University of Groningen

Multiomics Analysis Provides Novel Pathways Related to Progression of Heart Failure

Ouwerkerk, Wouter; Belo Pereira, Joao P.; Maasland, Troy; Emmens, Johanna E.; Figarska, Sylwia M.; Tromp, Jasper; Koekemoer, Andrea L.; Nelson, Christopher P.; Nath, Mintu; Romaine, Simon P.R.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

[Link to publication in University of Groningen/UMCG research database](#)

Multiomics Analysis Provides Novel Pathways Related to Progression of Heart Failure

Wouter Ouwerkerk, PHD,[a,b,*] Joao P. Belo Pereira, PHD,[c,d,*] Troy Maasland, MSC,[c,d,e] Johanna E. Emmens, MD, PHD,[f] Sylwia M. Figarska, PHD,[f] Jasper Tromp, MD, PHD,[f,g,h] Andrea L. Koekemoer, PHD,[i,j] Christopher P. Nelson, PHD,[i,j] Mintu Nath, PHD,[k] Simon P.R. Romaine, MBCHB,[i,j] John G.F. Cleland, MD, PHD,[l,m] Faiez Zannad, MD, PHD,[n,o,p] Dirk J. van Veldhuisen, MD, PHD,[f] Chim C. Lang, MD, PHD,[q] Piotr Ponikowski, MD, PHD,[r] Gerasimos Filippatos, MD, PHD,[s] Stefan Anker, MD, PHD,[t,u,v] Marco Metra, MD, PHD,[w] Kenneth Dickstein, MD, PHD,[x] Leong L. Ng, MD, PHD,[i,j] Rudolf A. de Boer, MD, PHD,[f] Natal van Riel, PHD,[e,y] Max Nieuwdorp, MD, PHD,[y] Albert K. Groen, PHD,[y] Erik Stroes, MD, PHD,[y] Aeilko H. Zwinderman, PHD,[z] Nilesh J. Samani, MD, PHD,[i,j] Carolyn S.P. Lam, MD, PHD,[b] Evgeni Levin, PHD,[c,d,†] Adriaan A. Voors, MD, PHD[f,†]

ABSTRACT

BACKGROUND Despite major advances in pharmacological treatment for patients with heart failure, residual mortality remains high. This suggests that important pathways are not yet targeted by current heart failure therapies.

OBJECTIVES We sought integration of genetic, transcriptomic, and proteomic data in a large cohort of patients with heart failure to detect major pathways related to progression of heart failure leading to death.

METHODS We used machine learning methodology based on stacked generalization framework and gradient boosting algorithms, using 54 clinical phenotypes, 403 circulating plasma proteins, 36,046 transcript expression levels in whole blood, and 6 million genomic markers to model all-cause mortality in 2,516 patients with heart failure from the BIOSTAT-CHF (Systems BIOlogy Study to TAilored Treatment in Chronic Heart Failure) study. Results were validated in an independent cohort of 1,738 patients.

RESULTS The mean age of the patients was 70 years (Q1-Q3: 61-78 years), 27% were female, median N-terminal pro-B-type natriuretic peptide was 4,275 ng/L (Q1-Q3: 2,360-8,486 ng/L), and 7% had heart failure with preserved ejection fraction. During a median follow-up of 21 months, 657 (26%) of patients died. The 4 major pathways with a significant association to all-cause mortality were: 1) the PI3K/Akt pathway; 2) the MAPK pathway; 3) the Ras signaling pathway; and 4) epidermal growth factor receptor tyrosine kinase inhibitor resistance. Results were validated in an independent cohort of 1,738 patients.

CONCLUSIONS A systems biology approach integrating genomic, transcriptomic, and proteomic data identified 4 major pathways related to mortality. These pathways are related to decreased activation of the cardioprotective ERBB2 receptor, which can be modified by neuregulin. (J Am Coll Cardiol 2023;82:1921–1931) © 2023 by the American College of Cardiology Foundation.

ABBREVIATIONS AND ACRONYMS

ACE = angiotensin-converting enzyme

ARB = angiotensin receptor blocker

EGFR = epidermal growth factor receptor

GO = Gene Ontology

HFrEF = heart failure with reduced ejection fraction

Despite all recent advances in pharmacological treatment of heart failure, residual mortality remains high. This suggests that there are still important pathways not adequately targeted by current guideline therapies.[1]

Capturing these pathways requires integration of genetic, transcriptomic, proteomic, and phenotypic markers using a systems biology -omics approach.[2] With recent advancements in bioinformatics,[3] and high-throughput -omics data, the integration and subsequent interpretation of multiple high-dimensional -omic data sets have become increasingly feasible tools for revealing novel biological insights. Previous efforts revealed putative markers related to pathological lipid abundance[4] and cancer.[5]

Such a deep analysis of heart failure requires an enormous repository of data with robust and reproducible observations that can also be validated in an independent population.[6] The BIOSTAT-CHF (Systems BIOlogy Study to TAilored Treatment in Chronic Heart Failure) study has already generated multiple new insights partially unraveling processes on single -omics data.[7-16] However, integration of multiple -omics data sets in the BIOSTAT-CHF study has never been done before.

Here, we present a multiomics approach integrating phenotypic, proteomic, transcriptomic, and genetic data sets using advanced machine learning. The primary aim of this study is to identify, validate, and understand disease pathways that are associated with progression of heart failure leading to early mortality.

METHODS

PATIENT POPULATION AND STUDY DESIGN. The BIOSTAT-CHF study was designed to identify pathophysiological pathways related to heart failure progression using a systems biology approach on multiomics data. The design and baseline characteristics of this study have been previously reported.[17] Briefly, the BIOSTAT-CHF study consists of 2 independent (index and validation) cohorts. Inclusion criteria were similar in both cohorts. The index cohort consisted of 2,516 patients with worsening signs and/or symptoms of heart failure, included from 69 centers in 11 European countries during 2010-2014. The validation cohort consisted of a comparable cohort of 1,738 patients from 6 centers in Scotland, United Kingdom. Patients were enrolled as inpatients or outpatients, with a median follow-up in each cohort of 21 months (Q1-Q3: 15-27 months). The endpoint of interest for the present study was 1-year all-cause mortality. Patient characteristics of both cohorts are presented in the Supplemental Appendix.[17]

The study complied with the Declaration of Helsinki and was approved by the participating centers' medical ethics committees. All patients provided written informed consent.

The BIOSTAT-CHF study has generated a large repository of clinical phenotypic data and biological data over the years. Biological data consisted of high-dimensional proteomic, transcriptomic, and genetic data.

SAMPLE HANDLING. Samples were collected and centrifuged and stored at preferably −80 °C or −70 °C, else −20 °C for a maximum of 1 month. Shipments from the site to the central biobank were done periodically.

Blood was drawn by venipuncture. First, plastic vacutainer tubes were used to collect approximately 83.5 mL of blood. Second, the following samples were collected: 1) EDTA plasma, 56 samples of 10 mL content each; 2) serum with gel, 12 samples of 8.5 mL

Initiative-Cardiovascular and Renal Clinical Trialists, French Institute of Health and Medical Research, Vandoeuvre-lès-Nancy, France; [q]Cardiology, Ninewells Hospital and Medical School, Dundee, United Kingdom; [r]Institute for Heart Diseases, Medical University, Wroclaw, Poland; [s]Attikon University Hospital, National and Kapodistrian University of Athens, Athens, Greece; [t]Department of Cardiology, Charité Universitätsmedizin Berlin, Berlin, Germany; [u]Berlin Institute of Health Center for Regenerative Therapies, Charité Universitätsmedizin Berlin, Berlin, Germany; [v]German Centre for Cardiovascular Research, partner site Berlin, Charité Universitätsmedizin Berlin, Berlin, Germany; [w]Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, Institute of Cardiology, University of Brescia, Brescia, Italy; [x]Stavanger University Hospital, University of Bergen, Stavanger, Norway; [y]Department of Vascular Medicine, Amsterdam UMC, Amsterdam, the Netherlands; and the [z]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands. *Drs Ouwerkerk and Bereira contributed equally to this work as joint first authors. †Drs Levin and Voors contributed equally to this work as joint senior authors.

content; 3) and PAXgene Blood RNA tubes, 2 samples of approximately 2.5 mL content each.

Details on the entire process of sample handling is described in the Laboratory Handbook in the Supplemental Appendix.

PHENOTYPIC (CLINICAL) PANEL. We collected 54 clinical markers in the BIOSTAT-CHF study (see Phenotypic Parameters in the Supplemental Appendix). Phenotypic data consisted of demographic data (eg, age, sex, medical history, comorbidities) and data derived during physical examination (eg, body mass index, systolic and diastolic blood pressure, left ventricular ejection fraction). The estimated glomerular filtration rate (eGFR) is calculated using the Chronic Kidney Disease Epidemiology Collaboration formula.[18] Clinical data were coordinated by the Trial Coordination Center, a contract research organization affiliated with the University Medical Center Groningen. A local investigator collected clinical data and was monitored by the Trial Coordination Center. Data were electronically stored in a centralized database at the University Medical Center Groningen.

PROTEIN PANEL FROM PERIPHERAL BLOOD: PERFORMED BY OLINK AND LOCALLY AT CENTER LABORATORIES. We measured 403 serum/plasma biomarkers (see Protein Listings in the Supplemental Appendix) from several pathophysiological domains, including markers of inflammation, apoptosis, remodeling, myocyte stress/injury, angiogenesis, endothelial function, and several markers of renal function. The protein biomarker data used for this study have been described in recent papers.[7-9] In brief, the biomarkers included standard biochemical blood parameters (eg, hemoglobin, hematocrit, blood urea nitrogen, heart failure–related markers [N-terminal pro-B-type natriuretic peptide and B-type natriuretic peptide]). In addition, 4 biomarker panels comprising each of 92 protein biomarkers provided by the Olink Bioscience analysis service were measured. These respective panels were cardiovascular II, cardiovascular III, immune response, and oncology II panels. The proteins were profiled using Olink Proseek Multiplex Inflammatory 96×96 platform. The Proseek kit uses proximity extension assay technology, whereby oligonucleotide-labeled antibody probe pairs bind to their respective targets. Quantification was achieved using a Fluidigm BioMark real-time polymerase chain reaction platform. The platform provides normalized protein expression (log2-normalized), rather than an absolute quantification.

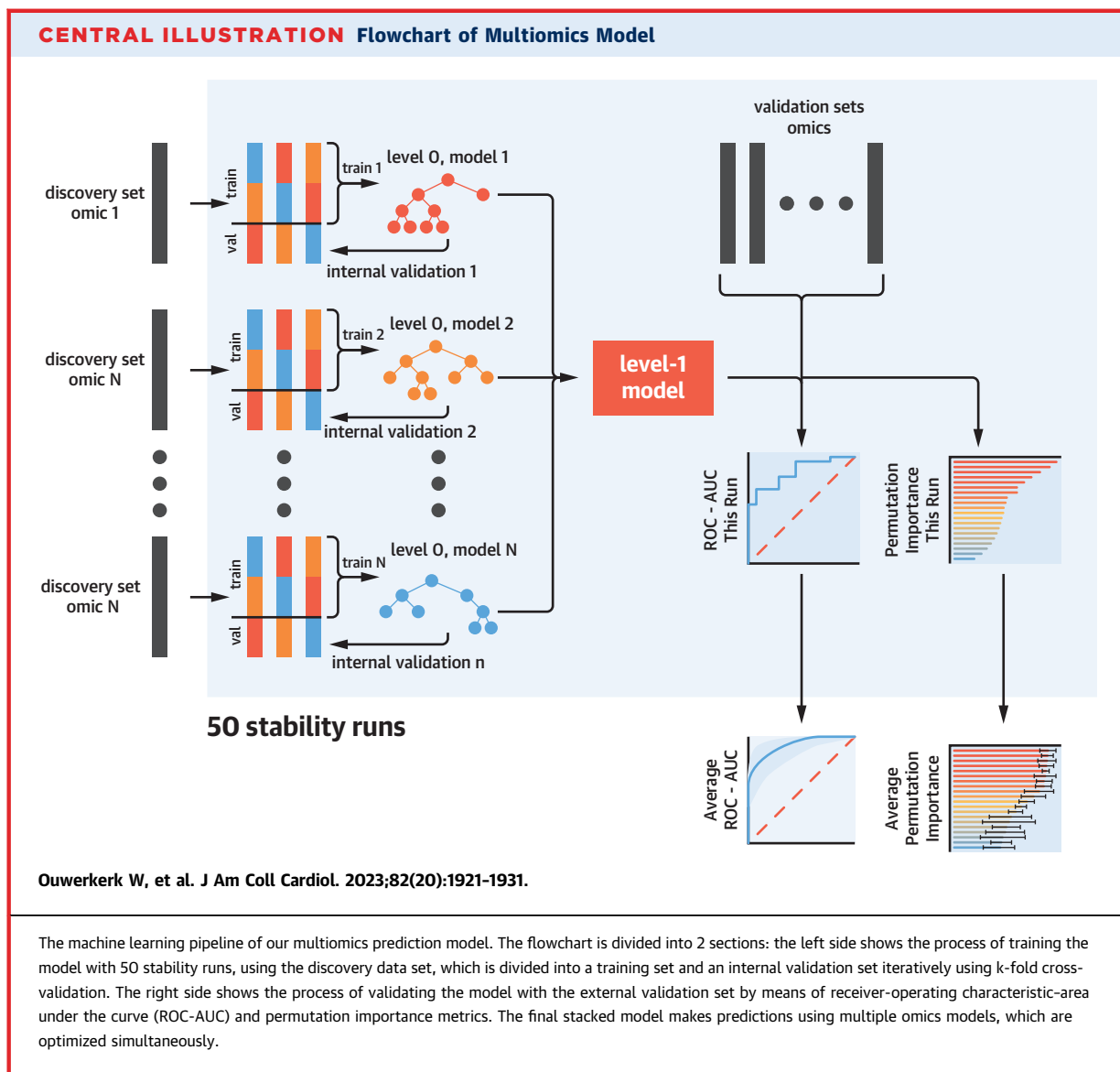**TRANSCRIPTOMIC PANEL: PERFORMED BY UNIVERSITY OF LEICESTER.** Whole blood transcriptomic profiles from 944 patients (626 survivors and 318 non-survivors who died from cardiovascular causes) from the index cohort were obtained using the GeneChip Human Transcriptomic Array 2.0 (HTA 2.0) developed by Affymetrix, Inc (part of Thermo Fisher Scientific). Patients were age- and sex-matched. Details on the protocols and methodology used to assess and confirm the quality of the raw transcriptomic data and the processes used to integrate signals from individual probes on the array to determine the expression levels of each gene and to assess the quality of the summarized RNA expression set data have been previously published.[19] In total, 36,046 (17,924 protein coding and 18,122 nonprotein coding) transcripts were analyzed.

GENOMIC PANEL: PERFORMED BY UNIVERSITY OF LEICESTER. Both cohorts were processed, genotyped, quality controlled, and imputed independently using identical protocols.[12] Genotyping of all patients was performed using the Affymetrix Axiom Genome-Wide UKB WCSG genotyping array. Sample-level quality control was performed for X chromosome homozygosity (sex mismatch) and identity by descent estimates (relatedness and duplicates). Before imputation, variants were removed if their call rate was <95% for variants with minor allele frequency ≥5% or <99% with minor allele frequency <5%, or a Hardy-Weinberg equilibrium $P < 1 \times 10^{-6}$. Imputation was performed using SHAPEIT2[20] and IMPUTE2[21] with the phase 3 release 1000G reference panel.[22]

STATISTICAL ANALYSES. We used machine learning methods, particularly gradient boosting (with tailored loss functions), with stacked regularization,[23] to model all-cause mortality. This method combines multiple *-omics* data in a nonlinear manner by learning how to combine predictions given by models trained on the individual *-omics* sets into a single coherent output.

It is specifically designed, in contrast to standard modern statistical methods (Benefits of Machine Learning for Multi-*omics* Analysis in the Supplemental Appendix), to handle not only high-dimensional *-omics* data, in which the number of patients is significantly smaller than the number of variables (n<<p), but also when different data sources are collectively used to estimate the core mechanism present in all data sources.

In brief, we used a combination of stacking generalization framework with multiple gradient boosting classifiers to improve prediction accuracy. For each *-omics* set, we built a level 0 model. These level 0 models were subsequently combined to form

CENTRAL ILLUSTRATION Flowchart of Multiomics Model

Ouwerkerk W, et al. J Am Coll Cardiol. 2023;82(20):1921–1931.

The machine learning pipeline of our multiomics prediction model. The flowchart is divided into 2 sections: the left side shows the process of training the model with 50 stability runs, using the discovery data set, which is divided into a training set and an internal validation set iteratively using k-fold cross-validation. The right side shows the process of validating the model with the external validation set by means of receiver-operating characteristic-area under the curve (ROC-AUC) and permutation importance metrics. The final stacked model makes predictions using multiple omics models, which are optimized simultaneously.

the level 1 model.[23] This allows us to use all data available in each panel (eg, phenotype data from all 2,516 patients are used to create the level 0 model of the phenotype panel, and the level 0 transcriptomic model was estimated using 944 patients). A figure visually summarizing this approach can be seen in the **Central Illustration**.

One of the challenges in machine learning is tuning the various models' hyperparameters. Typically, each model is optimized separately, leading to a local optimum. We optimized all the models simultaneously using Bayesian optimization,[24] achieving a global optimum. To avoid overfitting, we used stratified cross-validation over the training partition.[25]

To ensure the feature signatures' reliability and robustness, we conducted stability selection.[26] The complete analysis was repeated 50 times. Receiver-operating characteristic-area under curves were computed each time and averaged over the repeated analyses in both the index and validation cohorts. A permutation (randomization test)[27] was used to evaluate the results' statistical validity.

We evaluated the model's quality separately in the validation cohort. In the validation cohort, transcriptomic data were not measured and did not include the corresponding level 0 model. Nevertheless, our approach can validate the results of the other phenotype, protein, and genomic panels.

PATHWAY ENRICHMENT. To identify novel pathways related to mortality, we performed an overrepresentation analysis. We determined the effect of

TABLE 1 Baseline Demographics of the Index and Validation Cohorts

	Index			Validation		
	Alive (n = 1,859, 74%)	Died (n = 657, 26%)	P Value	Alive (n = 1,214, 75%)	Died (n = 401, 25%)	P Value
Age, y	68.0 ± 11.9	73.0 ± 11.2	<0.0001	73.0 ± 10.5	78.0 ± 9.7	<0.0001
Male	1,370 (74)	476 (72)	0.57	801 (66)	270 (67)	0.66
LVEF, %	31.0 ± 9.8	32.0 ± 12.5	0.03	41.0 ± 13	41.0 ± 13.3	0.63
BMI, kg/m^2	28.0 ± 5.5	27.0 ± 5.5	0.001	29.0 ± 6.4	28.0 ± 6.1	<0.0001
Ischemic heart disease	946 (51)	412 (63)	<0.0001	776 (64)	286 (71)	0.008
Heart failure hospitalization in last year	531 (29)	263 (40)	<0.0001	301 (25)	130 (32)	0.003
Myocardial infarction	657 (35)	306 (47)	<0.0001	575 (48)	223 (56)	0.006
DM	577 (31)	242 (37)	0.007	367 (30)	155 (39)	0.002
COPD	279 (15)	157 (24)	<0.0001	191 (16)	104 (26)	<0.0001
History of renal disease	402 (22)	294 (45)	<0.0001	491 (41)	241 (61)	<0.0001
NYHA functional class I	50 (2)	6 (1)	<0.0001	15 (1)	1 (0)	<0.0001
NYHA functional class II	711 (39)	157 (24)		575 (47)	92 (23)	
NYHA functional class III	853 (47)	375 (58)		516 (43)	200 (50)	
NYHA functional class VI	196 (10)	98 (15)		108 (9)	107 (27)	

Values are mean ± SD or n (%).
BMI = body mass index; COPD = chronic obstructive pulmonary disease; DM = diabetes mellitus; LVEF = left ventricular ejection fraction.

each variant using Ensembl Variant Effect Predictor[28] and converted all selected markers to Ensembl IDs.

We performed enrichment using Gprofiler for Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, Reactome, CORUM, and WikiPathways pathways. We report a corrected value of $P < 0.05$ as significant.

To test statistical differences between patients with and without activated pathways and their association to up-titration, we used principal component analysis, using missMDA,[29,30] to reduce the dimensionality of the biomarkers present in the pathways, in a similar manner as previously published.[10] A weighted score (first principal component) was generated to which each selected biomarker contributed to a greater or lesser extent, based on how much population variance they explained. First, we used the weighted score to identify associated clinical characteristics in a penalized linear regression.[31] Second, we performed multivariable regression analyses including percentage achieved target doses of angiotensin-converting enzyme (ACE) inhibitors/angiotensin receptor blockers (ARBs), beta-blockers, and mineralocorticoid receptor antagonists,[1] and age and sex. For this analysis, we only included heart failure with reduced ejection fraction (HFrEF) patients. Last, we associated activation of pathways to all-cause mortality, using Cox proportional hazards modeling.

Data are presented as mean ± SD when normally distributed, median (Q1-Q3) for skewed variables, and frequency and percentage for categorical variables. Differences between patients who died and those who did not in the index and validation cohort were tested using the Student's independent Student's t-test or Mann-Whitney U test, where appropriate, for continuous parameters. Differences in categorical variables were tested with chi-square tests.

We used Python version 3.8 (Python Software Foundation), with packages Numpy, Scipy, and scikit-learn for implementing the stacking model and R version 4.0 (R Foundation for Statistical Computing) for visualizations.

RESULTS

CLINICAL CHARACTERISTICS. Data were available for 2,516 patients in the phenotypic and protein panels, 944 in the transcriptomic and 2,470 in the genomic panels in the index cohort (Supplemental Figure 1A). The validation cohort had data available for 1,738 patients in the phenotypic and protein panels and for 1,693 patients in the genomic panel (Supplemental Figure 1B).

During a median follow-up of 21 months (Q1-Q3: 11-32 months) and 21 months (Q1-Q3: 15-27 months), 657 (26%) and 501 (32%) patients died in the index and validation cohorts, respectively. Baseline characteristics of the patients who died and those who survived in the index and validation cohorts are presented in Table 1. Patients who died in the index cohort were older (73 ± 11 years vs 68 ± 12 years; $P < 0.001$), had a higher NYHA functional class

FIGURE 1   The 15 Most Predictive Variables of Mortality in Each Panel
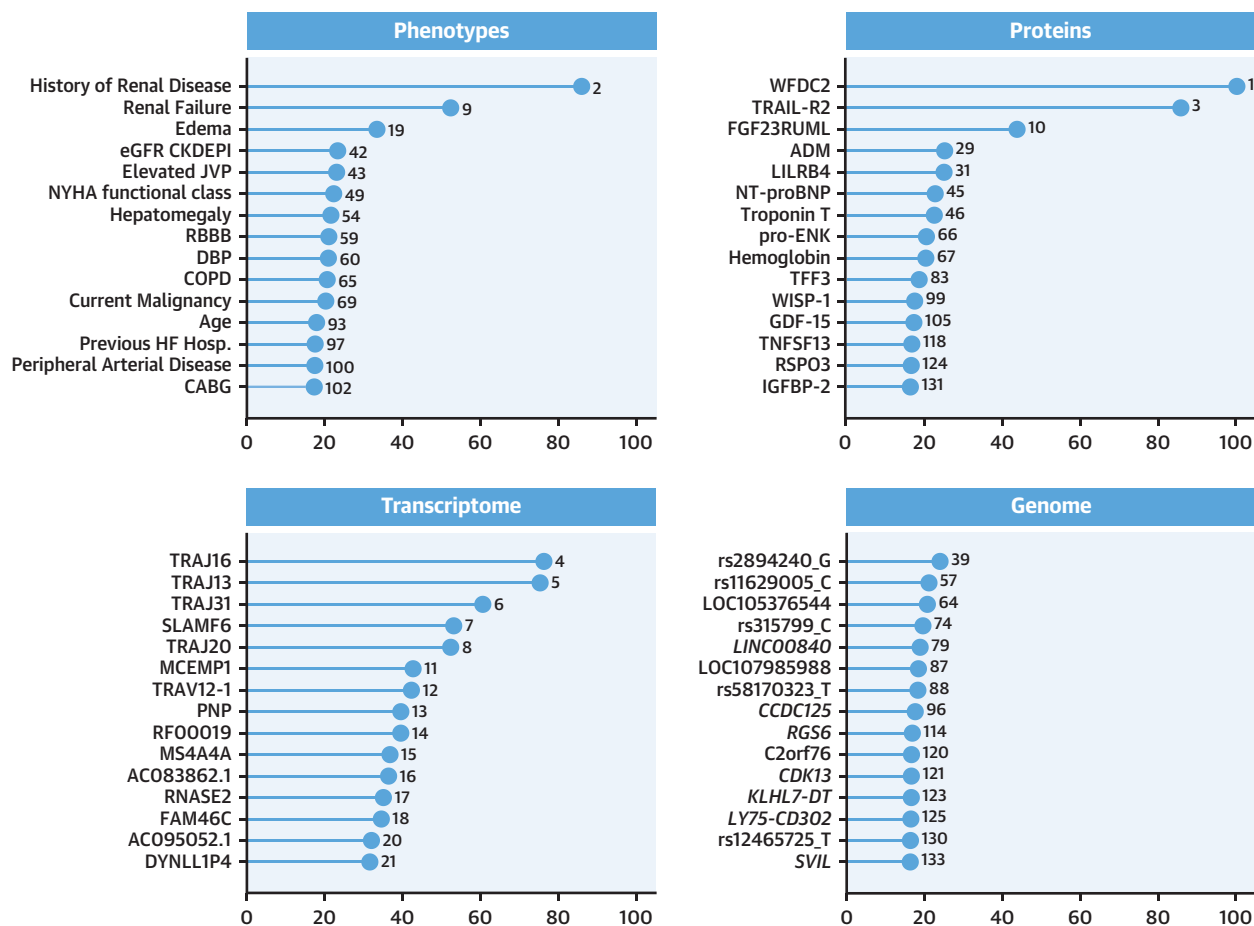
For every panel, the top 15 markers are shown in the lollipop plots including their overall ranking. The overall importance ranking was calculated by scaling all relative importance to the importance of each panel. CABG = coronary artery bypass grafting; COPD = chronic obstructive pulmonary disease; CKDEPI = Chronic Kidney Disease Epidemiology Collaboration; DBP = diastolic blood pressure; eGFR = estimated glomerular filtration rate; Hosp. = hospitalization; JVP = elevated jugular venous pressure; NT-proBNP = N-terminal pro-B-type natriuretic peptide; RBBB = right bundle branch block.

(III/IV 74% vs 58%; $P < 0.001$), and more comorbidities. These differences were similar in the validation cohort (**Table 1**).

MULTIOMICS MORTALITY MODEL. Our final risk prediction model, combining phenotypic, proteomic, transcriptomic, and genomic data achieved a significant receiver-operating characteristic-area under curve value of 0.81 ± 0.02 in the stratified cross-validated part of the index cohort and 0.85 ± 0.03 in the validation cohort (Supplemental Figure 2), both $P < 0.001$ in permutation tests (Supplemental Figure 3A).

This optimal model consisted of 60 markers per panel with a total of 240 markers, all associated with mortality and closely related to each other

(Supplemental Table 1, Supplemental Figure 3B). The relative importance of the top 15 markers for each data set and in the level 1 model is visualized in **Figure 1**. The direction of the association between each marker and mortality is presented in the spider plot of Supplemental Figure 4.

ENRICHMENT. We performed overrepresentation analysis, using 180 markers (60 proteomic, 60 transcriptomic, and 60 genomic) (see Supplemental Table 1 for selected markers from our level 1 model set against all the markers in our data set). We found that there were 177 pathways significantly overrepresented (Supplemental Figure 5A, Supplemental Table 2) These pathways were associated with immunological processes (eg, immune system

process [Gene Ontology (GO) and Reactome], inflammatory response [GO]), involved various cell surface receptor signaling pathways (eg, cell surface receptor signaling pathway [GO:BP (biological process)], cytokine-cytokine receptor interaction [Kyoto Encyclopedia of Genes and Genomes]). Among the most significant pathways were the closely related PI3K/Akt signaling pathway, MAPK Akt signaling pathway, Ras signaling pathway, and EGFR tyrosine kinase inhibitor resistance (Supplemental Table 1). Supplemental Figure 5B visualizes the close relation between the markers (n = 18) and their connection to the pathways.

Although not directly connected in the pathways, markers from the cytokine-cytokine receptor interaction pathways are also involved. IL1RL1 is involved in negative regulation of the PI3K/Akt network and TNFRSF6B in turn is upregulated by the PI3K/Akt-dependent pathway. The strongest of these connecting markers were GDF15-ERBB2 and VEGFR2-S1PR1-ERK1/2-PKC-alpha complexes. These complexes are located at the intersection of the PI3K-Akt, MAPK, and EGFR tyrosine kinase pathways (Supplemental Figure 5C).

Clinical parameters associated with activation of the selected pathways are presented in Table 2. These characteristics are comparable to the features selected in the phenotype panel of our level 1 model. History of renal disease and peripheral edema are most significant related to the pathways, but diabetes also plays a role. Based on these clinical parameters, we could reasonably predict which patients would have activation of pathways (C-statistic: 0.75; 95% CI: 0.74-0.77). Pathways are significantly less activated in patients with HFrEF compared with patients with higher ejection fractions (P < 0.0001). Please note that the measurement of pathway activation is performed at baseline and before up-titration. Mean ACE inhibitor dose of the group of patients with activated pathways (n = 892) was 42.8% ± 37.6% of the target dose, compared with 55.5% ± 38.0% in the patients who did not have activated pathways (n = 1,208). Patients with activated pathways had higher risk of dying, with an HR of 2.67 (95% CI: 2.25-3.16; all P < 0.0001 corrected for age and sex).

DISCUSSION

This systems biology multiomics approach integrating genomic, transcriptomic, and proteomic data identified 4 major pathways in the progression of heart failure: PI3K-Akt, MAPK, and Ras signaling pathways, and EGFR tyrosine kinase inhibitor resistance. These pathways were identified using 18 of the

TABLE 2 Results of Penalized Linear Regression of Pathway Activation and Clinical Parameters From the Index Cohort and Multivariable Analysis of Pathway Activation on Drug Up-Titration Levels in Patients With Heart Failure With Reduced Ejection Fraction From the Index Cohort

	Estimate	SE	T Value	P Value
Penalized linear regression				
(Intercept)	−4.76	1.017	−4.67	0.001
Age	0.04	0.01	5.24	<0.001
AF	0.50	0.15	3.24	0.001
DM	0.51	0.16	3.10	0.002
Renal failure	2.05	0.17	12.30	<0.0001
DBP	−0.03	0.01	−4.20	<0.0001
Pulmonary congestion	0.11	0.09	1.32	0.188
Edema	0.92	0.10	9.07	<0.0001
Hepatomegaly	0.63	0.22	2.83	0.005
Third heart tone	1.08	0.25	4.35	<0.0001
NYHA functional class	0.46	0.12	3.78	0.001
Orthopnea	0.43	0.18	2.35	0.020
Multivariable analysis				
(Intercept)	−4.15	0.70	−5.91	<0.0001
% ACE inhibitor/ARB target dose	−1.53	0.29	−5.25	<0.0001
% BB target dose	−0.12	0.38	−0.33	0.741
% MRA target dose	0.63	0.34	1.81	0.070
Age	0.07	0.01	6.90	<0.0001
Female	−0.70	0.27	−2.57	0.010

ACE = angiotensin-converting enzyme; AF = atrial fibrillation; ARB = angiotensin receptor blocker; BB = beta-blocker; DBP = diastolic blood pressure; DM = diabetes mellitus; MRA = mineralocorticoid receptor antagonist.

180 significant markers in our model spread over all biomarker panels. Activation of these signaling pathways was strongly and independently associated with higher mortality.

Interestingly, these pathways are known to be strongly related to each other, and the Ras/Raf/MAPK pathway cascade is able to stimulate angiogenesis through changes in expression of genes directly involved in the formation of new blood vessels.[32,33] Signaling through the Ras/Raf/MAPK also regulates a variety of cellular functions that are important for tumorigenesis. Ras also interacts with the PI3K/AKT and EGFR tyrosine kinase inhibitor pathways.[34,35]

MAPK is activated in response to a wide variety of extracellular stimuli and induces changes in critical intracellular processes promoting cell growth, apoptosis, and transformation. It can transduce multiple extracellular signals through various receptors, such as hypertrophic signals mediated by G-protein–coupled receptors, transforming growth factor-β signals mediated by receptor serine/threonine kinases, and insulin-like growth factor-1 signals mediated by receptor tyrosine kinase.[36]

The PI3K/Akt pathway is important in mediating signals of cell growth and proliferation. It plays an important role in regulating cardiac growth,

myocardial angiogenesis, glucose metabolism, and cell death in cardiac myocyte.[37,38]

The EGFR tyrosine kinase inhibitor resistance pathway is important in the treatment of various cancers. These treatments report a high risk of cardiotoxicity. Although a mechanistic explanation for the cardiotoxicity of EGFR tyrosine kinase inhibitor is not fully understood, disruption of ERBB family receptors impairs downstream signaling to Ras-ERK and PI3K/Akt pathways and normal cardiac myocyte stress response.[39,40] The association between these pathways and cardiotoxicity became apparent after the presentation of phase 3 trials on the use of trastuzimab (herceptin) in patients with breast cancer. Unexpectedly, an increased risk of the development of congestive heart failure was observed in patients treated with trastuzumab.[41,42] This finding was explained by the ERBB signaling pathway. Activation of the ERBB2 and ERBB4 receptors leads to downstream activation of the PI3K/Akt and MAPK pathways,[43] which on the one hand promotes proliferation of tumor cells but on the other hand promotes cardiomyocyte survival. These mechanisms are supported by data from cardiac myocyte–specific ERBB2$^-$ and ERBB4$^-$ conditional knockout mice, who developed a cardiomyopathy by 8 to 12 weeks of life.[44] In other words, stimulation of the ERBB2 and ERBB4 receptors seem to exert cardioprotective effects through activation of the PI3K and MAPK pathways.[45] Importantly, the ERBB receptors can be stimulated by neuregulin and parenteral administration of recombinant human neuregulin-1 to patients with stable chronic heart failure resulted in an increase in stroke volume and cardiac output.[46-49]

FUTURE PERSPECTIVES

With the present study, we showed the role and opportunities of systems biology in unravelling underlying pathology of complex diseases, which is attracting increasing attention in the field of in cardiology.[50-53] However, as far as we know, there has not yet been a study using this advanced methodology in such a data-rich cohort with the ability of validating the results.[54]

This comprehensive picture of markers involved in the pathophysiological disease processes underlying all-cause mortality already yielded a potential future therapeutic intervention target, which is currently in the early phase of development for the treatment of patients with HFrEF. The literature also suggests that neuregulins might also be beneficial in patients with preserved ejection fraction and other cardiovascular diseases.[55,56]

STUDY LIMITATIONS. The transcriptomic panel consisted of 944 patients selected from the index cohort and matched on age and sex.[19] This is an extensive transcriptomic data set, but unfortunately, data were measured in a preselected group of patients from the index cohort and none from the validation cohort. The selection of patients was not random and skewed toward cardiovascular mortality. Transcriptomic markers are therefore better suited to predict cardiovascular mortality. This might explain the lower contribution of the markers from the transcriptomic panel in our combined systems biology model. The absence of this panel had no impact on the (level 0) model development and validation of the other panels because our methods are able to handle changes in data sources. Also, despite the rigorous selection process, the effects of patient selection cannot be determined.

Unfortunately, because of the nature of this study, we are not able to draw causal conclusions on the pathways we found. However, it is apparent that when developing the models for predicting mortality, so many markers from all panels are independently selected that are associated with the outcome.

The BIOSTAT-CHF study was carried out between 2010 and 2015 and treatments were based on the guidelines that were applicable at that particular time.[57] They did not include treatment with sacubitril/valsartan and sodium-glucose cotransporter-2 inhibitors. Therefore, theoretically, additional use of sacubitril-valsartan and sodium-glucose cotransporter-2 inhibitors might have yielded different results. Although there are limited data on the interaction between these treatments and the activated pathways,[58-61] it is unknown to what extent our results would have been different.

Due to the relatively low percentage of patients in the index cohort (7%) and the validation cohort (34%), we were not able to discriminate between heart failure with preserved ejection fraction and HFrEF.[17]

CLINICAL IMPLICATIONS. Treatment of HFrEF has tremendously improved over the past decades. With the current therapies, life expectancy for a 70-year-old patient with HFrEF has increased by 5 years.[62] Nevertheless, residual mortality remains high, even in well-treated patients. This implies that our current therapies do not adequately target all disease pathways that are related to its progression. The present study identified pathways that remained to be activated in patients with HFrEF despite treatment with ACE inhibitors, ARBs, mineralocorticoid receptor antagonists, and beta-blockers. Although the current

analysis does not show causality, these data might stimulate to identify potential novel treatments for HFrEF, targeting the pathways that were identified in the present study. This might further improve outcomes of patients who remained to have a high risk of early mortality.

CONCLUSIONS

Integrating genomic, transcriptomic, proteomic, and clinical data from a large cohort of patients with heart failure identified pathways related to progression of heart failure leading to early mortality. The strongest pathways were related to the ERBB receptors and their downstream effects on the PI3K and MAPK pathways leading to cardioprotective effects. Neuregulin, a ligand of the ERBB receptors, is currently in the early phase of clinical development in patients with HFrEF.

ADDRESS FOR CORRESPONDENCE: Dr Wouter Ouwerkerk, Bart van der leckplantsoen 31, Amsterdam, the Netherlands. E-mail: w.ouwerkerk@amsterdamumc.nl. @wouterou.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: A systems biology approach combining multiple large -omics data sets and machine learning methodology identified 4 major pathways related to mortality in patients with HFrEF. These pathways are related to decreased activation of the cardioprotective ERBB2 receptor, which can be modified by neuregulin.

TRANSLATIONAL OUTLOOK: Further research is needed to establish the causal relationships between activation of the ERBB2 receptor and adverse clinical outcomes in patients with HFrEF and identify potential treatment pathways.

REFERENCES

1. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2021;42:3599-3726.

2. Kitano H. Computational systems biology. *Nature.* 2002;420:206-210.

3. Valenzuela O, Rojas F, Rojas I, Glosekotter P. Main findings and advances in bioinformatics and biomedical engineering- IWBBIO 2018. *BMC Bioinformatics.* 2020;21:153.

4. Parker BL, Calkin AC, Seldin MM, et al. An integrative systems genetic analysis of mammalian lipid metabolism. *Nature.* 2019;567:187-193.

5. Yarden Y, Pines G. The ERBB network: at last, cancer therapy meets systems biology. *Nat Rev Cancer.* 2012;12:553-563.

6. Bayes-Genis A, Liu PP, Lanfear DE, et al. Omics phenotyping in heart failure: the next frontier. *Eur Heart J.* 2020;41:3477-3484.

7. Santema BT, Kloosterman M, Van Gelder IC, et al. Comparing biomarker profiles of patients with heart failure: atrial fibrillation vs. sinus rhythm and reduced vs. preserved ejection fraction. *Eur Heart J.* 2018;39:3867-3875.

8. Tromp J, Ouwerkerk W, Demissei BG, et al. Novel endotypes in heart failure: effects on guideline-directed medical therapy. *Eur Heart J.* 2018;39:4269-4276.

9. Ouwerkerk W, Zwinderman AH, Ng LL, et al. Biomarker-guided versus guideline-based treatment of patients with heart failure: results from BIOSTAT-CHF. *J Am Coll Cardiol.* 2018;71:386-398.

10. Markousis-Mavrogenis G, Tromp J, Ouwerkerk W, et al. Multimarker profiling identifies protective and harmful immune processes in heart failure: findings from BIOSTAT-CHF. *Cardiovasc Res.* 2022;118:1964-1977. https://doi.org/10.1093/cvr/cvab235

11. Markousis-Mavrogenis G, Tromp J, Ouwerkerk W, et al. The clinical significance of interleukin-6 in heart failure: results from the BIOSTAT-CHF study. *Eur J Heart Fail.* 2019;21: 965–973. https://doi.org/10.1002/ejhf.1482

12. Kloosterman M, Santema BT, Roselli C, et al. Genetic risk and atrial fibrillation in patients with heart failure. *Eur J Heart Fail.* 2020;22:519–527.

13. Tromp J, Westenbrink BD, Ouwerkerk W, et al. Identifying pathophysiological mechanisms in heart failure with reduced versus preserved ejection fraction. *J Am Coll Cardiol.* 2018;72:1081–1090.

14. Sama IE, Woolley RJ, Nauta JF, et al. A network analysis to identify pathophysiological pathways distinguishing ischaemic from non-ischaemic heart failure. *Eur J Heart Fail.* 2020;22:821–833. https://doi.org/10.1002/ejhf.1811

15. Romaine SPR, Denniff M, Codd V, et al. Telomere length is independently associated with all-cause mortality in chronic heart failure. *Heart.* 2022;108:124–129.

16. Ter Maaten JM, Voors AA, Damman K, et al. Fibroblast growth factor 23 is related to profiles indicating volume overload, poor therapy optimization and prognosis in patients with new-onset and worsening heart failure. *Int J Cardiol.* 2018;253:84–90.

17. Voors AA, Anker SD, Cleland JG, et al. A systems BIOlogy Study to TAilored Treatment in Chronic Heart Failure: rationale, design, and baseline characteristics of BIOSTAT-CHF. *Eur J Heart Fail.* 2016;18:716–726.

18. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009;150:604–612.

19. Nath M, Romaine SPR, Koekemoer A, et al. Whole blood transcriptomic profiling identifies molecular pathways related to cardiovascular mortality in heart failure. *Eur J Heart Fail.* 2022;24:1009–1019.

20. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods.* 2013;10:5–6.

21. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5:e1000529.

22. Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. *Nature.* 2016;526:75–81.

23. Pereira J, Stroes ESG, Groen AK, Zwinderman AH, Levin E. Manifold mixing for stacked regularization. In: Oliver N, Pérez-Cruz F, Kramer S, Read J, Lozano JA, eds. *Machine Learning and Knowledge Discovery in Databases.* Springer; 2020:444–452.

24. Frazier PI. A tutorial on Bayesian optimization. Preprint. *arXiv.* Posted online July 8, 2018. 02811. https://doi.org/10.48550/arXiv.1807.02811

25. Stoppiglia H, Dreyfus G, Dubois R, Oussar Y. Ranking a random feature for variable and feature selection. *J Mach Learn Res.* 2003;3: 1399–1414.

26. Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Series B Stat Methodol.* 2010;72: 417–473.

27. Marques T, Buckland S, Borchers D, Rexstad E, Thomas L. Distance sampling. In: Lovric M, ed. *International Encyclopedia of Statistical Science.* Vol. 1. Springer; 2010:398–400.

28. McLaren W, Gil L, Hunt SE, et al. The Ensembl variant effect predictor. *Genome Biol.* 2016;17:122.

29. Josse J, Husson F. Selecting the number of components in principal component analysis using cross-validation approximations. *Comput Stat Data Anal.* 2012;56:1869–1879.

30. Bro R, Kjeldahl K, Smilde AK, Kiers HAL. Cross-validation of component models: a critical look at current methods. *Anal Bioanal Chem.* 2008;390: 1241–1251.

31. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Series B Stat Methodol.* 1994;58:267–288.

32. Molina JR, Adjei AA. The Ras/Raf/MAPK pathway. *J Thorac Oncol.* 2006;1:7–9.

33. Kranenburg O, Gebbink MFBG, Voest EE. Stimulation of angiogenesis by Ras proteins. *Biochim Biophys Acta Rev Cancer.* 2004;1654:23–37.

34. Vitiello PP, Cardone C, Martini G, et al. Receptor tyrosine kinase-dependent PI3K activation is an escape mechanism to vertical suppression of the EGFR/RAS/MAPK pathway in KRAS-mutated human colorectal cancer cell lines. *J Exp Clin Cancer Res.* 2019;38:41.

35. Castellano E, Downward J. Ras interaction with PI3K: More than just another effector pathway. *Genes Cancer.* 2011;2:261–274.

36. Heineke J, Molkentin JD. Regulation of cardiac hypertrophy by intracellular signalling pathways. *Nat Rev Mol Cell Biol.* 2006;7:589–600.

37. Walkowski B, Kleibert M, Majka M, Wojciechowska M. Insight into the role of the PI3K/Akt Pathway in Ischemic Injury and Post-Infarct Left Ventricular Remodeling in Normal and Diabetic Heart. *Cells.* 2022;11:1553.

38. Ghafouri-Fard S, Khanbabapour Sasi A, Hussen BM, et al. Interplay between PI3K/AKT pathway and heart disorders. *Mol Biol Rep.* 2022;49:9767–9781.

39. Piper-Vallillo AJ, Costa DB, Sabe MA, Asnani A. Heart Failure Associated With the Epidermal Growth Factor Receptor Inhibitor Osimertinib. *J Am Coll Cardiol CardioOnc.* 2020;2:119–122.

40. Lemmens K, Doggen K, De Keulenaer GW. Role of neuregulin-1/ErbB signaling in cardiovascular physiology and disease: implications for therapy of heart failure. *Circulation.* 2007;116: 954–960.

41. Inoue K, Nakagami K, Mizutani M, et al. Randomized phase III trial of trastuzumab monotherapy followed by trastuzumab plus docetaxel versus trastuzumab plus docetaxel as first-line therapy in patients with HER2-positive metastatic breast cancer: the JO17360 Trial Group. *Breast Cancer Res Treat.* 2010;119:127–136.

42. Nemeth BT, Varga ZV, Wu WJ, Pacher P. Trastuzumab cardiotoxicity: from clinical trials to

experimental studies. *Br J Pharmacol.* 2017;174: 3727–3748.

43. Kerkela R, Force T. p38 mitogen-activated protein kinase: a future target for heart failure therapy? *J Am Coll Cardiol.* 2006;48:556–558.

44. Ozcelik C, Erdmann B, Pilz B, et al. Conditional mutation of the ErbB2 (HER2) receptor in cardiomyocytes leads to dilated cardiomyopathy. *Proc Natl Acad Sci U S A.* 2002;99:8880–8885.

45. Cote GM, Sawyer DB, Chabner BA. ERBB2 Inhibition and Heart Failure. *N Engl J Med.* 2012;367: 2150–2153.

46. Jabbour A, Hayward CS, Keogh AM, et al. Parenteral administration of recombinant human neuregulin-1 to patients with stable chronic heart failure produces favourable acute and chronic haemodynamic responses. *Eur J Heart Fail.* 2011;13:83–92.

47. Xu J, Sun P, Zhao X, et al. Safety, Tolerability, and Pharmacokinetics of Recombinant Human Neuregulin-1 in Healthy Chinese subjects. *Am J Cardiovasc Drugs.* 2023;23:419–428. https://doi.org/10.1007/s40256-023-00585-6

48. Zensun Sci. & Tech. Co. Ltd. Survival Study of the Recombinant Human Neuregulin-1β in subjects with chronic heart failure. NCT03388593. Accessed August 16, 2023. https://classic.clinicaltrials.gov/ct2/show/NCT03388593

49. Gao R, Zhang J, Cheng L, et al. A phase II, Randomized, Double-Blind, Multicenter, Based on Standard Therapy, Placebo-Controlled Study of the Efficacy and Safety of Recombinant Human Neuregulin-1 in Patients With Chronic Heart Failure. *J Am Coll Cardiol.* 2010;55:1907–1914. https://doi.org/10.1016/j.jacc.2009.12.044

50. Trachana K, Bargaje R, Glusman G, Price ND, Huang S, Hood LE. Taking systems medicine to heart. *Circ Res.* 2018;122:1276–1289.

51. Leopold JA, Loscalzo J. Emerging Role of Precision Medicine in Cardiovascular Disease. *Circ Res.* 2018;122:1302–1315.

52. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One.* 2017;12:e0174944.

53. Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. *Nat Rev Cardiol.* 2021;18: 313–330. https://doi.org/10.1038/s41569-020-00477-1

54. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol Adv.* 2021;107739.

55. Geissler A, Ryzhov S, Sawyer DB. Neuregulins: protective and reparative growth factors in multiple forms of cardiovascular disease. *Clin Sci.* 2020;134:2623–2643.

56. Wang Y, Wei J, Zhang P, et al. Neuregulin-1, a potential therapeutic target for cardiac repair. *Front Pharmacol.* 2022;13:945206.

57. McMurray JJ, Adamopoulos S, Anker SD, et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure

2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail*. 2012;14:803-869.

58. Ding J, Cui S, Li SY, et al. The angiotensin receptor neprilysin inhibitor LCZ696 attenuates renal fibrosis via ASK1/JNK/p38 MAPK-mediated apoptosis in unilateral ureteral obstruction. *PLoS One*. 2023;18:e0286903.

59. Zhang X, Yan C, Zheng M. Sacubitril-valsartan ameliorates heart failure by inhibiting cardiac remodeling potentially via MAPK/ERK signaling. *Ann Clin Lab Sci*. 2022;52:391-938.

60. Mai Z, Li H, Chen G, et al. A bioinformatics investigation into the pharmacological mechanisms of sodium-glucose co-transporter 2 inhibitors in diabetes mellitus and heart failure based on network pharmacology. *Cardiovasc Drugs Ther*. 2022;36:713-726.

61. Weintraub MA, Liu D, DeMatteo R, Goncalves MD, Flory J. Sodium-glucose cotransporter-2 inhibitors for hyperglycemia in phosphoinositide 3-kinase pathway inhibition. *Res Sq*. Published online March 21, 2023. https://doi.org/10.21203/rs.3.rs-2655905/v1

62. Tromp J, Ouwerkerk W, van Veldhuisen DJ, et al. A systematic review and network meta-analysis of pharmacological treatment of heart failure with reduced ejection fraction. *J Am Coll Cardiol HF*. 2022;10:73-84.

KEY WORDS heart failure, machine learning, omics, systems biology

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