



Published in final edited form as:

J Am Coll Cardiol. 2020 March 24; 75(11): 1281–1295. doi:10.1016/j.jacc.2019.12.069.

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

Julio A. Chirinos, MD, PhD^{*,1,2}, Alena Orlenko, PhD^{*,2}, Lei Zhao, MD, PhD³, Michael D. Basso, MS³, Mary Ellen Cvijic, PhD³, Zhuyin Li, PhD³, Thomas E. Spires, MS³, Melissa Yarde, MS³, Zhaoqing Wang, MS³, Dietmar A. Seiffert, MD³, Stuart Prenner, MD^{1,2}, Payman Zamani, MD, MTR^{1,2}, Priyanka Bhattacharya, MD^{1,2}, Anupam Kumar, MD⁴, Kenneth B. Margulies, MD^{1,2}, Bruce D. Car, PhD³, David A. Gordon, PhD³, Jason H. Moore, PhD^{1,2}, Thomas P. Cappola, MD, ScM^{1,2}

¹Division of Cardiovascular Medicine, Hospital of the University of Pennsylvania, Philadelphia Pennsylvania

²University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania

³Bristol-Myers Squibb Company, Lawrenceville, New Jersey

⁴Vanderbilt University Medical Center, Nashville, Tennessee.

Abstract

Background: Better risk stratification strategies are needed to enhance clinical care and trial design in HFpEF.

Objective: To assess the value of a targeted plasma multi-marker approach to enhance our phenotypic characterization and risk prediction in HFpEF.

Methods: We measured 49 plasma biomarkers from TOPCAT trial participants ($n = 379$) using a Multiplex assay. We assessed the relationship between biomarkers and the risk of all-cause death or heart failure-related hospital admission (DHFA). 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 Penn Heart Failure study (PHFS, $n = 156$).

Address for correspondence: Julio A. Chirinos, MD, PhD, South Tower, Rm. 11-138., Perelman Center for Advanced Medicine, 3400, Civic Center Blvd., Philadelphia, PA. 19104., Telephone: 215-573-6606, Fax: 215-746-7415, julio.chirinos@uphs.upenn.edu, Twitter: @JulioChirinosMd.

*These authors contributed equally to this manuscript

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosures: J.A.C. has received consulting honoraria from Sanifit, Microsoft, Fukuda-Denshi, Bristol-Myers Squibb, OPKO Healthcare, Ironwood Pharmaceuticals, Pfizer, Akros Pharma, Merck and Bayer. He received research grants from National Institutes of Health, American College of Radiology Network, Fukuda-Denshi, Bristol-Myers Squibb and Microsoft. J.A.C. 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. K.B.M. has received research grants from Merck, Sanofi, Glaxo-Smith-Kline, AstraZeneca and Luitpold and has received consulting honoraria from Merck, Glaxo-Smith-Kline and Luitpold. The remaining authors have nothing to disclose.

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 neurohormonal regulators of mineral metabolism, intermediary metabolism, and biomarkers of myocardial injury. Multiple biomarkers predicted incident DHFA, including two biomarkers related to mineral metabolism/calcification (FGF-23 and OPG), 3 inflammatory biomarkers (TNF-alpha, sTNFR1 and IL-6), YKL-40 (related to liver injury and inflammation), two biomarkers related to intermediary metabolism and adipocyte biology (FABP-4 and GDF-15), angiotensin-2 (related to angiogenesis), MMP-7 (related to extracellular matrix turnover), ST-2 and NT-proBNP. A machine-learning derived model using a combination of biomarkers was strongly predictive of the risk of DHFA (Standardized HR = 2.85; 95%CI = 2.03-4.02; $P < 0.0001$) and markedly improved the risk prediction when added to the MAGGIC risk score. In an independent cohort (PHFS), the model strongly predicted the risk of DHFA (Standardized HR=2.74; 95%CI=1.93-3.90; $P < 0.0001$), which was also independent of the MAGGIC risk score.

Conclusions: Various novel circulating biomarkers in key pathophysiologic domains are predictive of outcomes in HFpEF and a multi-marker approach coupled with machine-learning represents a promising approach for enhancing risk stratification in HFpEF.

Condensed Abstract:

We assessed the relationship between 49 plasma biomarkers and the risk of death of heart failure admission (DHFA) in HFpEF. We used machine-learning to construct a multi-marker-based predictive model in the TOPCAT trial and validated the model in the Penn HF Study. Two main biomarker clusters were found, which included biomarkers of fibrosis/tissue remodeling, inflammation, renal injury/dysfunction and liver fibrosis. Multiple individual biomarkers predicted the risk of DHFA, including several biomarkers not previously studied in HFpEF. The multimarker model markedly improved the prediction of DHFA in the TOPCAT and Penn HF Study cohorts, above and beyond the MAGGIC risk score.

Keywords

HFpEF; biomarkers; fibrosis; inflammation; kidney; liver; TOPCAT trial; Penn Heart Failure Study

Introduction

The prevalence of heart failure (HF) has markedly increased and now represents an enormous clinical and public health problem. HF with a preserved ejection fraction (HFpEF) accounts for approximately half of all HF cases, a proportion that will likely increase as the population ages. To date, no pharmacologic 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 multi-marker approaches may offer the capability to enhance risk prediction and to better understand underlying biologic abnormalities in HFpEF.

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

Methods

Study population

We primarily utilized data and biosamples from the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist Trial (TOPCAT). We also utilized data and samples from the Penn Heart Failure Study (PHFS) for external validation of a ML-based outcome prediction model derived from the TOPCAT cohort.

TOPCAT: 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 multi-center, international, randomized, double-blinded, placebo-controlled trial of spironolactone that enrolled 3,445 adults with HFpEF across 6 countries from 2006-2012. 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 heart failure 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 online supplemental section. All study participants provided written informed consent.

In this analysis, we examined the relationship between biomarkers 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: The PHFS is a prospective cohort study of ambulatory HF patients recruited between 2003-2011 at the University of Pennsylvania (Philadelphia, PA), Case Western Reserve University (Cleveland, OH), and the University of Wisconsin (Madison, WI).(5,6) Patients with a clinical diagnosis of HF as determined by a 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 non-cardiac condition, as judged by their treating physician, mechanical circulatory support, or inability 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, rather than HFrEF or HF with recovered EF. 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, NJ). Analytes were chosen to represent a diverse number of physiologic 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 Online Table 1. We note that this assay is different from the one previously utilized by Small *et al.*(7) All biomarkers were measured from the same aliquot of each patient's baseline sample.

Machine learning 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 v13 for Mac (SAS Institute; Cary, NC), 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 biomarkers and the edges (connections) between nodes representing the correlation coefficient between a given biomarker (node) pair. In order 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 data set of interest. TPOT is an automated ML tool that employs “genetic programming” to build pipelines of ML methods for classification or regression along with preprocessing operators such as data transformers and feature selectors (9). This technique was inspired by biologic 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 preprocessing 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 1000 generations or 24 hours, 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 data set, 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 non-redundant 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 above.

Statistical Analyses

Participant characteristics were summarized using mean (SD) for normally distributed variables and median (interquartile range) for non-normally distributed continuous variables. Categorical variables are expressed as counts (percentages). Since not all TOPCAT participants had available samples, we compared subjects who had available samples for measurement of the biomarkers of interest vs. those who did not. We used the non-paired t test for normally-distributed variables, the Kruskal-Wallis test for non-normally distributed variables and the chisquare test or Fisher's exact test, as appropriate, for categorical variables.

We assessed the relationship between individual biomarkers and the risk of DHFA using Cox regression. In order to provide an intuitive unit-independent comparison between the biomarkers, hazard ratios for all biomarkers are standardized (expressed per standard-deviation increase, or 1-point increase 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 log₁₀ 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, one 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 multi-marker 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 heart failure diagnosis, for which no points were given to any subject. The model was independently validated in the PHFS. For each validation cohort participant, a 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 models. The Harrel's c index, which is analogous to the receiver-operator characteristic curve, was computed to compare various models.

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; Natwick, MA) and SPSS for Mac v22 (SPSS Inc., Chicago, IL).

Results

TOPCAT population

A comparison of trial participants who had vs. 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 males. Subjects with available samples were also more obese, and exhibited a higher prevalence hypertension, ACE inhibitor/ARB, and statin use, lower blood pressure and a higher prevalence of atrial fibrillation, previous myocardial infarction and advanced NYHA 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, also representing the relationships between biomarkers. Online 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 Online Table 2). Two large related clusters (labeled as clusters 1 and 4) were found, which included biomarkers of fibrosis/tissue remodeling (MMP-2, 3 and 9, Tenascin C, TIMP-1, Galectin-3), inflammation (Fas, sTNFR1I, MPO), liver fibrosis (YKL-40/chitinase 3-like 1 [CHI3L1]) and renal injury/function (NGAL, cystatin C). Another cluster (cluster 3) included neurohormonal regulators of intermediary (FGF-21, GDF-15) and mineral (OPG, FGF-23) metabolism. As expected, biomarkers of myocardial injury (troponin T and hFABP) clustered together, along with osteopontin. Endoglin, sFLT-1 and KIM also clustered together (cluster 5) and demonstrated strong interrelationships. Finally, a less-well defined cluster was identified (cluster 2), which included inflammatory mediators related to

the TNF-alpha pathway (TNF-a, sTNFR2), 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 biomarkers levels and outcomes

During a median follow-up of 2.86 years, 94 subjects in the sample experienced death or a HF-related admission. Online Table 3 shows standardized hazard ratios and 95% CIs for DHFA for all examined biomarkers in unadjusted analyses (one model per biomarker). Figure 3A shows a volcano plot for these hazard ratios plotted against the log₁₀ P-value, showing the biomarkers that were predictive at the Bonferroni-corrected level of significance.

In non-adjusted analyses, multiple biomarkers predicted DHFA, including ST-2, three inflammatory biomarkers (TNF-alpha, sTNFR1 and IL-6), two biomarkers related to metabolism and adipocyte biology (FABP-4 and GDF-15), two 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 NT-proBNP.

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, sTNFR2, cystatin-C, IL-10, VEGF-A, osteopontin (OPN) and syndecan-4. The latter tended to be negatively associated with risk of DHFA (Online 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 a ML pipeline which contained Stacking Estimator operator as a feature transformer, Robust Scaler as a feature preprocessor, and Bernoulli Naive Bayes as a machine learning 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=2.03-4.02; $P<0.0001$; Figure 5). In a model that included both the MAGGIC risk score and the ML score, the latter was a strong predictor of DHFA (Standardized HR=2.61; 95%CI=1.84-3.71; $P<0.0001$), whereas the MAGGIC risk score was no longer a significant predictor DHFA (Standardized HR=1.23; 95%CI=0.98-1.54; $P=0.07$; 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= 0.56-0.682). The addition of the ML model score increased the Harrel's c index to 0.73 (95%CI= 0.669-0.790). The c index for a model containing the ML score only (0.743; 95%CI=0.682-0.803) 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

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 was strongly predictive of the risk of DHFA (Standardized HR=2.74; 95%CI=1.93-3.90; $P<0.0001$). 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=1.94-4.38; $P<0.0001$). 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=1.71-4.22; $P<0.0001$), whereas the MAGGIC risk score was no longer a significant predictor DHFA (Standardized HR=1.15; 95%CI=0.81-1.63; $P=0.43$; 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= 0.557-0.687). The addition of the ML model score increased the Harrel's c index to 0.73 (95%CI= 0.646-0.814). The c index for a model containing the ML score only (0.717; 95%CI=0.643-0.791) 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 multi-marker 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 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 non-linear combinations of biomarkers for risk prediction and found that a multi-marker 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 multi-marker approaches that can be implemented for enhancing risk stratification in this condition.

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), inflammation (Fas, sTNFR1I, 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 above, 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 non-alcoholic fatty liver disease.

Biomarkers as predictors of outcomes

We examined the relationship between circulating levels of biomarkers and prognosis. Various biomarkers significantly predicted DHFA, including two biomarkers related to mineral metabolism/calcification (FGF-23 and OPG), three inflammatory biomarkers (TNF-alpha, sTNFR1 and IL-6), YKL-40 (related to liver injury and inflammation), two 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 NT-proBNP. 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 with incident events. The relationship between inflammatory biomarkers is important, because it supports a role for inflammation in HFpEF, as discussed above. 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 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 above.

We also found that angiopoietin-2, a biomarker for angiogenesis and endothelial dysfunction, was associated with incident DHFA in non-adjusted analyses. This adds to previous studies assessing the significance of biomarkers of angiogenesis in HFpEF. In a study of acute heart failure, biomarkers of angiogenesis, specifically angiogenin, were hubs of biomarker clusters, but only in HFpEF and not HFrEF or HFmrEF.(25) Neupilin, 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 multi-marker 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 US-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, since 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, since 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

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 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, 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 pathologic 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 machine learning 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 multi-marker technology coupled with machine learning provides robust prognostic information.

Conclusion

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

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

This manuscript was prepared using TOPCAT (Treatment of Pre- served Cardiac Function Heart Failure With an Aldosterone Antagonist) 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. We appreciate the technical support from Terry Delmonte and Karl Kammerhoff from the BMS Biorepository.

Funding: this work was founded by an Investigator-Initiated research grant from Bristol-Myers Squibb (JAC) and NIH grant R01HL088577 (TC). JAC is also supported by NIH grants R01-HL 121510-01A1, R61-HL-146390, R01-AG058969, 1R01-HL104106, P01-HL094307, R03-HL146874-01 and R56-HL136730. JHM is supported by NIH grant LM010098. PZ is supported by K23-HL-130551

Abbreviations:

HF	Heart failure
HFpEF	Heart failure with preserved ejection fraction
CI	Confidence Interval
ANOVA	analysis of variance
ANCOVA	analysis of covariance
TOPCAT	Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist
PHFS	Penn Heart Failure Study

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;162:966–972 e10. [PubMed: 22137068]
- 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. [PubMed: 25406305]
- Pitt B, Pfeffer MA, Assmann SF et al. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med* 2014;370:1383–92. [PubMed: 24716680]
- 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;39:1770–1780. [PubMed: 29390051]

5. 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. [PubMed: 21757116]
6. Lakshmi Kannan ARC. Thyroid Dysfunction in Heart Failure and Cardiovascular Outcomes. *Circ Heart Fail* 2018;11.
7. 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;73:1364–1365. [PubMed: 30898214]
8. Hidalgo CA, Klinger B, Barabasi AL, Hausmann R. The product space conditions the development of nations. *Science* 2007;317:482–7. [PubMed: 17656717]
9. Olson RS R U, Andrews PC, Lavender NA, Kidd LC, Moore JH. Automating biomedical data science through tree-based pipeline optimization. *Applications of Evolutionary Computation* 2016:123–137.
10. 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;23:460–471. [PubMed: 29218905]
11. Urbanowicz RJ, Moore JH. ExSTraCS 2.0: Description and Evaluation of a Scalable Learning Classifier System. *Evol Intell* 2015;8:89–116. [PubMed: 26417393]
12. 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. [PubMed: 23095984]
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. [PubMed: 23684677]
14. 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. [PubMed: 26472682]
15. 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. [PubMed: 24484911]
16. Almahmoud MF, Soliman EZ, Bertoni AG et al. Fibroblast Growth Factor-23 and Heart Failure With Reduced Versus Preserved Ejection Fraction: MESA. *J Am Heart Assoc* 2018;7:e008334. [PubMed: 30371180]
17. 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. [PubMed: 26273098]
18. Faul C, Amaral AP, Oskouei B et al. FGF23 induces left ventricular hypertrophy. *J Clin Invest* 2011;121:4393–408. [PubMed: 21985788]
19. 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. [PubMed: 24668259]
20. 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.
21. 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. [PubMed: 25142635]
22. 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. [PubMed: 23598540]
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. [PubMed: 17510463]
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. [PubMed: 27739482]

25. Tromp J, Khan MAF, Mentz RJ et al. Biomarker Profiles of Acute Heart Failure Patients With a Mid-Range Ejection Fraction. *JACC Heart Fail* 2017;5:507–517. [PubMed: 28624483]
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.
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. [PubMed: 25469731]
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. [PubMed: 21750125]

Clinical Perspectives

Competency in Patient Care: In patients with heart failure and preserved left ventricular ejection fraction (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.

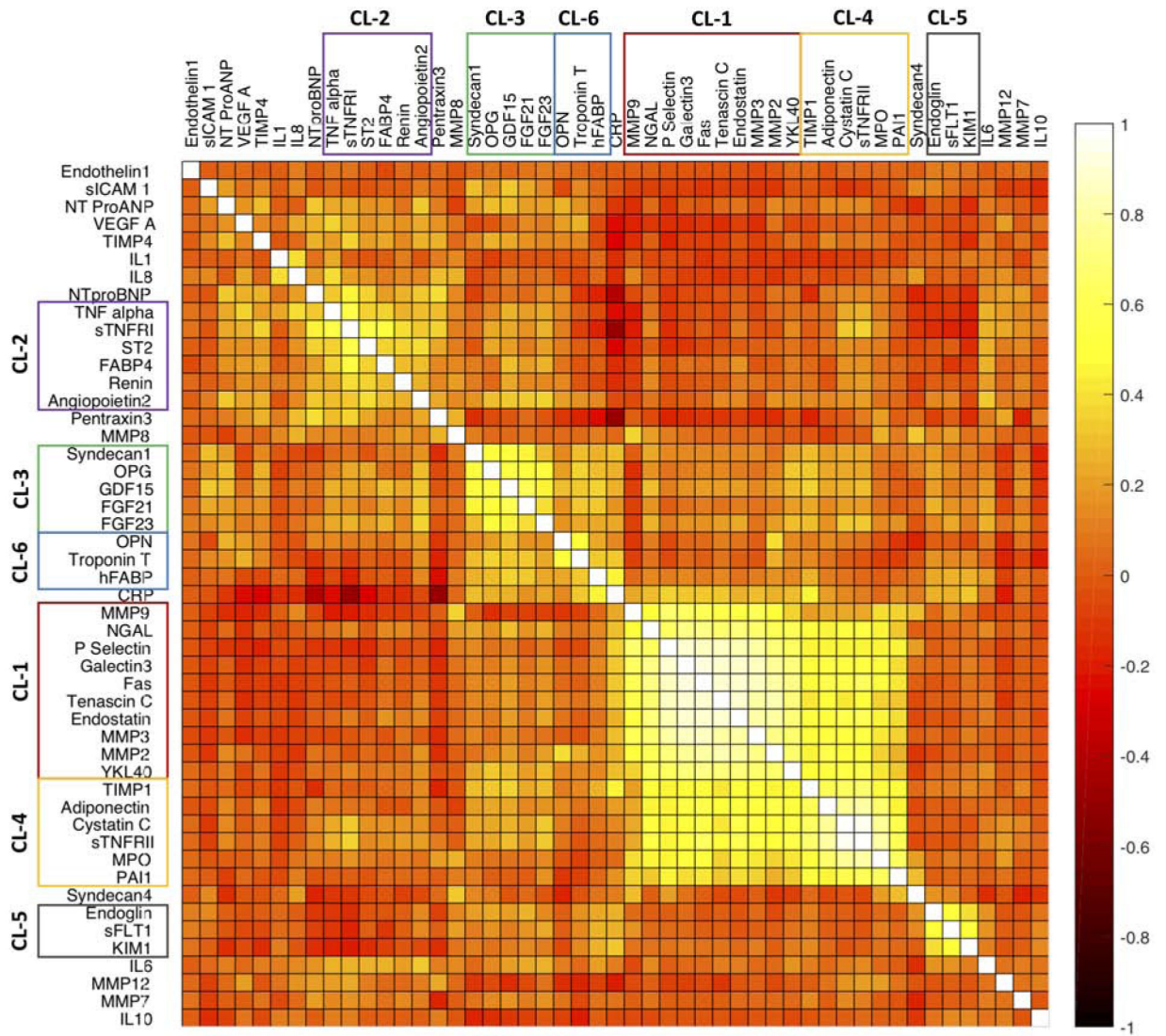


Figure 1. Correlations between biomarkers.

The heatmap represents the correlation between the biomarkers. The most important biomarker clusters, derived from cluster analyses are shown.

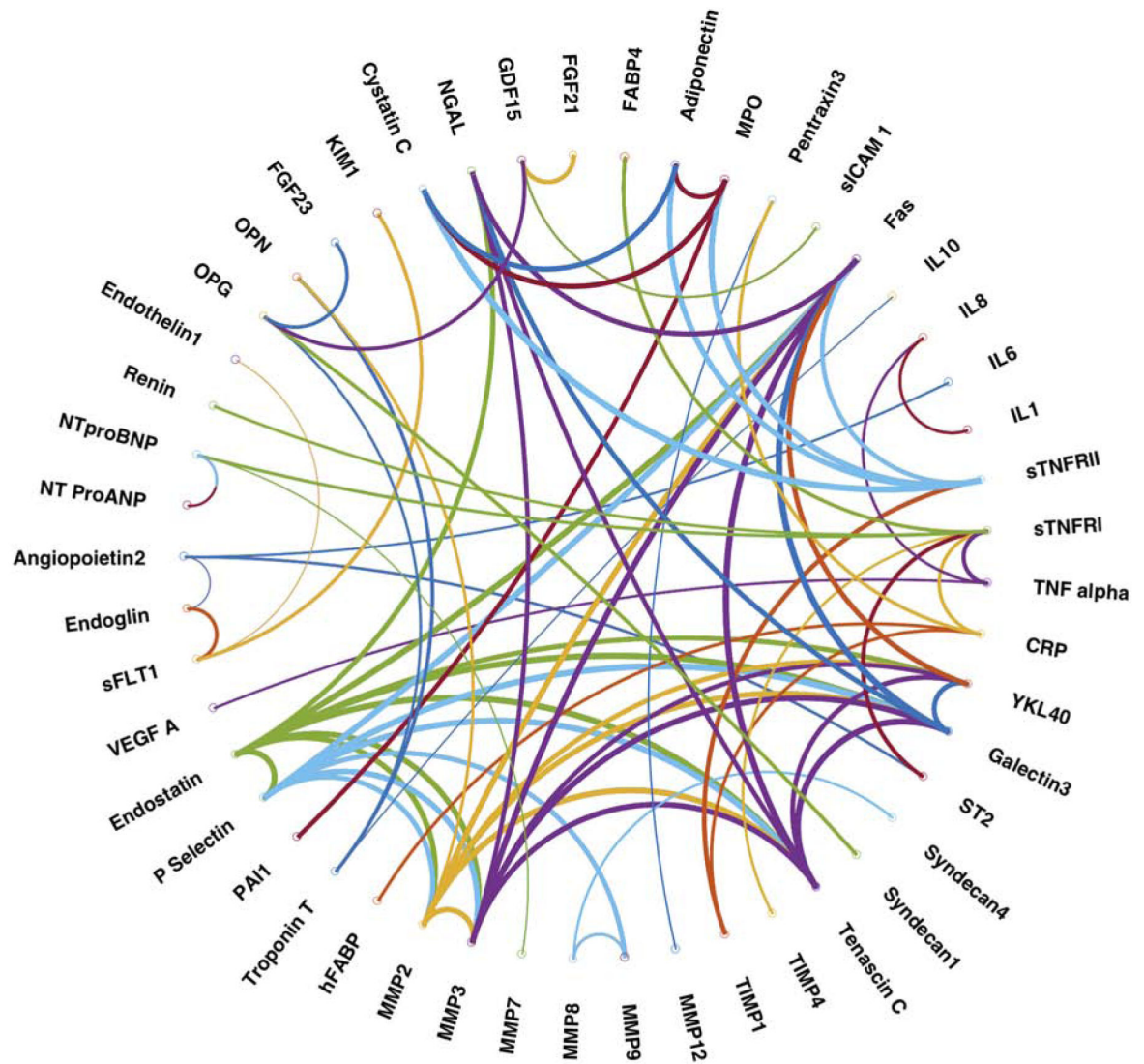


Figure 2. Network connectivity backbone of all measured biomarkers.

The nodes representing individual biomarkers and the edges (connections) between nodes representing the correlation coefficient between a given biomarker (node) pair.

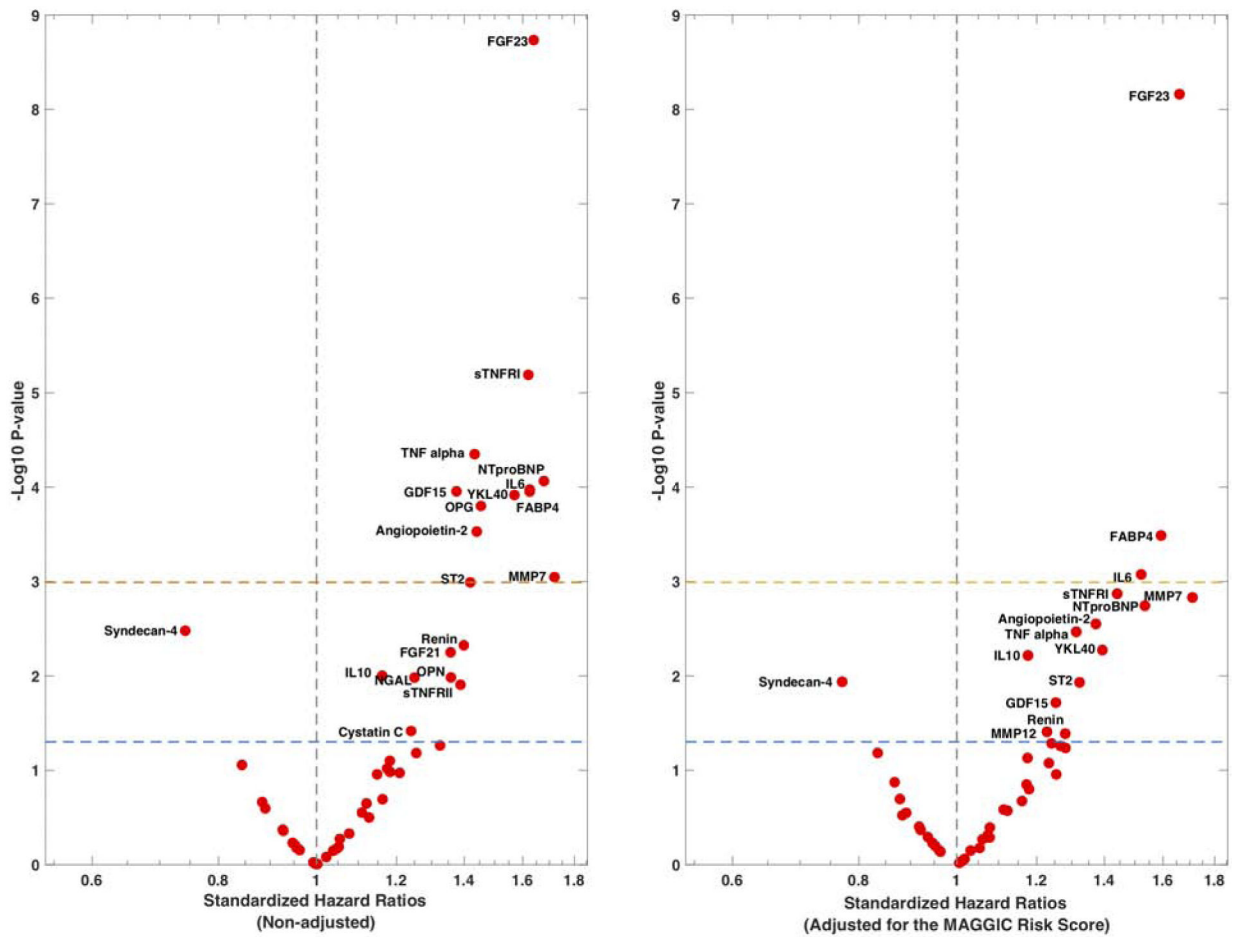


Figure 3. Standardized hazard ratios for examined biomarkers. The volcano plots show the standardized HRs for DHFA (one model per biomarker) in unadjusted analyses (left) and adjusted for the MAGGIC risk score (right), plotted against the Log-10 P-value. The dashed lines indicate the non-corrected (lower line) and Bonferroni-corrected (upper line) level of significance.

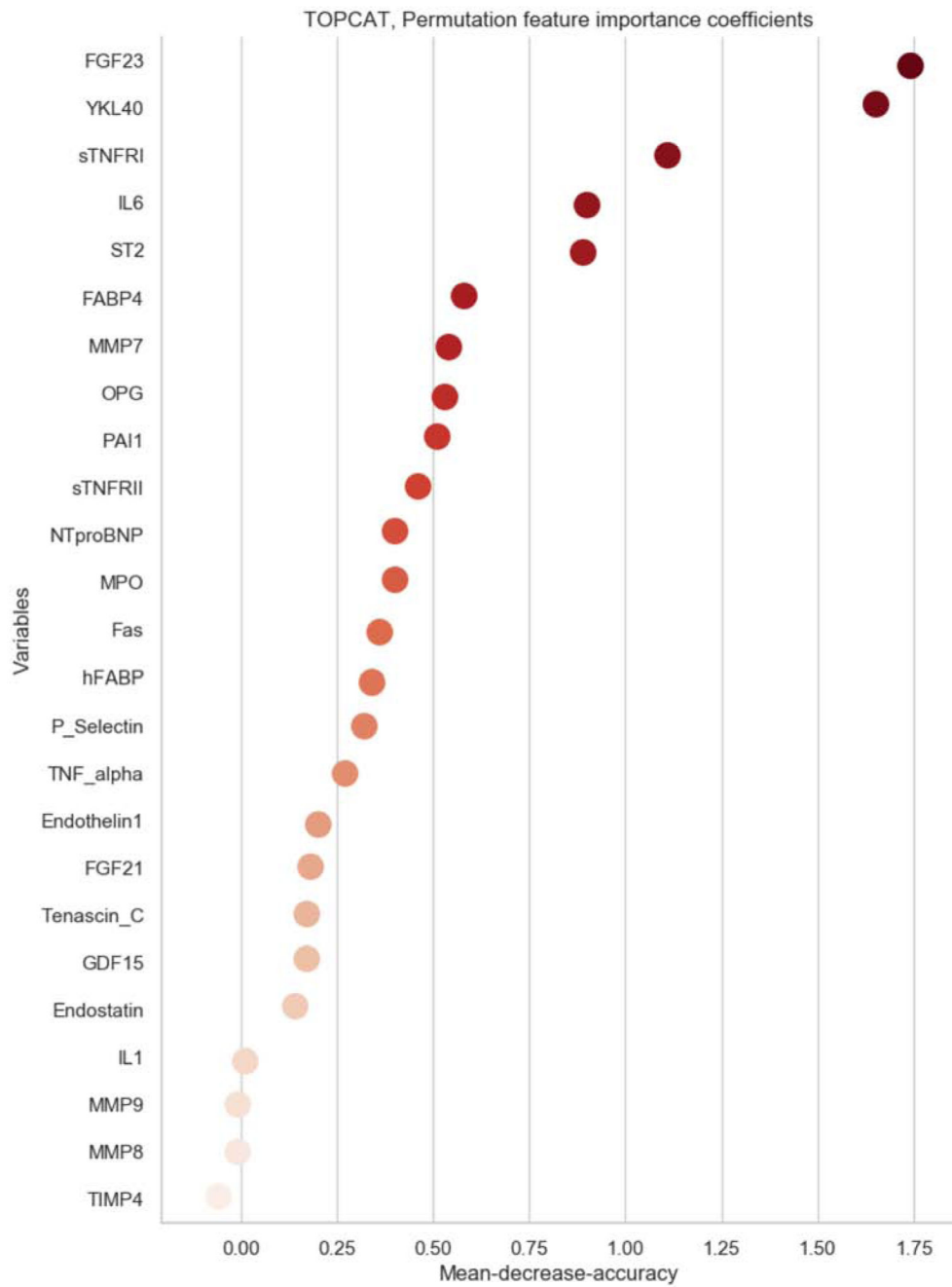


Figure 4. Permutation feature importance coefficients for biomarkers in the machine-learning model.

Biomarkers are ranked according to importance.

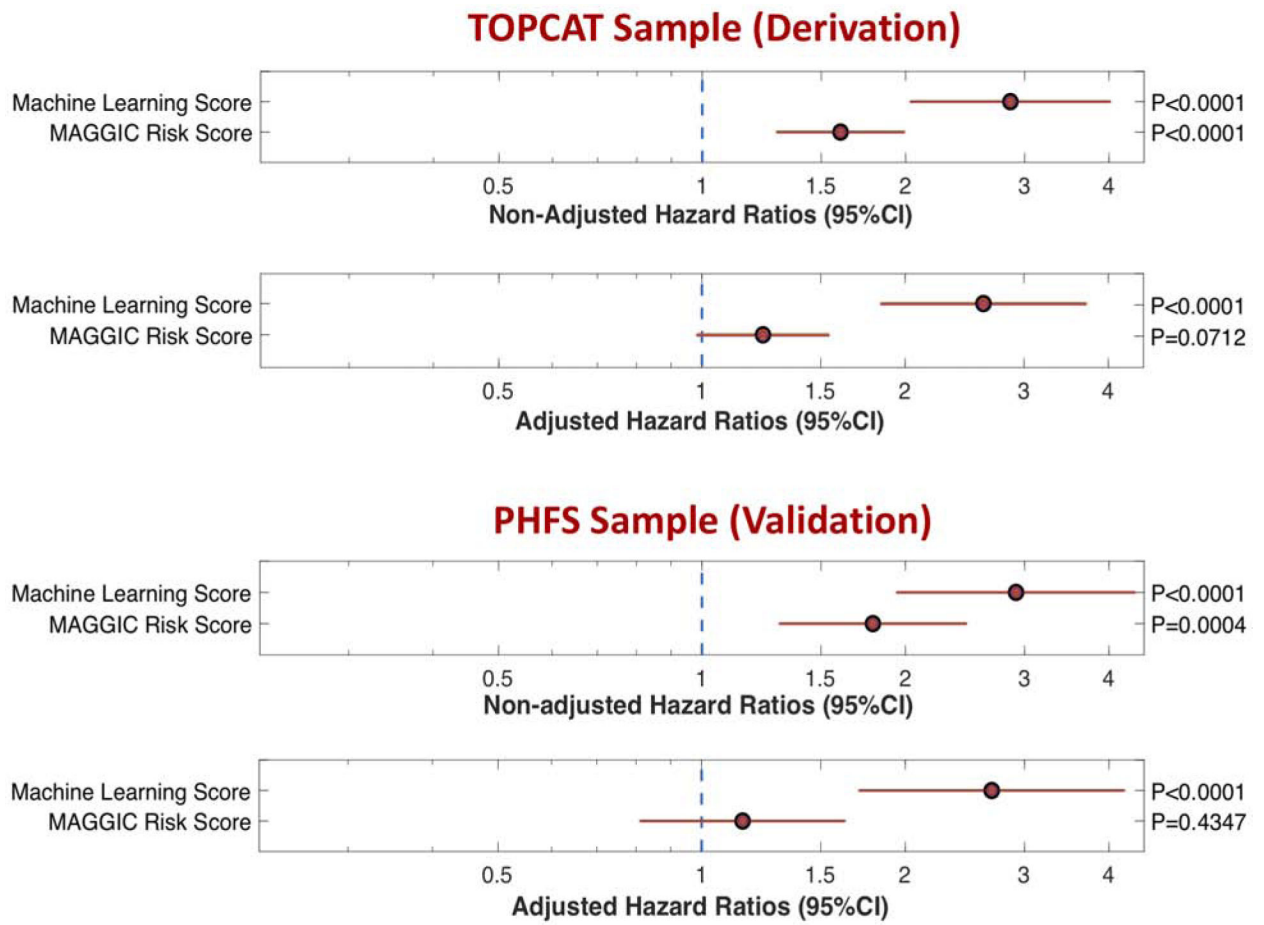


Figure 5. Standardized hazard ratios and 95% CIs for the risk of DHFA. HRs for the machine learning score vs. the MAGGIC risk score are presented for non-adjusted analyses and analyses adjusted for each other, in the derivation (TOPCAT) and validation (PHFS) samples.

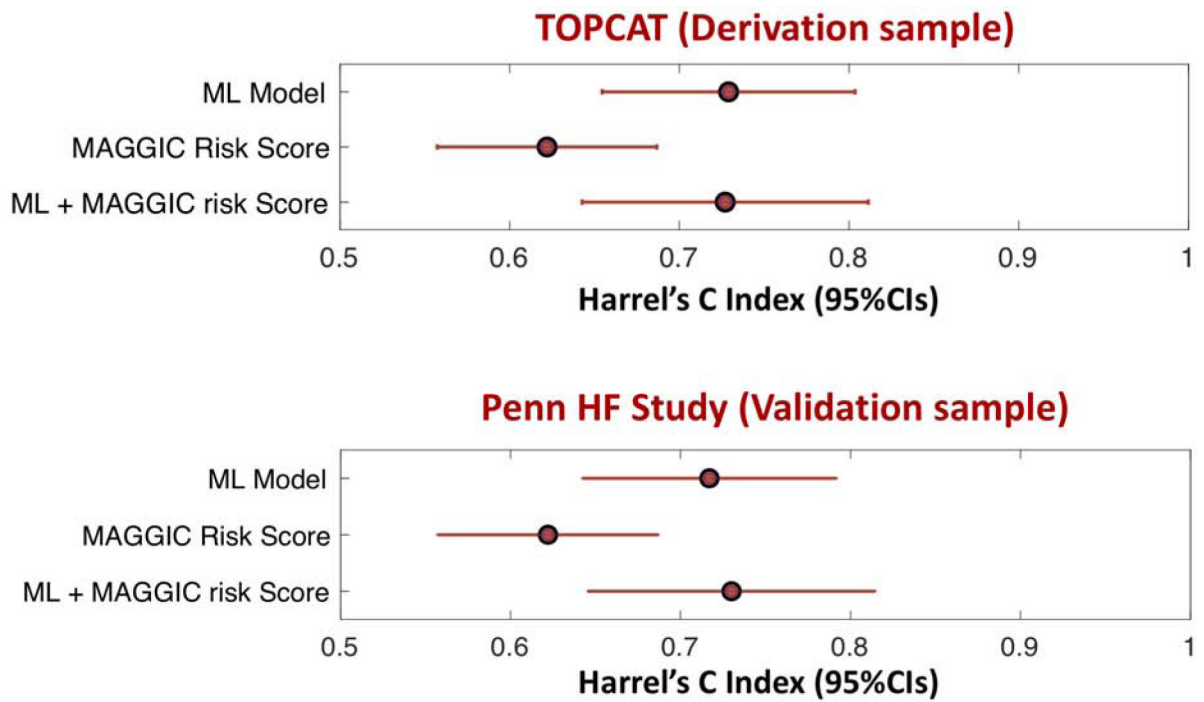
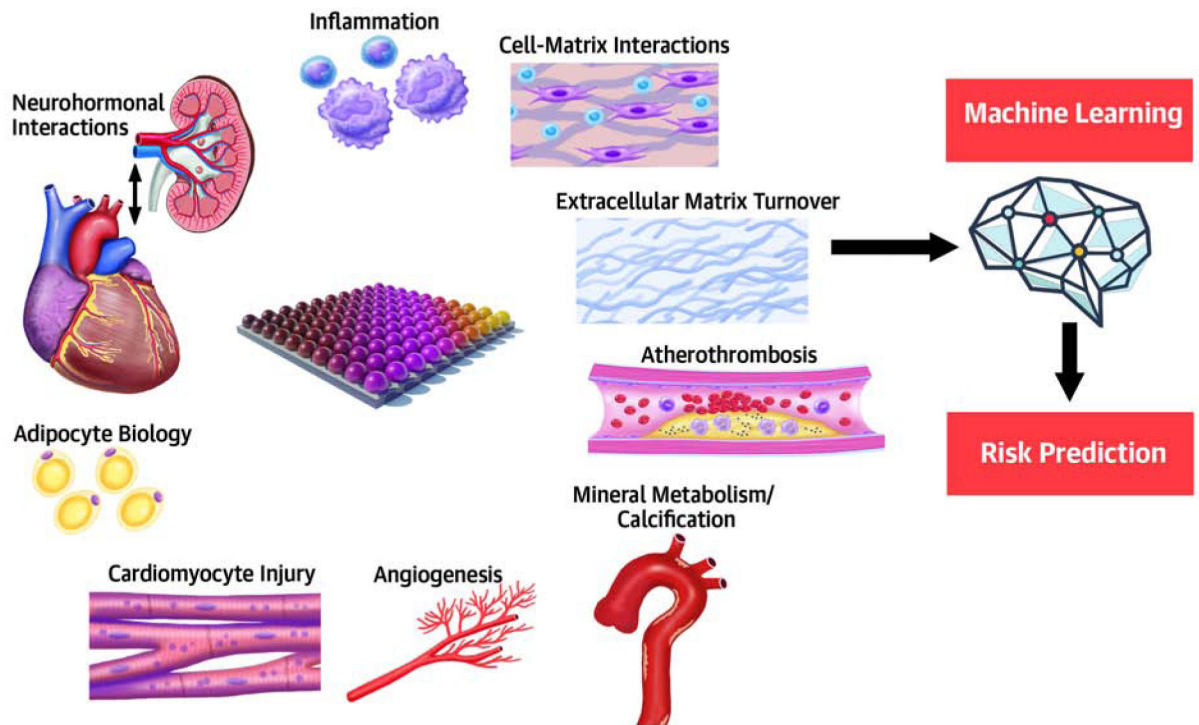


Figure 6. Harrel's concordance statistic (c index) and 95% CIs for the prediction of DHFA. Values are shown for the derivation (TOPCAT) and validation (PHFS) samples.

Multiplex Plasma Protein Measurements



Central Illustration. Multimarker-based Machine Learning Approach For Risk Prediction in Heart Failure with Preserved Ejection Fraction.

We performed multiplex-based measurements of 49 proteins related to key biologic pathways in the TOPCAT 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).

Table 1.Biomarkers included in the Luminex panel (selected *a priori*)

Pathophysiologic domain	Biomarker	Other common names
<i>Angiogenesis</i>	Angiopoietin	
	Endoglin	
	Vascular Endothelial Growth Factor (VEGF) A	
	Soluble fms-like tyrosine kinase-1 (sFLT1)	Soluble VEGF receptor-1
<i>Athero-thrombosis</i>	Endostatin	
	Soluble P-selectin (sP-selectin)	
<i>Cardiomyocyte injury</i>	Plasminogen activator inhibitor-1 (PAI-I)	
	Troponin T	
<i>Extracellular Matrix turnover</i>	Heart-type Fatty Acid Binding Protein (hFABP)	Fatty acid binding protein-3
	Matrix metalloproteinase (MMP)-2	
	MMP-3	
	MMP-7	
	MMP-8	
	MMP-9	
	MMP-12	
	Tissue inhibitor of metalloproteinases (TIMP)-1	
	TIMP-4	
	Tenascin-C	
<i>Cell-matrix interactions</i>	Syndecan-1	
	Syndecan-4	
<i>Tissue remodeling, inflammation and fibrosis</i>	sST-2	Soluble interleukin 1 receptor-like 1
	Galectin-3	
	Liver fibrosis: YKL-40/chitinase 3-like 1 (CHI3L1)	Chitinase-3-like protein-1 (CH3L-1)
<i>Inflammation</i>	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	Chemokine (CXC) ligand 8
	IL-10	
	Fas	Apoptosis antigen-1, CD95
	sICAM	
<i>Adipocyte biology</i>	Pentraxin-3 (PTX-3)	
	Myeloperoxidase (MPO)	
	Adiponectin	
<i>Intermediary Metabolism</i>	Fatty acid binding protein-4 (FABP-4)	
	Fibroblast growth factor 21 (FGF21)	

Pathophysiologic domain	Biomarker	Other common names
<i>Kidney function or injury</i>	Growth differentiation factor-15 (GDF-15)	
	Neutrophil gelatinase-associated lipocalin (NGAL)	Lipocalin-2
	Cystatin-C	
	Kim-1	
<i>Mineral metabolism/Calcification</i>	Fibroblast growth factor 21 (FGF21)	T-cell immunoglobulin and mucin domain 1 (TIM-1), Hepatitis A virus cellular receptor 1 (HAVcr-1)
	Osteopontin	
	Osteoprotegerin	
<i>Neurohormonal regulation and myocyte stretch</i>	Endothelin-1	
	Renin	
	NT-pro-BNP	
	NT-pro-ANP	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

General characteristics of study participants with vs. without available plasma samples. Numbers represent Mean (SD), Median (IQR) or counts (%).

	Participants without available samples (n=3063)	Participants with available samples (n=379)	P value
Demographic Characteristics			
<i>Age, years</i>	69 (61,76)	70 (62,77)	0.0298
<i>Male Sex</i>	1465 (47.83%)	203 (53.56%)	0.0351
<i>Race</i>			0.115
<i>White</i>	2712 (88.54%)	347 (91.56%)	
<i>Black</i>	274 (8.95%)	28 (7.39%)	
<i>Other</i>	77 (2.5%)	4 (1.1%)	
<i>BMI, kg/m²</i>	30.8 (27.1,35.7)	31.8 (27.8,36)	0.0384
<i>Heart rate, bpm</i>	68 (62,76)	68 (60,73.8)	0.0080
<i>Systolic BP, mmHg</i>	130 (120,140)	126 (120,135)	<0.0001
<i>Diastolic PB, mmHg</i>	80 (70,81)	75 (66,80)	<0.0001
<i>Country</i>			<0.0001
<i>United States</i>	980 (32%)	160 (44%)	
<i>Canada</i>	273 (8.9%)	53 (14%)	
<i>Russia</i>	908 (29.6%)	157 (41.4%)	
<i>Georgia, Brazil, Argentina</i>	902 (29.4%)	0	
Medical History			
<i>NYHA class III-IV</i>	992 (32.42%)	144 (37.99%)	0.0295
<i>Myocardial Infarction</i>	777 (25.38%)	116 (30.61%)	0.0284
<i>Stroke</i>	235 (7.67%)	30 (7.92%)	0.8682
<i>COPD</i>	358 (11.69%)	45 (11.87%)	0.9174
<i>Hypertension</i>	2788 (91.05%)	358 (94.46%)	0.0254
<i>Peripheral Arterial Disease</i>	281 (9.18%)	38 (10.03%)	0.5907
<i>Atrial Fibrillation</i>	1051 (34.32%)	162 (42.74%)	0.0012
<i>Diabetes Mellitus</i>	990 (32.33%)	128 (33.77%)	0.5720
Medication Use			
<i>Beta Blockers</i>	2375 (77.56%)	301 (79.42%)	0.4124
<i>Calcium Channel Blockers</i>	1159 (37.85%)	134 (35.36%)	0.3441
<i>Diuretics</i>	2504 (81.78%)	312 (82.32%)	0.7951
<i>Glucose-lowering agents</i>	849 (27.73%)	113 (29.82%)	0.3928
<i>ACE Inhibitors or ARBs</i>	2593 (84.68%)	306 (80.74%)	0.0468
<i>Statins</i>	1555 (50.78%)	250 (65.96%)	<0.0001