greater proportion of men. Subjects with available samples

were also more obese and exhibited a higher prevalence of hypertension, angiotensin-converting enzyme inhibitor/angiotensin receptor blocker, and statin use; lower blood pressure; and a higher prevalence of atrial fibrillation, previous myocardial infarction, and advanced New York Heart Association functional class (III/IV). There were no subjects with available samples from Georgia, Brazil, or Argentina, whereas 29.4% of subjects without available samples were enrolled in these countries.

CLUSTERING OF BIOMARKERS. Figure 1 shows a heatmap representing the correlation between different biomarkers in the study population. Figure 2 shows a plot of the network connectivity backbone,

TABLE 2 General Characteristics of Study Participants With Versus Without Available Plasma Samples

	Participants Without Available Samples (n = 3,063)	Participants With Available Samples (n = 379)	p Value
Demographic characteristics			
Age, yrs	69 (61-76)	70 (62-77)	0.0298
Male	1,465 (47.83)	203 (53.56)	0.0351
Race			0.115
White	2,712 (88.54)	347 (91.56)	
Black	274 (8.95)	28 (7.39)	
Other	77 (2.5)	4 (1.1)	
BMI, kg/m ²	30.8 (27.1-35.7)	31.8 (27.8-36)	0.0384
Heart rate, beats/min	68 (62-76)	68 (60-73.8)	0.0080
Systolic BP, mm Hg	130 (120-140)	126 (120-135)	<0.0001
Diastolic BP, mm Hg	80 (70-81)	75 (66-80)	<0.0001
Country			<0.0001
United States	980 (32)	160 (44)	
Canada	273 (8.9)	53 (14)	
Russia	908 (29.6)	157 (41.4)	
Georgia, Brazil, Argentina	902 (29.4)	0 (0.0)	
Medical history			
NYHA functional class III-IV	992 (32.42)	144 (37.99)	0.0295
Myocardial infarction	777 (25.38)	116 (30.61)	0.0284
Stroke	235 (7.67)	30 (7.92)	0.8682
COPD	358 (11.69)	45 (11.87)	0.9174
Hypertension	2,788 (91.05)	358 (94.46)	0.0254
Peripheral arterial disease	281 (9.18)	38 (10.03)	0.5907
Atrial fibrillation	1,051 (34.32)	162 (42.74)	0.0012
Diabetes mellitus	990 (32.33)	128 (33.77)	0.5720
Medication use			
Beta-blockers	2,375 (77.56)	301 (79.42)	0.4124
Calcium-channel blockers	1,159 (37.85)	134 (35.36)	0.3441
Diuretics	2,504 (81.78)	312 (82.32)	0.7951
Glucose-lowering agents	849 (27.73)	113 (29.82)	0.3928
ACE inhibitors or ARBs	2,593 (84.68)	306 (80.74)	0.0468
Statins	1,555 (50.78)	250 (65.96)	<0.0001

Values are median (interquartile range) or n (%).

ACE = angiotensin-converting enzyme; ARB = angiotensin receptor blocker; BMI = body mass index; COPD = chronic obstructive pulmonary disease; NYHA = New York Heart Association.

also representing the relationships between biomarkers. [Supplemental Table 2](#) shows the results of formal variable cluster analyses. There were 6 dominant clusters observed, as shown in [Figure 1](#) (and numbered according to the results of variable cluster analysis in [Supplemental Table 2](#)). Two large related clusters (labeled as clusters 1 and 4) were found, which included biomarkers of fibrosis/tissue remodeling (matrix metalloproteinase [MMP]-2, -3 and -9, Tenascin C, tissue inhibitor of metalloproteinases [TIMP]-1, Galectin-3), inflammation (Fas, sTNFR1I, myeloperoxidase [MPO]), liver fibrosis (YKL-40/chitinase 3-like 1 [CHI3L1]), and renal injury/function (neutrophil gelatinase-associated lipocalin [NGAL], cystatin C). Another cluster (cluster 3)

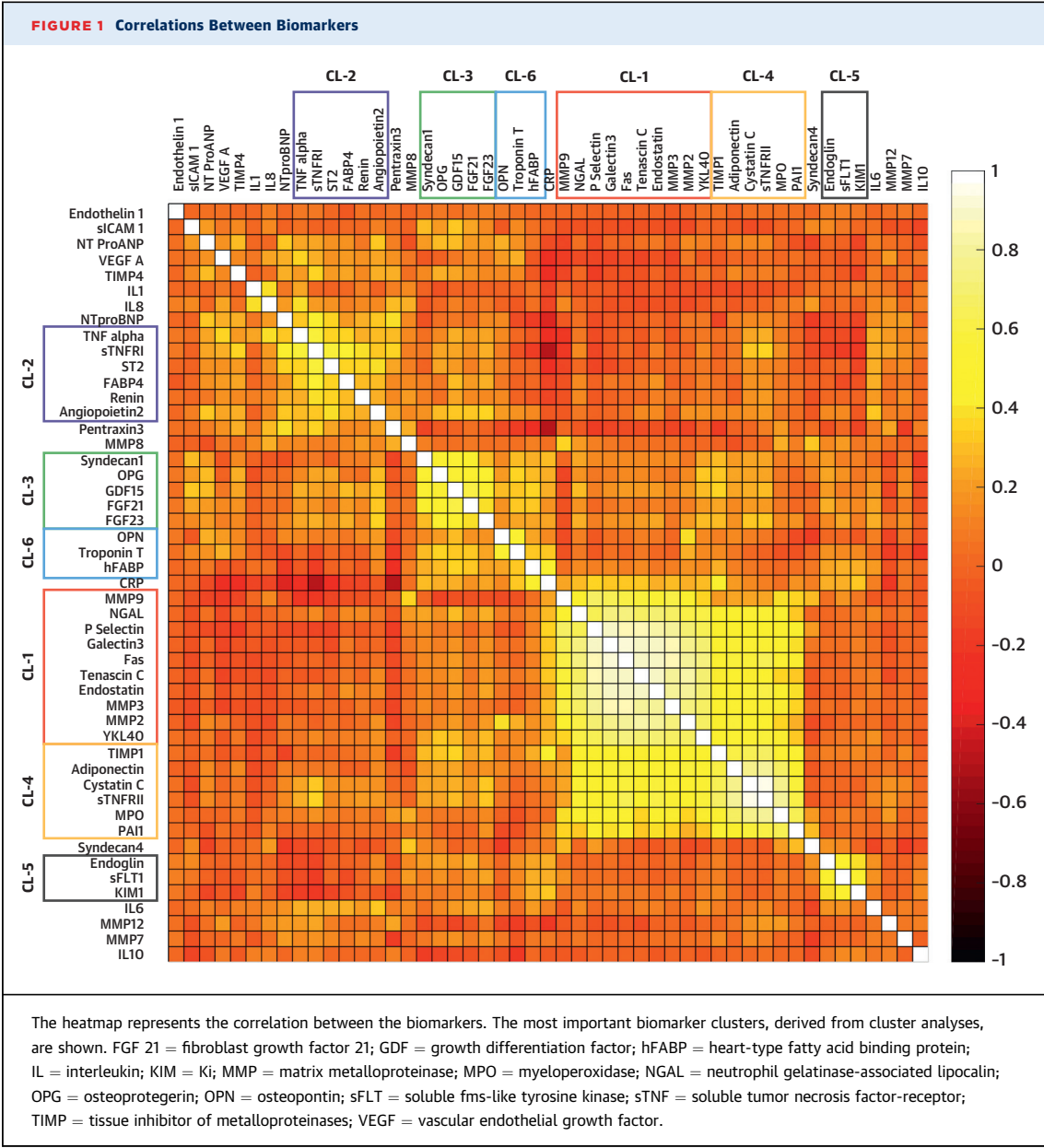
included neurohormonal regulators of intermediary (fibroblast growth factor [FGF]-21, growth differentiation factor-15 [GDF-15]) and mineral (OPG, FGF-23) metabolism. As expected, biomarkers of myocardial injury (troponin T and heart-type fatty acid binding protein-[FABP]) clustered together, along with osteopontin. Endoglin, soluble fms-like tyrosine kinase-1, and KIM also clustered together (cluster 5) and demonstrated strong inter-relationships. Finally, a less well-defined cluster was identified (cluster 2), which included inflammatory mediators related to the tumor necrosis factor (TNF)-alpha pathway (TNF- α , sTNFR1I), ST2, FABP4 (adipocyte-related protein), renin, and angiopoietin-2 (related to angiogenesis). The results of network analyses in general demonstrated similar patterns of biomarker connectivity ([Figure 2](#)).

RELATIONSHIPS BETWEEN BIOMARKER LEVELS AND OUTCOMES. During a median follow-up of 2.86 years, 94 subjects in the sample experienced death or an HF-related admission. [Supplemental Table 3](#) shows standardized HRs and 95% confidence intervals (CIs) for DHFA for all examined biomarkers in unadjusted analyses (1 model per biomarker). [Figure 3A](#) shows a volcano plot for these hazard ratios plotted against the log-10 p value, showing the biomarkers that were predictive at the Bonferroni-corrected level of significance.

In nonadjusted analyses, multiple biomarkers predicted DHFA, including ST-2, 3 inflammatory biomarkers (TNF-alpha, sTNFR1I, and interleukin [IL]-6), 2 biomarkers related to metabolism and adipocyte biology (FABP-4 and GDF-15), 2 biomarkers related to mineral metabolism/calcification (FGF-23 and OPG), angiopoietin-2 (related to angiogenesis), MMP-7 (related to extracellular matrix turnover), YKL-40 (related to liver injury and inflammation), and N-terminal pro-B-type natriuretic peptide.

A number of additional biomarkers tended to predict DHFA, meeting nominal uncorrected significance, but without meeting significance at the multiple comparison-corrected alpha level, including FGF21, NGAL, renin, sTNFR1I, cystatin-C, IL-10, VEGF-A, osteopontin (OPN), and syndecan-4. The latter tended to be negatively associated with risk of DHFA ([Supplemental Table 3, Figure 3A](#)).

In analyses adjusted for the MAGGIC risk score ([Figure 3B](#)), FGF-23, FABP-4, and IL-6 were independently predictive of DHFA at the Bonferroni-corrected level of significance. Various other biomarkers tended to be associated with DHFA (only at nominal levels of significance) in adjusted analyses ([Figure 3B](#)).



INTERACTIONS WITH RANDOMIZED ARM. Interactions between randomized arm and endostatin ($p = 0.0322$), TIMP-1 ($p = 0.0417$), sTNFR2 ($p = 0.03$), MPO ($p = 0.0387$), adiponectin ($p = 0.0242$), and cystatin C ($p = 0.0492$) were found for death/HFA, in all cases suggesting greater benefit with higher biomarker levels. However, none of these interactions reached statistical significance after accounting for multiple comparisons.

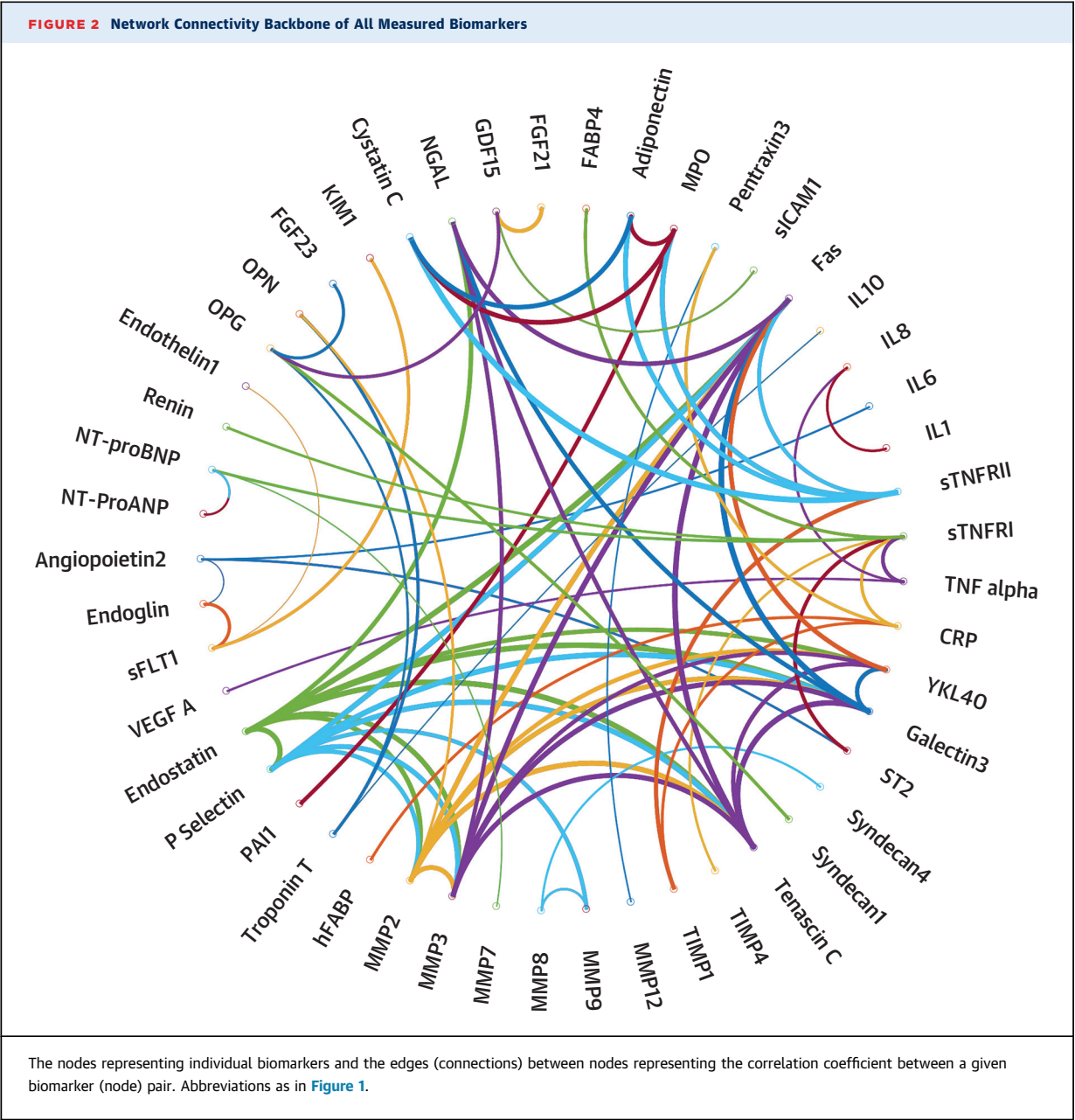
COMBINATION OF BIOMARKERS AS PREDICTORS OF DHFA. The TPOT optimization process using all biomarkers in the panel produced an ML pipeline, which

contained Stacking Estimator operator as a feature transformer, Robust Scaler as a feature preprocessor, and Bernoulli Naïve Bayes as an ML classifier.

Figure 4 shows the PFI coefficients of all 49 biomarkers included in the ML multimarker model for the prediction of DHFA. The PFI coefficient is a measure of the importance of each variable in the model, which in turn is influenced by its relationship with the outcome and any redundancy with other biomarkers.

The ML model was strongly predictive of the risk of DHFA (standardized HR: 2.86; 95% CI: 2.03 to 4.02; $p < 0.0001$) (**Figure 5**). In a model that included both

FIGURE 2 Network Connectivity Backbone of All Measured Biomarkers



the MAGGIC risk score and the ML score, the latter was a strong predictor of DHFA (standardized HR: 2.61; 95% CI: 1.84 to 3.71; $p < 0.0001$), whereas the MAGGIC risk score was no longer a significant predictor DHFA (standardized HR: 1.23; 95% CI: 0.98 to 1.54; $p = 0.07$) ([Figure 5](#)).

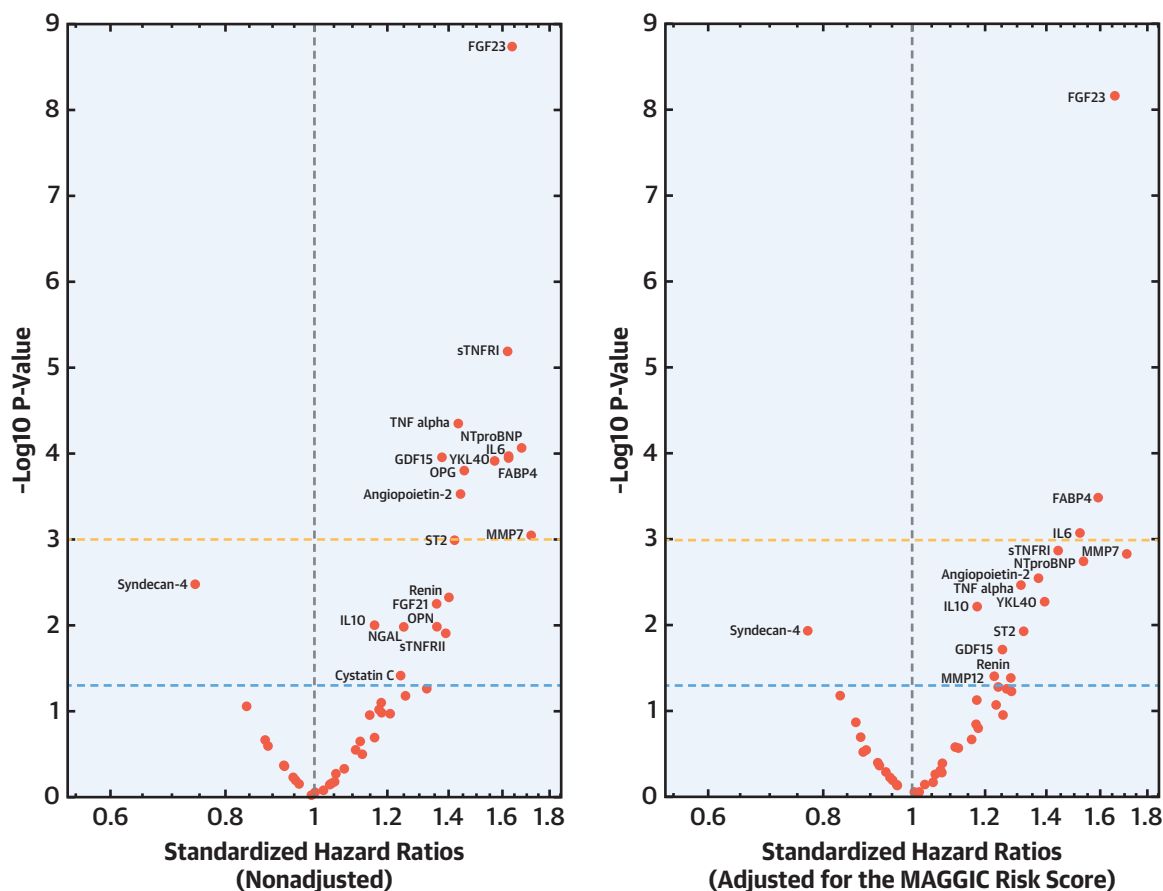
The ML score markedly improved the prediction of the endpoint when added to the MAGGIC risk score ([Figure 6](#)). The Harrel's c index for the MAGGIC risk score was 0.621 (95% CI: 0.560 to 0.682). The addition of the ML model score increased the Harrel's c index to 0.73 (95% CI: 0.669 to 0.790). The c index for a

model containing the ML score only (0.743; 95% CI: 0.682 to 0.803) was similar to the c index of the model that included both the ML score and the MAGGIC risk score ([Figure 6](#)).

We found no significant interaction between the ML score and randomized spironolactone therapy in the prediction of DHFA ($p = 0.34$).

MODEL VALIDATION IN THE PHFS. General characteristics of PHFS participants with HFpEF are shown in [Supplemental Table 4](#). During a median follow-up of 2.83 years, 69 subjects in the PHFS sample experienced a DHFA event. In this cohort, the ML score

FIGURE 3 Standardized Hazard Ratios for Examined Biomarkers



The red volcano plots show the standardized HRs for DHFA (1 model per biomarker) in unadjusted analyses (left) and adjusted for the MAGGIC risk score (right), plotted against the Log-10 p value. The dashed lines indicate the noncorrected (blue line) and Bonferroni-corrected (yellow line) level of significance. Abbreviations as in Figure 1.

was strongly predictive of the risk of DHFA (standardized HR: 2.74; 95% CI: 1.93 to 3.90; $p < 0.0001$). Figure 5 shows standardized HRs for the ML score and the MAGGIC risk score among subjects with available MAGGIC risk score data ($n = 125$; 59 events). The ML score was strongly predictive of the risk of DHFA (standardized HR: 2.91; 95% CI: 1.94 to 4.38; $p < 0.0001$). In a model that included both the MAGGIC risk score and the ML score, the latter was a strong predictor of DHFA (standardized HR: 2.68; 95% CI: 1.71 to 4.22; $p < 0.0001$), whereas the MAGGIC risk score was no longer a significant predictor DHFA (standardized HR: 1.15; 95% CI: 0.81 to 1.63; $p = 0.43$) (Figure 5).

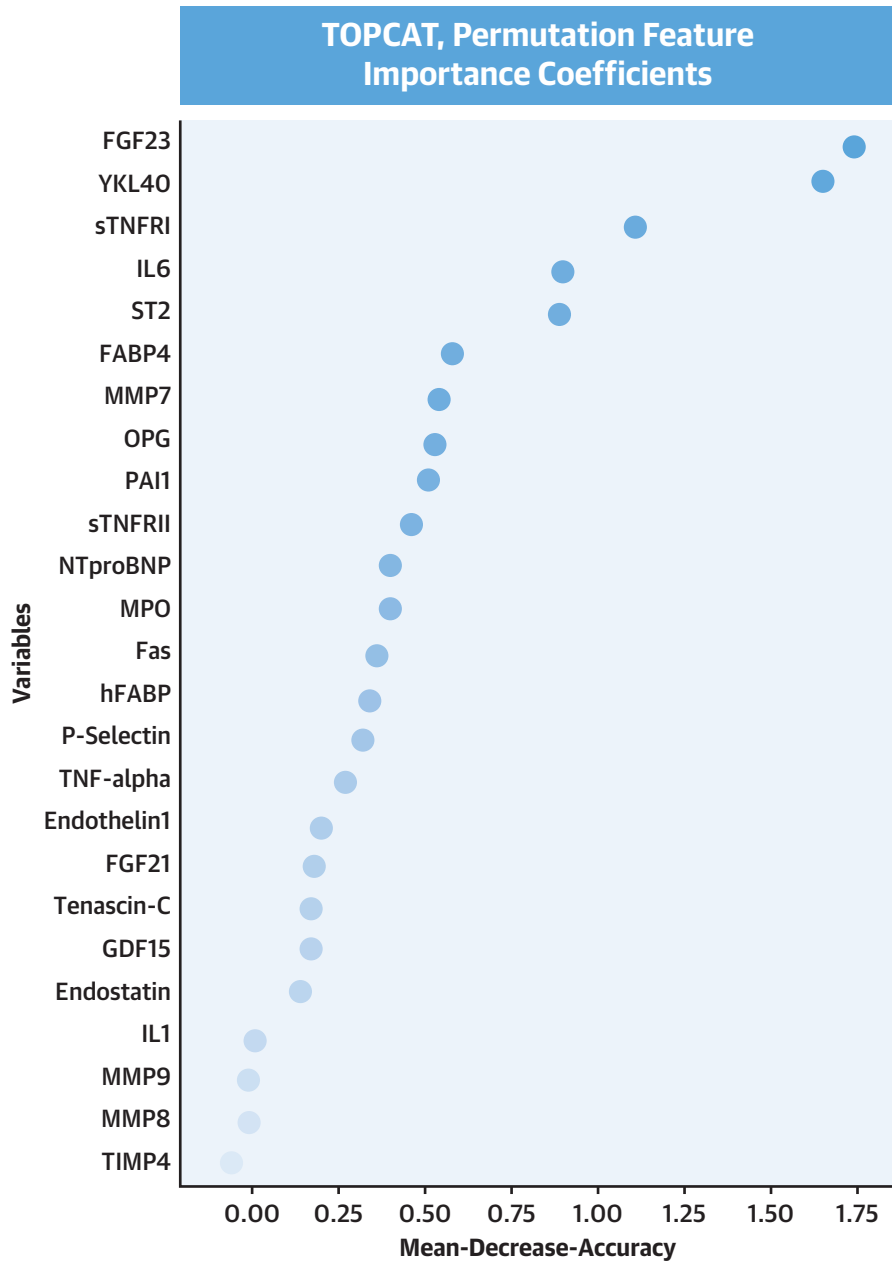
The ML score markedly improved the prediction of the endpoint when added to the MAGGIC risk score (Figure 6B). The Harrel's c index for the MAGGIC risk

score was 0.622 (95% CI: 0.557 to 0.687). The addition of the ML model score increased the Harrel's c index to 0.73 (95% CI: 0.646 to 0.814). The c index for a model containing the ML score only (0.717; 95% CI: 0.643 to 0.791) was similar to the c index of the model that included both the ML score and the MAGGIC risk score (Figure 6B).

DISCUSSION

In the current study, we assessed the prognostic value of a multimarker approach for risk stratification in HFpEF. We measured 49 pre-selected proteins using a multiplex assay, using baseline visit plasma samples obtained from TOPCAT trial participants. We report on the clustering patterns of key biomarkers in HFpEF and the relationship

FIGURE 4 Permutation Feature Importance Coefficients for Biomarkers in the Machine-Learning Model

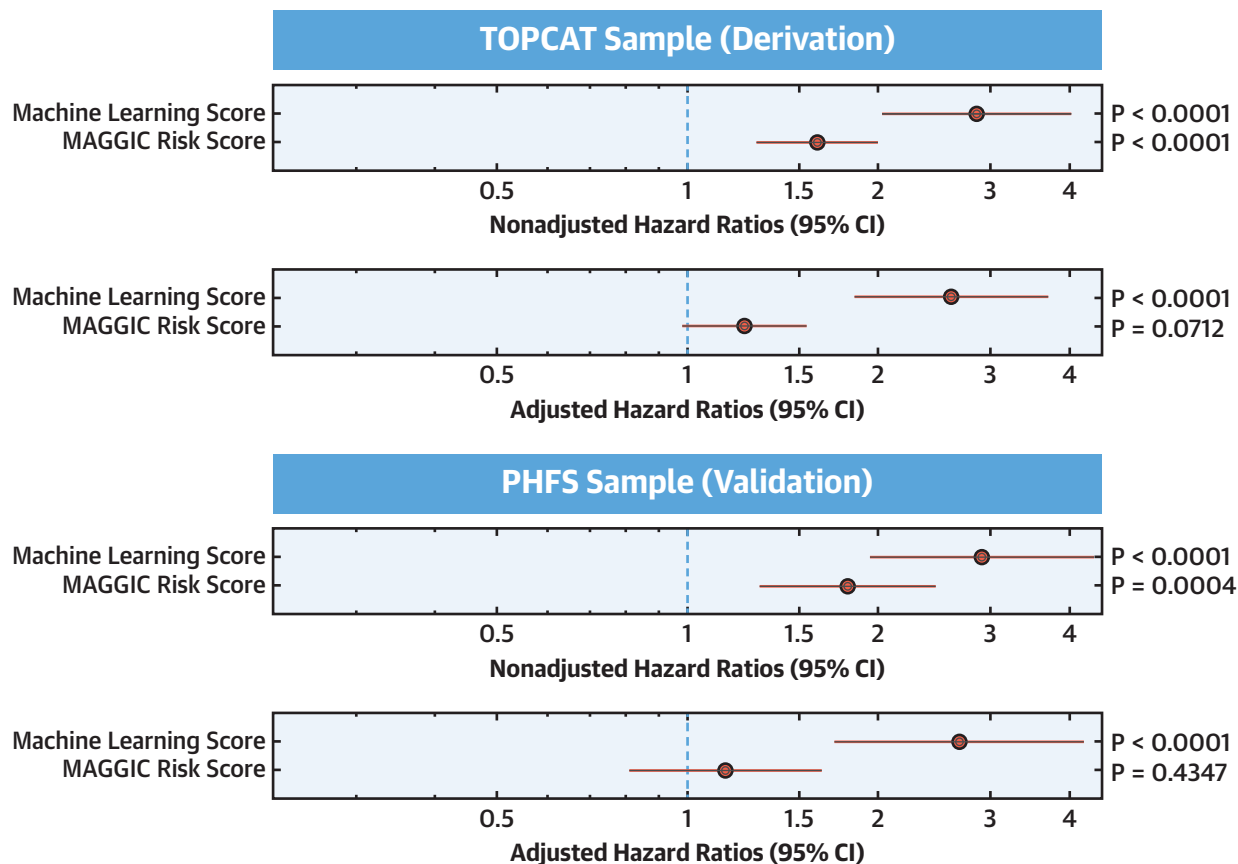


Biomarkers are ranked according to importance. Abbreviations as in [Figure 1](#).

between biomarker levels and risk of incident adverse outcomes. We found that several biomarkers related to mineral metabolism/calcification, liver fibrosis, inflammation, intermediary metabolism, myocardial fibrosis, adipocyte biology, and angiogenesis were predictive of DHFA. Finally, we utilized advanced ML techniques to assess the predictive value of optimal nonlinear combinations of

biomarkers for risk prediction and found that a multimarker approach markedly improved prediction above the MAGGIC risk score. We validated this predictive model in an external cohort (PHFS). Our findings advance our understanding of circulating biomarker profiles in HFpEF and suggest that multimarker approaches can be implemented for enhancing risk stratification in this condition.

FIGURE 5 Standardized HRs and 95% CIs for the Risk of DHFA

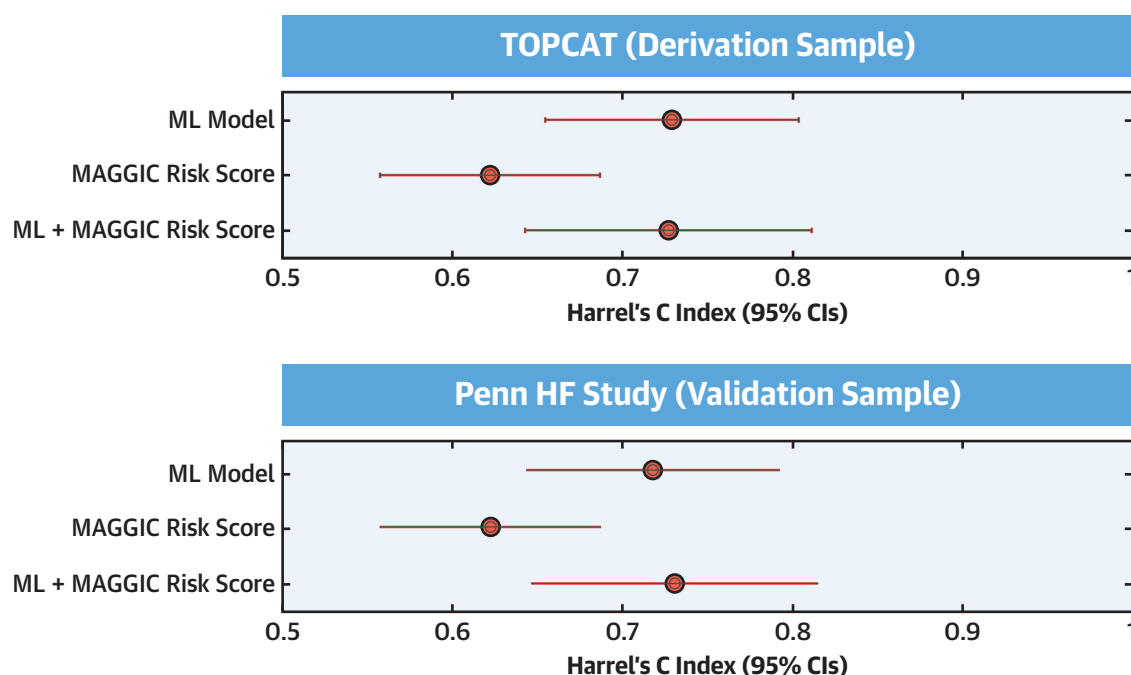


Hazard ratios (HRs) for the machine learning score versus the MAGGIC risk score are presented for nonadjusted analyses and analyses adjusted for each other, in the derivation (TOPCAT [Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist]) and validation (PHFS [Penn Heart Failure Study]) samples. CI = confidence interval.

BIOMARKER CLUSTERING. Our biomarker panel included proteins related to key biological pathways which have been implicated in the pathophysiology of HFpEF (Central Illustration). Interestingly, although many of these biomarkers are known to represent specific pathways, the significance, tissue specificity, and correlates of circulating levels in specific disease states have not been thoroughly investigated. Clustering patterns of specific biomarkers can provide insights regarding the phenotypic signatures related to various circulating proteins. Our cluster analyses demonstrated a large biomarker cluster composed of biomarkers implicated in inflammatory and extracellular matrix turnover pathways, specifically, biomarkers of fibrosis, tissue remodeling (MMP-2, -3 and -9, Tenascin C, TIMP-1, Galectin-3), and inflammation (Fas, sTNFR1, MPO). This pattern of clustering of inflammatory and tissue

remodeling biomarkers in our study is noteworthy, because it is consistent with the molecular underpinning proposed by a current hypothesis that chronic inflammation in HFpEF may serve to propagate myocardial fibrosis and target organ dysfunction (13). We also demonstrate that markers of renal injury, including cystatin C and NGAL, cluster with inflammatory and remodeling biomarkers, supporting a role for kidney injury/dysfunction in this systemic process. Interestingly, a biomarker of liver fibrosis (YKL-40/CHI3L1) also tightly clustered with the biomarkers previously mentioned, suggesting that profibrotic and inflammatory processes may extend beyond the heart and the kidney, and that the cardiac-hepatic axis requires further investigation in HFpEF, especially considering that HFpEF shares many risk factors with nonalcoholic fatty liver disease.

FIGURE 6 Harrel's Concordance Statistic (c Index) and 95% CIs for the Prediction of DHFA



Values are shown for the derivation (TOPCAT) and validation (PHFS) samples. Abbreviations as in [Figure 5](#).

BIOMARKERS AS PREDICTORS OF OUTCOMES. We examined the relationship between circulating levels of biomarkers and prognosis. Various biomarkers significantly predicted DHFA, including 2 biomarkers related to mineral metabolism/calcification (FGF-23 and OPG), 2 inflammatory biomarkers (TNF-alpha, sTNFRI, and IL-6), YKL-40 (related to liver injury and inflammation), 2 biomarkers related to intermediary metabolism and adipocyte biology (FABP-4 and GDF-15), angiopoietin-2 (related to angiogenesis), MMP-7 (related to extracellular matrix turnover), ST-2, and N-terminal pro-B-type natriuretic peptide. Some of these biomarkers have been previously reported to predict incident events in HFpEF, including ST2 (14) and GDF-15 (15). However, to the best of our knowledge, our study is the first to report a relationship between FGF-23, YKL-40, FABP-4, OPG, MMP7, and angiopoietin-2 and incident events. The relationship between inflammatory biomarkers is important, because it supports a role for inflammation in HFpEF, as discussed previously. Of note, IL-6 was predictive of DHFA independent of the MAGGIC risk score, along with FGF-23 and FABP-4.

Among the examined biomarkers, FGF-23 demonstrated the strongest association to adverse

outcomes, which was also independent of the MAGGIC risk score. FGF-23 is involved in phosphate homeostasis and increasing levels are observed as renal function decreases. FGF-23 has been shown to be a powerful predictor of incident HF (16) and has also been shown to be a strong predictor of mortality in heart failure with reduced ejection fraction (HFREF) (17). However, to our knowledge this has not been shown in HFpEF. The mechanisms by which FGF-23 is associated with incident cardiovascular events and HFpEF are likely multifactorial. In animal models, administration of FGF-23 led directly to cardiomyocyte hypertrophy (18), and increased FGF-23 levels are associated with left ventricular hypertrophy in humans (19). Additionally, FGF-23 has been shown to suppress angiotensin-converting enzyme-2, which normally degrades the vasoconstrictor angiotensin-II into vasodilator peptides.

We showed an association between FABP4 and incident risk of DHFA, which was independent of the MAGGIC risk score. FABP4 is expressed in adipocytes and macrophages and plays an important role in the development of insulin resistance and atherosclerosis in relation to systemic inflammation (20). Increased levels even in healthy individuals have been linked to

diastolic dysfunction (21) as well as in obese women (22). Levels of FABP4 have also been linked to atherosclerosis and increased carotid intimal medial thickness (23). In HFpEF, FABP4 may serve as an adipocyte-derived marker of the insulin resistance or inflammation, and may predict risk in this population on this basis.

We found YKL-40 to be associated with incident DHFA, which ranked second in importance to FGF-23 in the ML model. YKL-40 is considered a marker of liver fibrosis, and has been shown to be elevated in patients with NAFLD and more advanced fibrosis (24). There are little data regarding the role of liver fibrosis in HFpEF; therefore, the finding that increased YKL-40 is a predictor of outcomes in HFpEF supports the need for further work in understanding mechanisms behind liver fibrosis in this population, as mentioned previously.

We also found that angiopoietin-2, a biomarker of angiogenesis and endothelial dysfunction, was associated with incident DHFA in nonadjusted analyses. This adds to previous studies assessing the significance of biomarkers of angiogenesis in HFpEF. In a study of acute HF, biomarkers of angiogenesis, specifically angiogenin, were hubs of biomarker clusters, but only in HFpEF and not HFrEF or HF with mid-range ejection fraction (25). Neuropilin, another biomarker of angiogenesis, was associated with incident events in HFpEF, but not HFrEF, despite similarly increased levels in both groups (26). The mechanisms underlying these findings remain unclear and should be the subject of future research.

Finally, we found that osteoprotegerin, a member of the TNF-receptor superfamily that regulates both differentiation and function of osteoclasts, significantly predicted the risk of DHFA. Interestingly, osteoprotegerin clustered with FGF-23, which is also intimately involved in calcification processes, and in a previous study, osteoprotegerin was shown to be an independent predictor of death in decompensated HFpEF (27). Finally, we demonstrate a relationship between MMP7 and DHFA, which supports prior work in which various biomarkers related to collagen deposition (including PINP, PIIINP, and osteopontin) were shown to be associated with adverse clinical endpoints in HFpEF (26,28).

MULTIMARKER MODEL. We found that a multimarker ML model was strongly predictive of the risk of DHFA, and substantially improved the predictive power above and beyond the MAGGIC risk score, as shown by an important increase in the Harrel's *c* index, a measure of model discrimination.

Interestingly, the ML model alone was substantially more predictive than the MAGGIC risk score, and in a model that included both the ML score and the MAGGIC risk score, only the former was an independent predictor of DHFA. These findings were reproduced in an independent U.S.-based validation cohort (PHFS) in which the findings were very similar to those obtained from the primary cohort, increasing our confidence in the external validity of the model.

It is worth noting that the clustering patterns found at baseline were not necessarily informative regarding the predictive power of biomarkers in the context of outcome prediction. This is not surprising, because cluster analyses attempt to detect parallel variance between biomarkers that form a cluster, whereas outcome models attempt to maximize the orthogonal information provided by individual biomarkers.

The multimarker approach developed in this study can be applied in clinical trials and clinical practice, because analytical techniques that require minimal plasma volume to quantify a relatively large protein panel at reasonable cost are now available. The model developed in our study is suitable for clinical application upon refinement of automated assays, particularly if they can be deployed in standard clinical analyzers.

STUDY STRENGTHS AND LIMITATIONS. Strengths of our study include the inclusion of a well-characterized HFpEF cohort, the use of multiple biomarkers, advanced ML methods, and the validation of our model in an independent HFpEF sample. Our study also has several limitations. We did not have available plasma samples from all TOPCAT trial participants, and we had to restrict the study to a subpopulation with available samples. Although we found multiple highly significant associations between biomarkers and outcomes with strict Bonferroni correction, the power to detect weaker associations was limited. This may be particularly relevant for interactions with randomized treatment, which did not reach formal significance after correction for multiple testing. Because we do not know the tissue origins for most of the circulating biomarkers, we are uncertain about whether they are reflecting systemic or regional pathological responses. We note that our multiplex platform has assay-specific limits of detection that are not necessarily equivalent to established clinical assays, and the findings should be interpreted with this consideration in mind. In particular, our ML method was developed using this specific platform and is not intended for application using individually measured analytes with clinically

approved assays or other methods. Nevertheless, we provide convincing evidence that this multimarker technology coupled with ML provides robust prognostic information.

CONCLUSIONS

Our study demonstrates that a multimarker approach using circulating biomarkers can be a powerful tool for enhancing risk stratification in HFpEF. Future research should further examine this approach, including the use of broader proteomic panels in HFpEF.

ACKNOWLEDGMENTS This paper was prepared using TOPCAT trial research materials obtained from the National Heart, Lung, and Blood Institute Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the TOPCAT Trial or the National Heart, Lung, and Blood Institute. The authors are grateful for the technical support from Terrye Delmonte and Karl Kammerhoff from the BMS Biorepository.

ADDRESS FOR CORRESPONDENCE: Dr. Julio A. Chirinos, South Tower, Room 11-138, Perelman Center for Advanced Medicine, 3400 Civic Center Boulevard, Philadelphia, Pennsylvania 19104. E-mail: julio.chirinos@uphs.upenn.edu. Twitter: @JulioChirinosMd.

PERSPECTIVES

COMPETENCY IN PATIENT CARE AND PROCEDURAL OUTCOMES:

In patients with HFpEF, a combination of multiple plasma biomarkers centered around tissue remodeling, inflammation, renal dysfunction, and liver fibrosis is predictive of clinical outcomes.

TRANSLATIONAL OUTLOOK: A combination of biomarkers could be employed in the design of future studies to guide clinical management of patients with HFpEF.

REFERENCES

- Desai AS, Lewis EF, Li R, et al. Rationale and design of the treatment of preserved cardiac function heart failure with an aldosterone antagonist trial: a randomized, controlled study of spironolactone in patients with symptomatic heart failure and preserved ejection fraction. *Am Heart J* 2011;162:966-72.e10.
- Pfeffer MA, Claggett B, Assmann SF, et al. Regional variation in patients and outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) trial. *Circulation* 2015;131:34-42.
- Pitt B, Pfeffer MA, Assmann SF, et al. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med* 2014;370:1383-92.
- Lam CSP, Gamble GD, Ling LH, et al. Mortality associated with heart failure with preserved vs. reduced ejection fraction in a prospective international multi-ethnic cohort study. *Eur Heart J* 2018;39:1770-80.
- Ky B, French B, Ruparel K, et al. The vascular marker soluble fms-like tyrosine kinase 1 is associated with disease severity and adverse outcomes in chronic heart failure. *J Am Coll Cardiol* 2011;58:386-94.
- Lakshmi K, Shaw PA, Morley MP, et al. Thyroid dysfunction in heart failure and cardiovascular outcomes. *Circ Heart Fail* 2018;11:e005266.
- Small AM, Kiss DH, Anwaruddin S, et al. Soluble FMS-like tyrosine kinase-1 is a circulating biomarker associated with calcific aortic stenosis. *J Am Coll Cardiol* 2019;73:1364-5.
- Hidalgo CA, Klinger B, Barabasi AL, Hausmann R. The product space conditions the development of nations. *Science* 2007;317:482-7.
- Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Kidd LC, Moore JH. Automating biomedical data science through tree-based pipeline optimization. *Applications of Evolutionary Computation* 2016:123-37.
- Orlenko A, Moore JH, Orzechowski P, et al. Considerations for automated machine learning in clinical metabolic profiling: Altered homocysteine plasma concentration associated with metformin exposure. *Pac Symp Biocomput* 2018;23:460-71.
- Urbanowicz RJ, Moore JH. ExSTraCS 2.0: Description and evaluation of a scalable learning classifier system. *Evol Intell* 2015;8:89-116.
- Pocock SJ, Ariti CA, McMurray JJ, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J* 2013;34:1404-13.
- Paulus WJ, Tschope C. A novel paradigm for heart failure with preserved ejection fraction: comorbidities drive myocardial dysfunction and remodeling through coronary microvascular endothelial inflammation. *J Am Coll Cardiol* 2013;62:263-71.
- Sanders-van Wijk S, van Empel V, Davarzani N, et al. Circulating biomarkers of distinct pathophysiological pathways in heart failure with preserved vs. reduced left ventricular ejection fraction. *Eur J Heart Fail* 2015;17:1006-14.
- Izumiya Y, Hanatani S, Kimura Y, et al. Growth differentiation factor-15 is a useful prognostic marker in patients with heart failure with preserved ejection fraction. *Can J Cardiol* 2014;30:338-44.
- Almahmoud MF, Soliman EZ, Berton AG, et al. Fibroblast growth factor-23 and heart failure with reduced versus preserved ejection fraction: MESA. *J Am Heart Assoc* 2018;7:e008334.
- Koller L, Kleber ME, Brandenburg VM, et al. Fibroblast growth factor 23 is an independent and specific predictor of mortality in patients with heart failure and reduced ejection fraction. *Circ Heart Fail* 2015;8:1059-67.
- Faul C, Amaral AP, Oskoue B, et al. FGF23 induces left ventricular hypertrophy. *J Clin Invest* 2011;121:4393-408.
- Kestenbaum B, Sachs MC, Hoofnagle AN, et al. Fibroblast growth factor-23 and cardiovascular disease in the general population: the Multi-Ethnic Study of Atherosclerosis. *Circ Heart Fail* 2014;7:409-17.
- Furuhashi M, Saitoh S, Shimamoto K, Miura T. Fatty acid-binding protein 4 (FABP4): pathophysiological insights and potent clinical biomarker of metabolic and cardiovascular diseases. *Clin Med Insights Cardiol* 2014;8:23-33.
- Fuseya T, Furuhashi M, Yuda S, et al. Elevation of circulating fatty acid-binding protein 4 is independently associated with left ventricular diastolic dysfunction in a general population. *Cardiovasc Diabetol* 2014;13:126.
- Engeli S, Utz W, Haufe S, et al. Fatty acid binding protein 4 predicts left ventricular mass

and longitudinal function in overweight and obese women. *Heart* 2013;99:944-8.

23. Yeung DC, Xu A, Cheung CW, et al. Serum adipocyte fatty acid-binding protein levels were independently associated with carotid atherosclerosis. *Arterioscler Thromb Vasc Biol* 2007;27:1796-802.

24. Kumagai E, Mano Y, Yoshio S, et al. Serum YKL-40 as a marker of liver fibrosis in patients with non-alcoholic fatty liver disease. *Sci Rep* 2016;6:35282.

25. Tromp J, Khan MAF, Mentz RJ, et al. Biomarker profiles of acute heart failure patients with a mid-

range ejection fraction. *J Am Coll Cardiol HF* 2017;5:507-17.

26. Tromp J, Khan MA, Klip IT, et al. Biomarker profiles in heart failure patients with preserved and reduced ejection fraction. *J Am Heart Assoc* 2017;6:e003989.

27. Aramburu-Bodas O, Garcia-Casado B, Salamanca-Bautista P, et al. Relationship between osteoprotegerin and mortality in decompensated heart failure with preserved ejection fraction. *J Cardiovasc Med (Hagerstown)* 2015;16:438-43.

28. Krum H, Elsik M, Schneider HG, et al. Relation of peripheral collagen markers to death and

hospitalization in patients with heart failure and preserved ejection fraction: results of the I-PRESERVE collagen substudy. *Circ Heart Fail* 2011;4:561-8.

KEY WORDS HFpEF, biomarkers, fibrosis, inflammation, kidney, liver, Penn Heart Failure Study, TOPCAT trial

APPENDIX For an expanded Methods section and supplemental tables, please see the online version of this paper.