ESC European Society of Cardiology

Europace (2024) **26**, euae267
https://doi.org/10.1093/europace/euae267

CLINICAL RESEARCH

Prediction of new-onset atrial fibrillation in patients with hypertrophic cardiomyopathy using plasma proteomics profiling

Heidi S. Lumish [1†], Nina Harano [2†], Lusha W. Liang [1], Kohei Hasegawa [3], Mathew S. Maurer [1], Albree Tower-Rader[4], Michael A. Fifer [4], Muredach P. Reilly[1,5], and Yuichi J. Shimada [1*]

[1]Division of Cardiology, Department of Medicine, Columbia University Irving Medical Center, 622 West 168th Street, PH3-342, New York, NY 10032, USA; [2]Department of Biology, Columbia University, New York, NY, USA; [3]Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; [4]Cardiology Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; and [5]Irving Institute for Clinical and Translational Research, Columbia University Irving Medical Center, New York, NY, USA

Received 2 April 2024; accepted after revision 23 July 2024; online publish-ahead-of-print 23 October 2024

Aims Atrial fibrillation (AF) is the most common sustained arrhythmia among patients with hypertrophic cardiomyopathy (HCM), increasing symptom burden and stroke risk. We aimed to construct a plasma proteomics-based model to predict new-onset AF in patients with HCM and determine dysregulated signalling pathways.

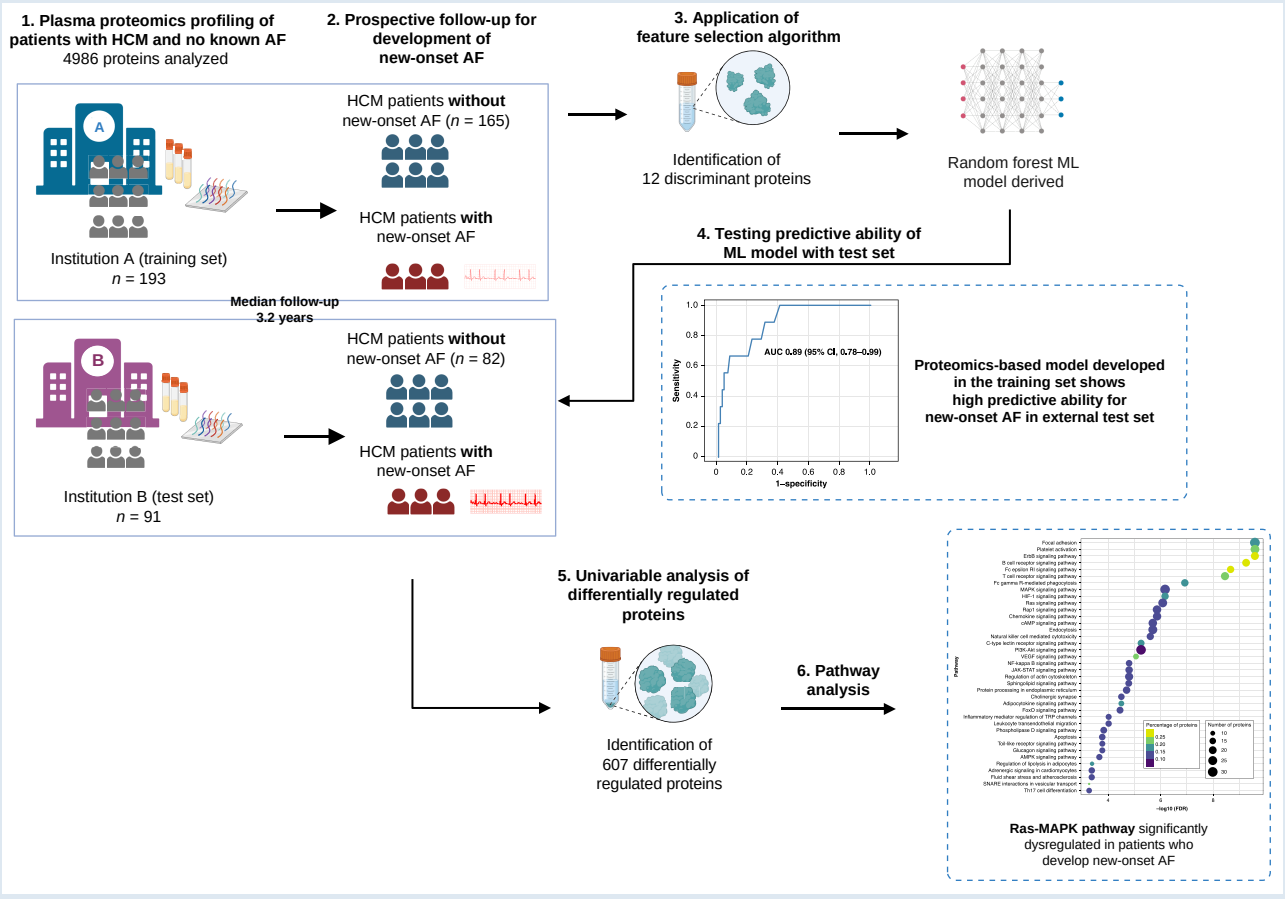
Methods and results In this prospective, multi-centre cohort study, we conducted plasma proteomics profiling of 4986 proteins at enrolment. We developed a proteomics-based machine learning model to predict new-onset AF using samples from one institution (training set) and tested its predictive ability using independent samples from another institution (test set). We performed a survival analysis to compare the risk of new-onset AF among high- and low-risk groups in the test set. We performed pathway analysis of proteins significantly (univariable $P < 0.05$) associated with new-onset AF using a false discovery rate (FDR) threshold of 0.001. The study included 284 patients with HCM (training set: 193, test set: 91). Thirty-seven (13%) patients developed AF during median follow-up of 3.2 years [25–75 percentile: 1.8–5.2]. Using the proteomics-based prediction model developed in the training set, the area under the receiver operating characteristic curve was 0.89 (95% confidence interval 0.78–0.99) in the test set. In the test set, patients categorized as high risk had a higher rate of developing new-onset AF (log-rank $P = 0.002$). The Ras-MAPK pathway was dysregulated in patients who developed incident AF during follow-up (FDR $< 1.0 \times 10^{-6}$).

Conclusion This is the first study to demonstrate the ability of plasma proteomics to predict new-onset AF in HCM and identify dysregulated signalling pathways.

* Corresponding author. Tel: 212-305-8013; fax: 212-342-3591. *E-mail address*: ys3053@cumc.columbia.edu

† The first two authors contributed equally to the study.

Graphical Abstract



Keywords Atrial fibrillation • Hypertrophic cardiomyopathy • Proteomics • Prediction

What's new?

- Atrial fibrillation (AF) is the most common sustained arrhythmia among patients with hypertrophic cardiomyopathy (HCM), leading to increased symptom burden and risk of stroke. In patients with HCM, concomitant AF is a Class 1 indication for anticoagulation regardless of the CHA2DS2-VASc score. There is an urgent need for more accurate tools to predict new-onset AF in HCM to facilitate intensified surveillance, earlier detection, and initiation of treatment.
- This study serves as the first to demonstrate the ability of plasma proteomics profiling to predict new-onset AF in HCM and further identifies dysregulated signalling pathways (e.g. Ras-MAPK) in patients with new-onset AF.
- The study contributes to our understanding of the molecular mechanisms underlying the development of AF in HCM.

Introduction

Hypertrophic cardiomyopathy (HCM) is the most common genetic cardiac disease, affecting 1 in 200–500 people.[1,2] Atrial fibrillation (AF) is the most common sustained arrhythmia among patients with HCM.[3,4] While AF only affects 1–2% of the general population, it is more common in patients with HCM, with a prevalence of 13–24% and an incidence of 2%/year.[5–10] Atrial fibrillation is associated with increased morbidity and mortality among patients with HCM. Patients with HCM are particularly vulnerable to AF-induced loss of atrial systole due to stiffening of the left ventricle (LV), which can lead to a reduction in cardiac output and heart failure (HF) symptoms.[5,9,11] Moreover, ischaemic strokes are eight times more common in patients with HCM and concomitant AF as compared to patients with HCM in sinus rhythm.[7] Importantly, AF is also associated with increased mortality among patients with HCM, independent of traditional risk factors for sudden cardiac death.[9]

Once patients with HCM at increased risk for AF are identified, treating physicians can intensify their surveillance strategies by performing more frequent rhythm monitoring. Furthermore, anticoagulation is an effective strategy for stroke prevention in patients with AF.[12] Thus, initiation of therapeutic anticoagulation for patients with HCM and AF is identified as a Class 1 recommendation in the 2024 AHA/ACC guidelines for the management of HCM, regardless of the CHA_2DS_2-VASc score.[13] Despite the importance of early detection of AF in patients with HCM, the currently available tool to predict AF in the HCM population—the HCM-AF model—is based only on clinical variables and is limited in its sensitivity {58% [95% confidence interval (CI) 50–67%]} and specificity [66% (95% CI 63–69%)].[14] Therefore, there remains an urgent need for more accurate AF risk stratification tools for patients with HCM.

Moreover, further understanding of the molecular signalling pathways through which HCM leads to AF can contribute to the development of novel targeted therapeutics to prevent HCM-related AF.

Proteomics profiling is a recently developed technology that simultaneously measures the concentration of thousands of proteins in a tissue or fluid sample. Proteomics profiling has been used in several cardiovascular diseases (e.g. HF, coronary artery disease, and hypertension), not only to study underlying disease mechanisms but also to predict adverse events.[15–17] However, proteomics profiling has never been used to predict AF in HCM or to identify dysregulated signalling pathways associated with new-onset AF in HCM. Therefore, in this multi-centre, prospective cohort study of patients with HCM, we aimed to (i) develop a plasma proteomics-based model to predict new-onset AF and (ii) determine signalling pathways dysregulated in patients with new-onset AF.

Methods

Study design and sample

This was a multi-centre prospective cohort study conducted among patients aged ≥18 years with HCM. Patients were recruited from the inpatient or outpatient setting at Massachusetts General Hospital (MGH, Boston, MA) and Columbia University Irving Medical Center (CUIMC, New York, NY) between October 2015 and September 2019. The training set for deriving the prediction model consisted of patients with HCM who were followed at MGH, while the external test set for validation consisted of patients with HCM who were followed at CUIMC. Informed consent was obtained from all participants, and the study followed protocols approved by the Mass General Brigham Human Research Office/Institutional Review Board and the CUIMC Institutional Review Board.

Hypertrophic cardiomyopathy was diagnosed based on echocardiographic evidence of LV hypertrophy, defined as the maximum end-diastolic wall thickness of 15 mm or greater out of proportion to systemic loading conditions.[13] Patients with suspected HCM phenocopies (e.g. Fabry disease, Danon disease, and cardiac amyloidosis) were excluded based on extensive history and clinical examination and, if indicated, additional genetic testing, nuclear scintigraphy, or cardiac magnetic resonance imaging.[13,18] Patients with suspected secondary LV hypertrophy due to other conditions (e.g. aortic stenosis, subaortic membrane, or athlete's heart) were excluded.[13] Patients diagnosed with AF prior to enrolment were also excluded.

Processing of blood samples and proteomics profiling

Blood samples were collected from the inpatient and outpatient setting following a standardized protocol.[19,20] Proteomics profiling of 4986 proteins was performed using the SomaScan assay (SOMALogic, Inc., Boulder, CO).[21–23] Unlike conventional methods based on liquid chromatography and mass spectrometry, which are limited in their ability to determine low-abundance protein concentrations (e.g. cytokines), the SomaScan assay can determine protein concentrations across 10^8 in abundance from femtomolar to micromolar.[21–23] The high reproducibility (median coefficient of variation 4.6%) and accuracy of the SomaScan are comparable to those of sandwich enzyme-linked immunosorbent assay.[21–23] Additional details pertaining to processing of the blood samples and proteomics profiling are available in the Supplementary material online, *Supplemental Methods*.

Outcome measure

The primary outcome was new-onset AF in patients with HCM detected through routine clinical care with 12-lead electrocardiograms, Holter monitors, ambulatory rhythm monitors, loop recorders, and intracardiac devices (i.e. permanent pacemakers and implantable cardioverter-defibrillators) during the follow-up period. The primary outcome was adjudicated prospectively by two attending cardiologists.

Univariate analysis

Statistical analyses were performed using R version 4.0.3. Continuous variables were presented as mean ± standard deviation for data with normal distribution and median [25–75 percentile] for data with non-normal or skewed distribution. Categorical variables were presented as count (%). Differences between patients with and without new-onset AF in clinical characteristics and in several cardiac biomarkers known to be associated with AF [i.e. B-type natriuretic peptide (BNP), troponins, and C-reactive protein] were tested for statistical significance using the unpaired Student's *t*-test for normally distributed continuous variables, the Mann–Whitney–Wilcoxon test for ordinal variables and non-normally distributed continuous variables, and the χ^2 test or Fisher's exact test for categorical variables.

Table 1 Baseline clinical characteristics of the study sample

Characteristics[a]	No new-onset AF ($n = 247$)	New-onset AF ($n = 37$)	*P* value
Demographics			
Age (year)	57 ± 16	61 ± 14	0.16
Male	157 (64)	20 (54)	0.27
Body mass index (kg/m$^{2)}$	30 ± 6	32 ± 7	0.04
NYHA functional Class ≥ 2	117 (47)	20 (54)	0.45
Race/ethnicity			0.94
Caucasian	182 (74)	31 (84)	
African-American	17 (7)	2 (5)	
Asian	7 (3)	1 (3)	
Other or unidentified	41 (17)	3 (8)	
Medical and family history			
Diabetes mellitus	22 (10)	3 (10)	>0.99
Hypertension	125 (51)	19 (51)	0.99
Serum creatinine (mg/dL)	1.0 ± 0.5	1.0 ± 0.2	0.49
Prior VT/VF	9 (4)	2 (5)	0.64

Continued

Table 1 *Continued*

Characteristics[a]	No new-onset AF (*n* = 247)	New-onset AF (*n* = 37)	*P* value
Prior non-sustained VT	33 (13)	8 (22)	0.18
Prior syncope	47 (19)	6 (16)	0.68
Prior septal myectomy	15 (6)	4 (11)	0.29
Prior alcohol septal ablation	17 (7)	3 (8)	0.73
Prior ICD implantation	57 (23)	13 (35)	0.11
Family history of sudden cardiac death	25 (10)	3 (8)	0.99
Family history of HCM	59 (24)	11 (30)	0.44
HCM-AF score	19.6 (3.4)	21.0 (3.1)	0.02
Medications			
β-Blocker	153 (62)	26 (70)	0.45
Calcium channel blocker	44 (18)	4 (11)	0.29
ACE inhibitor	21 (9)	4 (11)	0.55
ARB	34 (14)	4 (11)	0.80
Diuretic			
Loop diuretic	17 (7)	4 (11)	0.50
Thiazide	16 (6)	5 (14)	0.17
Potassium sparing diuretic	14 (6)	0 (0)	0.23
Disopyramide	9 (4)	3 (8)	0.20
Anticoagulant	6 (2.5%)	1 (2.9%)	>0.99
Blood pressure			
Systolic blood pressure (mmHg)	124 ± 15	129 ± 13	0.06
Diastolic blood pressure (mmHg)	74 ± 10	75 ± 8	0.66
Echocardiographic measurements			
Left atrial diameter (mm)	40 ± 6	42 ± 5	0.13
Left atrial volume index (mL/m^2)	39 (12)	43 (6)	0.47
Interventricular septum thickness (mm)	16 ± 4	17 ± 4	0.20
Posterior wall thickness (mm)	12 ± 2	13 ± 3	<0.001
LVOT gradient at rest (mmHg)	9 [0–40]	32 [0–81]	0.006
LVOT gradient with Valsalva manoeuvre (mmHg)	16 [0–70]	50 [0–87]	0.07
Left ventricular ejection fraction (%)	69 ± 10	71 ± 6	0.21
Left ventricular end-diastolic diameter (mm)	43 ± 6	41 ± 7	0.07
Left ventricular end-systolic diameter (mm)	27 ± 5	25 ± 5	0.06
Systolic anterior motion of mitral valve leaflet	106 (43)	21 (57)	0.11
Degree of mitral regurgitation[b]	2 [1–2.5]	2 [1–2]	0.28
Exercise stress test characteristics (*n* = 200)	*n* = 179	*n* = 21	
METs achieved with stress test	10.0 ± 3.8	7.9 ± 2.5	0.06
Peak heart rate during stress (b.p.m.)	142 ± 28	138 ± 25	0.57
Peak systolic blood pressure during stress (mmHg)	168 ± 29	172 ± 28	0.62
Peak diastolic blood pressure during stress (mmHg)	80 ± 13	83 ± 11	0.26
LVOT gradient with stress (mmHg)	34 [14–68]	54 [42–79]	0.12
Cardiac magnetic resonance imaging (*n* = 176)	*n* = 158	*n* = 18	
Late gadolinium enhancement	90 (57)	6 (33)	0.06
Genetic testing (*n* = 158)	*n* = 138	*n* = 20	
Pathogenic or likely pathogenic	42 (30)	4 (20)	0.80

[a]Data are expressed as number (percentage), mean ± standard deviation, or median [25 percentile–75 percentile].
[b]Degree of mitral regurgitation was converted to numerical values according to the following rule: none = 0, trace = 1, trace to mild = 1.5, mild = 2, mild to moderate = 2.5, moderate = 3, moderate to severe = 3.5, and severe = 4.
ACE, angiotensin converting enzyme; AF, atrial fibrillation; ARB, angiotensin II receptor blocker; HCM, hypertrophic cardiomyopathy; ICD, implantable cardioverter-defibrillator; MET, metabolic equivalent; NYHA, New York Heart Association; VT/VF, ventricular tachycardia or ventricular fibrillation.

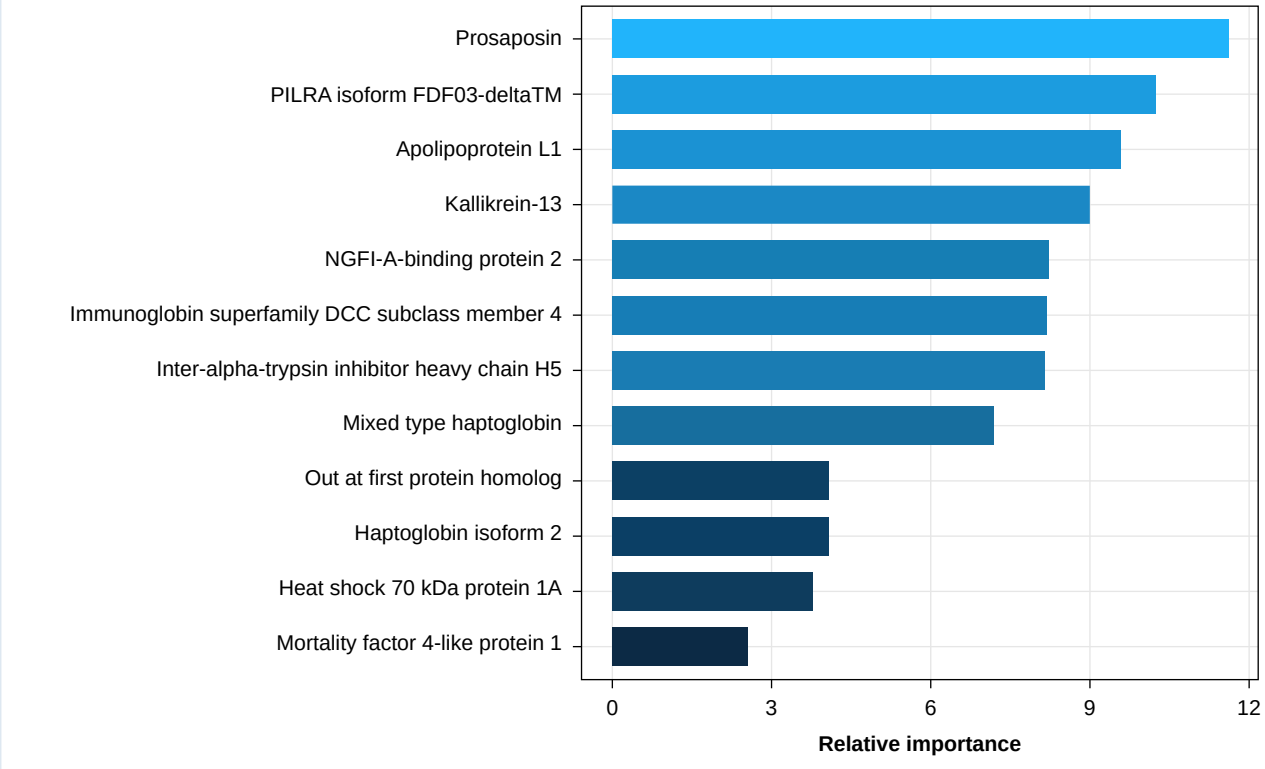


Figure 1 Relative importance of proteins included in the model to predict new-onset AF in patients with HCM. NGFI-A, nerve growth factor induced clone-A; PILRA, pared immunoglobulin-like type 2 receptor alpha.

Development of a proteomics-based model to predict new-onset atrial fibrillation in patients with hypertrophic cardiomyopathy

Due to the high-dimensional nature of the proteomics data, the Boruta algorithm was first applied for feature selection to the training set, using the *Boruta* package in R.[24] The Boruta algorithm is a wrapper around the random forest algorithm that iteratively removes features that are less important than random probes.[24] A total of 12 proteins were identified as important predictors of new-onset AF in patients with HCM, using the Boruta algorithm. Using five-fold cross-validation on the 12 important predictors, using the *caret* package, a random forest machine learning (ML) model was developed in the training set. Since random forest models are non-parametric and capable of handling skewed and multi-modal data, the data were not transformed. A hyperparameter tuning grid was used to identify the tuning parameters, optimizing the area under the receiver operating characteristic curve (AUC) in the training set. The 12-protein model was used to predict which patients with HCM would develop new-onset AF in the external test set. The *ROCR* package was used to calculate the AUC in the test set. The *OptimalCutpoints* package was used to determine the sensitivity and specificity of the model, and the threshold was selected by identifying the point on the receiver operating characteristic curve for new-onset AF that minimizes the distance to the point (0, 1), or the left upper corner of the graph.[25,26]

The association between clinical parameters in *Table 1* and the predicted probability derived from the proteomics-based ML model was tested using Pearson's correlation coefficient for continuous variables, analysis of variance for normally distributed categorical variables, and the Kruskal–Wallis test for non-normally distributed categorical variables.

Proteomics-based risk stratification

Patients in the test set were divided into high- and low-risk groups using the proteomics-based prediction model developed in the training set. The cut-off probability was determined by identifying the point on the receiver operating characteristic curve for new-onset AF that minimizes the distance to the point (0, 1), using the *OptimalCutpoints* package in R.[25,26] The difference in rates of new-onset AF in the high-risk as compared to the low-risk group was examined using the log-rank test.

Clinical prediction models

A clinical random forest ML model based on components of the HCM-AF score (i.e. New York Heart Association class ≥ 2, age at enrolment, age at HCM diagnosis, and left atrial diameter) was developed for comparison with the proteomics-based ML model.[14] The clinical model developed in the training set using five-fold cross validation in *caret* was used to predict new-onset AF in the test set. A second clinical prediction model was developed using the HCM-AF score. A combined clinical and proteomics model was developed incorporating the predicted probability of new-onset AF using the proteomics-based ML model, the HCM-AF score, and clinical variables from *Table 1* with a statistically significant difference between the two groups. For all clinical variables with missing values, we imputed the missing values using multivariate imputation by chained equations, with the R *mice* package. The DeLong test was used to compare the AUC of the proteomics-based ML model with each of the clinical and combined clinical and proteomics models.

Pathway analysis

The Mann–Whitney–Wilcoxon test was used to identify statistically significant differences in protein concentrations between patients with HCM who developed new-onset AF and those who did not. Pathway analysis was performed using STRING version 11 (String Consortium, Europe).[27] Associations between the significant proteins and canonical pathways from the Kyoto Encyclopedia of Genes and Genomes database were determined by examining the ratio of the number of proteins that map to a pathway divided by the total number of proteins that map to the pathway.

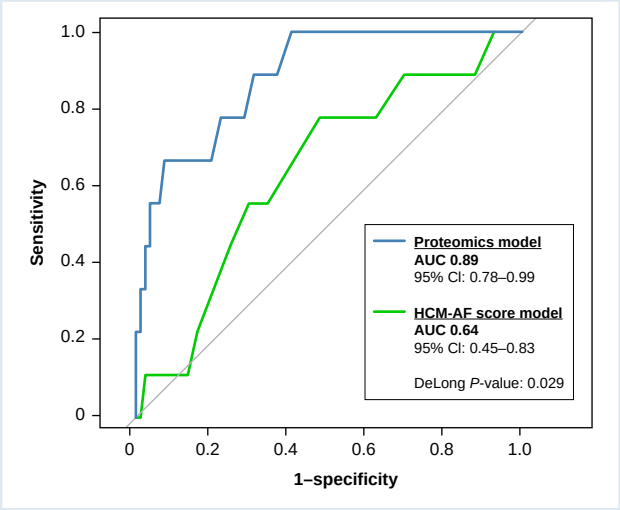


Figure 2 Receiver operating characteristic curves to predict new-onset AF using the proteomics-based ML model derived from the training set compared to the HCM-AF score, in the external test set. AUC, area under the receiver operating characteristic curve; CI, confidence interval.

Pathways with false discovery rate (FDR) < 0.001 were considered to be dysregulated.

Results

A total of 407 patients with HCM were identified. Among the 407 patients, 105 patients were excluded due to a diagnosis of AF prior to the start of the study, and 18 additional patients were excluded because outcome data were not available. Thus, the final analytic cohort consisted of 284 patients, who underwent proteomics profiling. The training set included 193 patients with HCM from MGH, and the test set consisted of 91 patients with HCM from CUIMC. *Table 1* demonstrates the baseline clinical characteristics of the cohort stratified by the primary outcome of new-onset AF. At baseline, patients with new-onset AF had higher body mass index (BMI), higher HCM-AF score,[14] greater LV posterior wall thickness, and higher left ventricular outflow tract (LVOT) gradient at rest, as compared to patients who did not develop new-onset AF. Concentrations of BNP, NT-proBNP, troponin I, and C-reactive protein were significantly higher in the new-onset AF group (see Supplementary material online, *Table S2*).

Median follow-up time was 3.17 [25–75 percentile: 1.83–5.17] years. A total of 37 patients developed new-onset AF during follow-up. New-onset AF occurred in 28 (14.5%) patients in the training set and 9 (9.9%) patients in the test set. The event rate was 4.00 per 100 patient-years in the training set and 3.08 per 100 patient-years for the test set. Methods of screening for new-onset AF are shown in the Supplementary material online, *Table S1*.

Twelve proteins were identified as important in predicting new-onset AF using the Boruta feature selection algorithm (*Figure 1*). A random forest ML model incorporating these 12 predictive proteins was derived from the training set. The AUC when applying the predictive model to the test set was 0.89 (95% CI: 0.78–0.99; *Figure 2*). Sensitivity was 78% (95% CI: 40–97%), and specificity was 78% (95% CI: 68–86%). Patients in the test set were classified as either high risk (25 patients) or low risk (66 patients) for developing AF according to the proteomics-based predictive model derived from the training set. The high-risk group had a significantly higher rate of developing new-onset AF compared to the low-risk group (log-rank $P = 0.002$; *Figure 3*). Follow-up time was comparable between the high- and low-risk groups [median follow-up time 2.67 (IQR 0.67–4.17) vs. 2.46 (IQR 1.48–3.81) years, respectively; $P = 0.786$]. Parameters known to be associated with new-onset AF or severe disease (e.g. BMI, HCM-AF score, and LVOT gradient at rest) were significantly associated with the predicted probability derived from the proteomics-based ML model (see Supplementary material online, *Table S3*).

The performance of the proteomics-based ML model was tested in comparison with several clinical and combined clinical and proteomics models. In a random forest ML model derived from the training set using only clinical variables from the HCM-AF score,[14] the AUC was 0.50 (95% CI: 0.31–0.69; Supplementary material online, *Figure S1*). The sensitivity was 67% (30–93%), and specificity was 46% (35–58%). The DeLong test demonstrated that the proteomics-based ML model achieved significantly better AUC than the clinical ML model based on the HCM-AF score components for predicting whether patients with HCM would develop new-onset AF ($P = 0.0007$). In a model derived from the HCM-AF score in the test set,[14] the AUC was 0.64 (95% CI: 0.45–0.83; *Figure 2*), with sensitivity 67% (44–100%) and specificity 66% (44–82%). The proteomics-based model outperformed the HCM-AF model for predicting new-onset AF in the test set ($P = 0.029$). Results were similar when the HCM-AF model was applied to the entire cohort (DeLong $P = 0.0003$; Supplementary material online, *Figure S2*). The addition of the HCM-AF score and the clinical variables with significant difference between the two groups did not significantly improve the predictive ability of the proteomics-based model (DeLong $P = 0.538$; Supplementary material online, *Figure S3*).

Among the 4986 proteins analysed, 607 proteins were differentially regulated between patients with HCM who developed new-onset AF and those who did not with $P < 0.05$. Pathway analysis of these 607 proteins revealed that the Ras-MAPK pathway was dysregulated (FDR < 1.0×10^{-6}), as well as several of its upstream and downstream pathways (e.g. ErbB, Rap1, PI3K-Akt, JAK-STAT, and mTOR; *Figure 4*). Additionally, several pathways known to be dysregulated in HCM, including pathways related to inflammation (*Figure 4*), were found to be dysregulated in patients with new-onset AF.

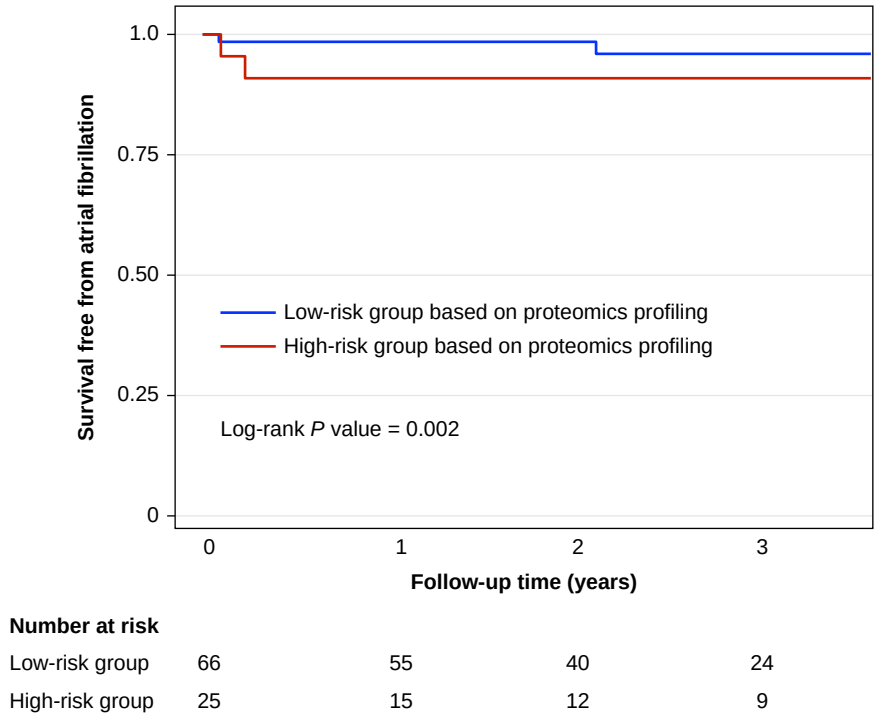
Discussion

Summary of findings

In this prospective, multi-centre cohort study of 284 patients with HCM, our 12-protein model derived from comprehensive proteomics profiling of 4986 plasma proteins demonstrated a high predictive ability for new-onset AF in HCM in both training and test sets. The high predictive ability of our model using only a small panel of plasma biomarkers (i.e. 12 proteins) suggests great potential for clinical application. Furthermore, proteomics profiling revealed both novel and previously recognized molecular signalling pathways associated with progression to AF in HCM. Notably, the Ras-MAPK pathway, as well as several of its associated upstream and downstream pathways, was dysregulated in patients with new-onset AF. This study serves as the first application of proteomics profiling to the HCM population to predict new-onset AF and to elucidate signalling pathways associated with the development of AF.

Results in context

Atrial fibrillation and its complications are significantly more frequent in patients with HCM compared to the general population. Based on recent studies, AF has a prevalence of 13–24% in patients with HCM as compared to 1–2% in the general population.[5,6,8,9] Our data



Number at risk				
Low-risk group	66	55	40	24
High-risk group	25	15	12	9

Figure 3 Kaplan–Meier curve for the risk of developing new-onset AF in the test set, using the proteomics-based prediction model derived from the training set.

demonstrated that the event rate was 3–4%/year, which is consistent with the reported incidence rate of 2%/year in a recent study.[7] Moreover, the risk of stroke is eight times higher in patients with HCM and concomitant AF than in patients with HCM in normal sinus rhythm.[7] Importantly, prevention of such devastating thromboembolic events with anticoagulation is possible once AF is detected.[12] Accordingly, initiation of therapeutic anticoagulation in patients with HCM and AF is a Class 1 recommendation, according to the 2024 AHA/ACC guidelines.[13]

In addition to the increased risk of stroke, the combination of AF and HCM is associated with markedly increased symptom burden and impaired quality of life.[14] Due to diastolic dysfunction, LV stiffness, and elevated LV filling pressures, patients with HCM are particularly susceptible to the loss of atrial systole caused by AF.[5] In addition, LVOT obstruction and resultant mitral valve regurgitation cause left atrial dilation, which in turn increases the inducibility of AF. There are several interventions available for rhythm control once AF has been diagnosed, including antiarrhythmic therapy, catheter ablation, and the Maze procedure.[5,28,29] Early detection and intervention are preferred, before left atrial dilation occurs.[30] Therefore, prediction, intensified surveillance, and early detection of AF in patients with HCM are critical.

Despite the importance of early AF detection and available treatments, AF is often asymptomatic and difficult to diagnose. In the USA alone, 13.1% of the 5.3 million AF cases were undiagnosed in the general population.[13,31,32] Therefore, establishing a patient-friendly, practical method for the identification of high-risk patients and early detection of AF in patients with HCM is crucial for modifying clinical management and avoiding devastating consequences. In this context, the plasma proteomic-based model developed in the present study can serve as a potential risk stratification tool for the prediction of new-onset AF in HCM.

Comparison between proteomics-based and clinical machine learning models to predict new-onset atrial fibrillation

Comprehensive proteomics profiling with the SomaScan assay is a recently developed technology that simultaneously measures the concentrations of thousands of proteins in a blood sample as small as 65 µL.[19] Proteomics profiling has been used to identify circulating plasma biomarkers and proteins associated with a variety of cardiovascular diseases and conditions, such as coronary artery disease,[17,33,34] hypertension,[16] HF,[15,35] and non-HCM cardiomyopathies.[36–38] Proteomics profiling has also been used to predict incident AF in non-HCM populations in the Framingham Heart Study.[39] Our prior studies have applied proteomics profiling to distinguish patients with HCM from those without.[19,20] These studies collectively support the ability of proteomics profiling to identify circulating biomarkers to predict disease progression. The present study serves as the first to apply proteomics profiling to identify circulating biomarkers to predict new-onset AF specifically among patients with HCM.

Underscoring the importance of predicting new-onset AF, several methods have been introduced to identify patients with HCM who are at high risk of developing new-onset AF using clinical parameters. The HCM-AF score is a recently developed model to predict AF in HCM using clinical variables.[14] The HCM-AF score model has a modest predictive ability [i.e. sensitivity of 58% (95% CI, 50–67%) and specificity of 66% (63–69%)].[14] In contrast, our model using proteomics profiling achieved significantly better accuracy to predict AF in HCM compared to the HCM-AF score. This is likely because proteomic profiles reflect biological changes more directly than clinical parameters. One may argue that the improvement in the predictive accuracy may be attributed to the use of ML methods rather than proteomics. However, the

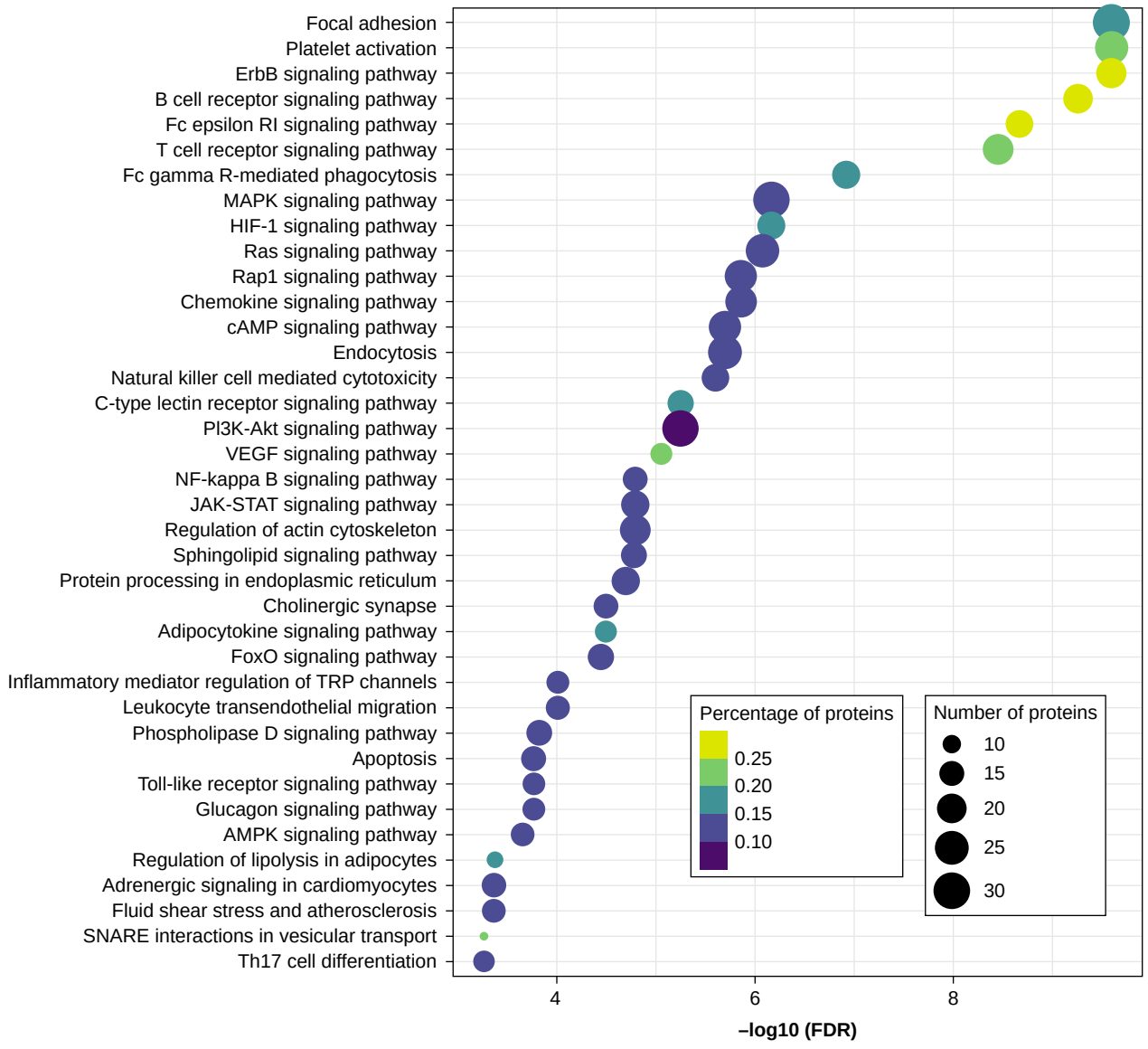


Figure 4 Pathways that were differentially regulated between patients with HCM who subsequently developed new-onset AF and those who did not. cAMP, cyclic adenosine monophosphate; ECM-receptor, extracellular matrix-receptor; EGFR, epidermal growth factor receptor; Fox, forkhead box protein; HIF, hypoxia inducible factor; JAK-STAT, Janus kinase signal transducer and activator of transcription proteins; MAPK, mitogen-activated protein kinase; NF-κB, nuclear factor-κB; PI3 K, phosphoinositide 3-kinase; Th17, T helper type 17; TRP, transient receptor potential; VEGF, vascular endothelial growth factor.

proteomics-based model performed significantly better even when ML methods were used to improve the accuracy of the clinical model in the present study.

The concentration of a small number of plasma proteins can be measured rapidly and at a low-cost (e.g. sandwich enzyme-linked immunosorbent assay). Thus, our panel of 12 predictive plasma proteins suggests a high potential to develop a clinically useful panel of circulating biomarkers that can be applied to facilitate early detection of AF in patients with HCM. In the future, it may become feasible to collect a blood sample at HCM diagnosis for assessment of the plasma proteome and prediction of future disease-related events, such as new-onset AF.

The Ras-MAPK signalling pathway, hypertrophic cardiomyopathy pathogenesis, and disease progression

This study demonstrated that the Ras-MAPK pathway and several of its upstream and downstream pathways were dysregulated at enrolment among patients with HCM who developed new-onset AF. Upregulation of the Ras-MAPK pathway is known to cause RASopathies, which are a group of syndromes that include HCM-like cardiac changes.[40] Previous case-control studies from our group have identified upregulation of the Ras-MAPK pathway in patients with HCM compared to several different control groups.[19,20] These suggest that there is an association

between the Ras-MAPK pathway and HCM pathogenesis. However, whether the Ras-MAPK pathway is related to HCM disease progression is not known. This study, which was conducted on a cohort of patients with pre-existing HCM, adds to this body of knowledge by demonstrating that the Ras-MAPK and related pathways are associated with not only HCM pathogenesis but also progression to AF. Furthermore, this study serves as the first to demonstrate an association between dysregulation of the Ras-MAPK pathway at baseline and new-onset AF in HCM.

Emerging evidence suggests that MAPK dysregulation promotes the development of arrhythmias, although the specific molecular mechanism underlying the arrhythmia development is yet to be understood.[41] Our results strengthen the claim that MAPK dysregulation underlies arrhythmia development, and we speculate that our discriminant proteins may serve as a trigger of AF pathological process.

Circulating proteins used to predict new-onset atrial fibrillation in hypertrophic cardiomyopathy

We identified a panel of 12 circulating proteins that demonstrated a high ability to predict new-onset AF in HCM in both the training and test sets. Consistent with our pathway analysis, which demonstrated dysregulation of the Ras-MAPK pathway in association with new-onset AF, several of our top discriminant proteins serve as upstream regulators of the MAPK pathway. Specifically, prosaposin is known to induce MAPK phosphorylation and activation in different cell types, and ablation of prosaposin is associated with MAPK inactivation.[42] Furthermore, heat shock protein 70 (hsp70) has cardioprotective properties, and its deletion has been implicated in activation of the MAPK pathway.[43] Lack of hsp70 has been reported to cause dysregulation in the Ca^{2+} content of the sarcoplasmic reticulum (SR), and loss of SR Ca^{2+} homeostasis is known to play an essential role in AF development in pressure-overloaded hearts.[44,45]

In addition, two of the proteins identified as part of the 12-protein panel have been found to be associated with AF in prior studies in non-HCM populations. In the Framingham Heart Study, lower concentrations of kallikrein were associated with incident AF.[46] Similarly, paired immunoglobulin-like type 2 receptor alpha has been identified in association with AF.[47,48] Our 12-protein panel also includes haptoglobin, which is a well-known marker for haemolysis.[49,50] Prior studies have suggested that LVOT obstruction in HCM may lead to haemolysis.[51] This phenomenon makes haptoglobin a potential biomarker for a severe form of HCM, as it correlates with a high level of LVOT obstruction. Importantly, we also found that patients with new-onset AF in our study had higher LVOT gradients than those without AF. Thus, haptoglobin serves as a positive control to confirm that our comprehensive proteomics profiling reduces the possibility of false negatives—i.e. failure to capture important biomarkers. Taken together, these prior studies buttress the plausibility of the findings in the present study both at the individual protein and the signalling pathway levels.

Strengths of the current study

This study employed a variety of strategies to strengthen internal and external validity and to reduce the rate of false positives and false negatives. First, to increase the external validity, the predictive ability of the training set in the ML model was validated using an external test set containing patients from a different institution than those in the training set. Second, the study employed the most comprehensive proteomics profiling to date (4986 proteins) on the largest number of patients with HCM ($n = 284$), ensuring that internal validity was maximized and that the possibility of missing important proteins (i.e. false negatives) was minimized.[19,20,52] Third, to further reduce the possibility of false positives, we only included proteins with $P < 0.05$ in our pathway analysis and only declared pathways with FDR < 0.001 to be positive. The pathway analysis ensures that the identified proteins are biologically interrelated, further bolstering the biological plausibility of identified proteins and reducing the likelihood of false positives. Fourth, the association between known cardiac biomarkers from the proteomics data set and the outcome of new-onset AF serves as a positive control, to confirm that the proteomics assay is not missing important predictors. Lastly, the accuracy of clinical event adjudication is further strengthened by the prospective cohort design of this study.

Limitations

The present study has several limitations. The first is potential misclassifications in the outcome event (i.e. AF in HCM). Although most patients underwent Holter monitoring or other types of cardiac monitoring, the possibility of ascertainment bias should be considered. The second is that generalizability to other HCM populations, such as those with less severe HCM symptoms, may be limited, as both hospitals that participated in this study are tertiary care institutions and referral centres for patients with advanced HCM symptoms. Validation of the findings from this study is warranted in a larger and more diverse patient population. Third, it is possible that sample acquisition time of the day affected protein concentrations, given the dynamic nature of proteins. Lastly, since myocardial samples were not available, molecular-level analysis with tissue specimens was not performed.

Conclusions

This study serves as the first to apply comprehensive proteomics profiling to identify plasma protein biomarkers to predict new-onset AF in patients with HCM. Moreover, we identified signalling pathways that were dysregulated in patients with HCM who developed new-onset AF. The prediction model may help physicians identify patients with HCM who are at high risk of developing new-onset AF, facilitate more frequent surveillance and earlier detection, and ultimately improve symptom burden and prevent stroke. Moreover, by elucidating dysregulated pathways associated with new-onset AF, this study contributes to our understanding of the molecular mechanisms underlying the development of AF in HCM.

Supplementary material

Supplementary material is available at *Europace* online.

Funding

This work was supported by the National Institutes of Health (R01 HL157216 and R01 HL168382 to Y.J.S., UL1 TR001873 and K24 HL107643 to M.P.R., and K24 AG036778 to M.S.M.), American Heart Association (two National Clinical and Population Research Awards, Career Development Award, and Transformational Project Award to Y.J.S.), Korea Institute of Oriental Medicine (W22005 to Y.J.S.), Feldstein Medical Foundation to Y.J.S., Columbia University Irving Medical Center Irving Institute for Clinical & Translational Research Precision Medicine Pilot Award to Y.J.S., and Columbia University Irving Medical Center Marjorie and Lewis Katz Cardiovascular Research Prize to Y.J.S. The funding organizations did not have any role in the study design, collection, analysis, or interpretation of data, in writing of the manuscript, or in the decision to submit the article for publication. The researchers were independent from the funding organizations.

Conflict of interest: Y.J.S. has received research funding from Bristol Myers Squibb and consulting income from Bristol Myers Squibb and Moderna Japan. M.S.M. has received consulting income from Akcea, Alnylam, Eidos Therapeutics, Pfizer, Prothena, Novo Nordisk, and Intellia. All remaining authors have declared no conflicts of interest.

Data availability

The data that support the findings of the present study are available from the corresponding author upon reasonable request. The machine learning model used in the present study will be publicly available on our website

(https://www.columbiacardiology.org/research-labs/shimada-lab) within 6 months of the publication date.

References

1. Maron BJ. Clinical course and management of hypertrophic cardiomyopathy. *N Engl J Med* 2018;**379**:655–68.
2. Semsarian C, Ingles J, Maron MS, Maron BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2015;**65**:1249–54.
3. Vaidya K, Semsarian C, Chan KH. Atrial fibrillation in hypertrophic cardiomyopathy. *Heart Lung Circ* 2017;**26**:975–82.
4. Lee SE, Park JK, Uhm JS, Kim JY, Pak HN, Lee MH *et al.* Impact of atrial fibrillation on the clinical course of apical hypertrophic cardiomyopathy. *Heart* 2017;**103**:1496–501.
5. Dragasis S, Vlachos K, Kariki O, Koskina S, Zygouri A, Patsiotis IG *et al.* Atrial fibrillation in hypertrophic cardiomyopathy—a contemporary mini-review. *Hellenic J Cardiol* 2022;**67**:66–72.
6. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV *et al.* Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. *JAMA* 2001;**285**:2370–5.
7. Olivotto I, Cecchi F, Casey SA, Dolara A, Traverse JH, Maron BJ. Impact of atrial fibrillation on the clinical course of hypertrophic cardiomyopathy. *Circulation* 2001;**104**:2517–24.
8. Rowin EJ, Hausvater A, Link MS, Abt P, Gionfriddo W, Wang W *et al.* Clinical profile and consequences of atrial fibrillation in hypertrophic cardiomyopathy. *Circulation* 2017;**136**:2420–36.
9. Siontis KC, Geske JB, Ong K, Nishimura RA, Ommen SR, Gersh BJ. Atrial fibrillation in hypertrophic cardiomyopathy: prevalence, clinical correlations, and mortality in a large high-risk population. *J Am Heart Assoc* 2014;**3**:e001002.
10. Zorner CR, Schjerning AM, Jensen MK, Christensen AH, Tfelt-Hansen J, Tonnesen J *et al.* Understanding the incidence of atrial fibrillation and stroke in hypertrophic cardiomyopathy patients: insights from Danish nationwide registries. *Europace* 2024;**26**:euae177.
11. Braunwald E, Seidman CE, Sigwart U. Contemporary evaluation and management of hypertrophic cardiomyopathy. *Circulation* 2002;**106**:1312–6.
12. Nasser MF, Gandhi S, Siegel RJ, Rader F. Anticoagulation for stroke prevention in patients with hypertrophic cardiomyopathy and atrial fibrillation: a review. *Heart Rhythm* 2021;**18**:297–302.
13. Ommen SR, Ho CY, Asif IM, Balaji S, Burke MA, Day SM *et al.* 2024 AHA/ACC/AMSSM/HRS/PACES/SCMR guideline for the management of hypertrophic cardiomyopathy: a report of the American Heart Association/American College of Cardiology joint committee on clinical practice guidelines. *Circulation* 2024;**149**:e1239–311.
14. Carrick RT, Maron MS, Adler A, Wessler B, Hoss S, Chan RH *et al.* Development and validation of a clinical predictive model for identifying hypertrophic cardiomyopathy patients at risk for atrial fibrillation: the HCM-AF score. *Circ Arrhythm Electrophysiol* 2021;**14**:e009796.
15. Lemesle G, Maury F, Beseme O, Ovart L, Amouyel P, Lamblin N *et al.* Multimarker proteomic profiling for the prediction of cardiovascular mortality in patients with chronic heart failure. *PLoS One* 2015;**10**:e0119265.
16. Zhang ZY, Thijs L, Petit T, Gu YM, Jacobs L, Yang WY *et al.* Urinary proteome and systolic blood pressure as predictors of 5-year cardiovascular and cardiac outcomes in a general population. *Hypertension* 2015;**66**:52–60.
17. Ngo D, Sinha S, Shen D, Kuhn EW, Keyes MJ, Shi X *et al.* Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation* 2016;**134**:270–85.
18. Bokhari S, Morgenstern R, Weinberg R, Kinkhabwala M, Panagiotou D, Castano A *et al.* Standardization of (99 m)Technetium pyrophosphate imaging methodology to diagnose TTR cardiac amyloidosis. *J Nucl Cardiol* 2018;**25**:181–90.
19. Shimada YJ, Hasegawa K, Kochav SM, Mohajer P, Jung J, Maurer MS *et al.* Application of proteomics profiling for biomarker discovery in hypertrophic cardiomyopathy. *J Cardiovasc Transl Res* 2019;**12**:569–79.
20. Shimada YJ, Raita Y, Liang LW, Maurer MS, Hasegawa K, Fifer MA *et al.* Comprehensive proteomics profiling reveals circulating biomarkers of hypertrophic cardiomyopathy. *Circ Heart Fail* 2021;**14**:e007849.
21. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 2010;**5**:e15004.
22. Kraemer S, Vaught JD, Bock C, Gold L, Katilius E, Keeney TR *et al.* From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. *PLoS One* 2011;**6**:e26332.
23. Hensley P. SOMAmers and SOMAscan—a protein biomarker discovery platform for rapid analysis of sample collections from bench top to the clinic. *J Biomol Tech* 2013;**24**:S5.
24. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw* 2010;**36**:1–13.
25. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, Gude-Sampedro F. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J Stat Softw* 2014;**61**:1–36.
26. Hassanzad M, Hajian-Tilaki K. Methods of determining optimal cut-point of diagnostic biomarkers with application of clinical data in ROC analysis: an update review. *BMC Med Res Methodol* 2024;**24**:84.
27. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;**47**:D607–D13.
28. Ezzeddine FM, Agboola KM, Hassett LC, Killu AM, Del-Carpio Munoz F, DeSimone CV *et al.* Catheter ablation of atrial fibrillation in patients with and without hypertrophic cardiomyopathy: systematic review and meta-analysis. *Europace* 2023;**25**:euad256.
29. Tzeis S, Gerstenfeld EP, Kalman J, Saad EB, Shamloo AS, Andrade JG *et al.* 2024 European Heart Rhythm Association/Heart Rhythm Society/Asia Pacific Heart Rhythm Society/Latin American Heart Rhythm Society expert consensus statement on catheter and surgical ablation of atrial fibrillation. *Europace* 2024;**26**:euae043.
30. Patel RC, Sweigart J, Sanchez JM, Cunningham JM. Clinical progress note: rhythm control for patients with atrial fibrillation. *J Hosp Med* 2022;**17**:194–8.
31. Turakhia MP, Shafrin J, Bognar K, Trocio J, Abdulsattar Y, Wiederkehr D *et al.* Estimated prevalence of undiagnosed atrial fibrillation in the United States. *PLoS One* 2018;**13**:e0195088.
32. MacIntyre C, Lakdawala NK. Management of atrial fibrillation in hypertrophic cardiomyopathy. *Circulation* 2016;**133**:1901–5.
33. Reiser H, Klingenberg R, Hof D, Cooksley-Decasper S, Fuchs N, Akhmedov A *et al.* Circulating FABP4 is a prognostic biomarker in patients with acute coronary syndrome but not in asymptomatic individuals. *Arterioscler Thromb Vasc Biol* 2015;**35**:1872–9.
34. Cheng JM, Akkerhuis KM, Meilhac O, Oemrawsingh RM, Garcia-Garcia HM, van Geuns RJ *et al.* Circulating osteoglycin and NGAL/MMP9 complex concentrations predict 1-year major adverse cardiovascular events after coronary angiography. *Arterioscler Thromb Vasc Biol* 2014;**34**:1078–84.
35. Berezin AE, Kremzer AA, Martovitskaya YV, Samura TA, Berezina TA, Zulli A *et al.* The utility of biomarker risk prediction score in patients with chronic heart failure. *Int J Clin Exp Med* 2015;**8**:18255–64.
36. Frustaci A, Ciccosanti F, Chimenti C, Nardacci R, Corazzari M, Verardo R *et al.* Histological and proteomic profile of diabetic versus non-diabetic dilated cardiomyopathy. *Int J Cardiol* 2016;**203**:282–9.
37. Hathout Y, Brody E, Clemens PR, Cripe L, DeLisle RK, Furlong P *et al.* Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. *Proc Natl Acad Sci U S A* 2015;**112**:7153–8.
38. Shi T, Moravec CS, Perez DM. Novel proteins associated with human dilated cardiomyopathy: selective reduction in alpha(1A)-adrenergic receptors and increased desensitization proteins. *J Recept Signal Transduct Res* 2013;**33**:96–106.
39. Ko D, Benson MD, Ngo D, Yang Q, Larson MG, Wang TJ *et al.* Proteomics profiling and risk of new-onset atrial fibrillation: Framingham Heart Study. *J Am Heart Assoc* 2019;**8**:e010976.
40. Rauen KA. The RASopathies. *Annu Rev Genomics Hum Genet* 2013;**14**:355–69.
41. Ai X, Yan J, Carrillo E, Ding W. The stress-response MAP kinase signaling in cardiac arrhythmias. *Rev Physiol Biochem Pharmacol* 2016;**172**:77–100.
42. Morales CR, Badran H. Prosaposin ablation inactivates the MAPK and Akt signaling pathways and interferes with the development of the prostate gland. *Asian J Androl* 2003;**5**:57–63.
43. Liu D, Han X, Zhang Z, Tse G, Shao Q, Liu T. Role of heat shock proteins in atrial fibrillation: from molecular mechanisms to diagnostic and therapeutic opportunities. *Cells* 2022;**12**:151.
44. Kim YK, Suarez J, Hu Y, McDonough PM, Boer C, Dix DJ *et al.* Deletion of the inducible 70-kDa heat shock protein genes in mice impairs cardiac contractile function and calcium handling associated with hypertrophy. *Circulation* 2006;**113**:2589–97.
45. Zhang Y, Qi Y, Li JJ, He WJ, Gao XH, Zhang Y *et al.* Stretch-induced sarcoplasmic reticulum calcium leak is causatively associated with atrial fibrillation in pressure-overloaded hearts. *Cardiovasc Res* 2021;**117**:1091–102.
46. Staerk L, Preis SR, Lin H, Lubitz SA, Ellinor PT, Levy D *et al.* Protein biomarkers and risk of atrial fibrillation: the FHS. *Circ Arrhythm Electrophysiol* 2020;**13**:e007607.
47. Shi W, Li X, Su Y, Liu D, Wu L, Li S *et al.* PILRA is associated with immune cells infiltration in atrial fibrillation based on bioinformatics and experiment validation. *Front Cardiovasc Med* 2023;**10**:1082015.
48. Zhou T, Liu J, Bao Y, Ling T, Lin C, Pan W *et al.* Soluble PILRalpha: a novel plasma biomarker for atrial fibrillation progression and recurrence after catheter ablation. *Clin Chim Acta* 2024;**553**:117703.
49. Tolosano E, Fagoonee S, Hirsch E, Berger FG, Baumann H, Silengo L *et al.* Enhanced splenomegaly and severe liver inflammation in haptoglobin/hemopexin double-null mice after acute hemolysis. *Blood* 2002;**100**:4201–8.
50. Nielsen MJ, Moestrup SK. Receptor targeting of hemoglobin mediated by the haptoglobins: roles beyond heme scavenging. *Blood* 2009;**114**:764–71.
51. Lin Y, Yu L, Liu F, Lin X, Li H, Xu X. Relationship between left ventricular outflow tract pressure gradient and hemoglobin in patients with hypertrophic cardiomyopathy. *Acta Cardiol Sin* 2020;**36**:343–50.
52. Captur G, Heywood WE, Coats C, Rosmini S, Patel V, Lopes LR *et al.* Identification of a multiplex biomarker panel for hypertrophic cardiomyopathy using quantitative proteomics and machine learning. *Mol Cell Proteomics* 2020;**19**:114–27.