Blood-based cardiometabolic phenotypes in atrial fibrillation and their associated risk: EAST-AFNET 4 biomolecule study

Larissa Fabritz[1,2,3,4,5], Winnie Chua[5], Victor R. Cardoso[5], Christoph Al-Taie[1,2,3], Katrin Borof[1,2], Anna Suling[6], Linda Krause[6], Shino Kany[1,3,7,8], Christina Magnussen[1,3,9], Karl Wegscheider[6], Guenter Breithardt[10], Harry J.G.M. Crijns[11], A. John Camm[12], George Gkoutos[5], Patrick T. Ellinor[7,8], Andreas Goette[13], Ulrich Schotten[4,14], Ursula-Henrike Wienhues-Thelen[15], Tanja Zeller[1,2,3], Renate B. Schnabel[1,2,3]Antonia Zapf[6], Paulus Kirchhof[1,2,4,5]

[1]Department of Cardiology, University Heart and Vascular Center Hamburg, University Medical Center Hamburg Eppendorf, Hamburg, Germany

[2]University Center of Cardiovascular Science, University Heart and Vascular Center Hamburg, University Medical Center Hamburg Eppendorf, Hamburg, Germany

[3]German Center for Cardiovascular Research (DZHK), partner site Hamburg/Kiel/Lübeck, Germany

[4]AFNET, Münster, Germany

[5]Institute of Cardiovascular Sciences, University of Birmingham, United Kingdom

[6]Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg Eppendorf, Hamburg, Germany

[7]Cardiovascular Disease Initiative, The Broad Institute of MIT and Harvard, Cambridge, MA, USA

[8]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA

[9]Center for Population Health Innovation (POINT), University Heart and Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[10]University Hospital Münster, Münster, Germany

[11]Department of Cardiology, University Hospital Maastricht, Maastricht, The Netherlands

[12]Clinical Sciences, St George´s University, London, UK

[13]Vincenz-Krankenhaus Paderborn, Germany

[14]Department of Physiology, Maastricht University, Maastricht, The Netherlands

[15]Roche Diagnostics, Penzberg, Germany

Short title: Biomolecule-based patient clusters in atrial fibrillation

Corresponding author:
Prof. Dr. Paulus Kirchhof
Department of Cardiology
University Heart and Vascular Centre Hamburg-Eppendorf
University Medical Center Hamburg Eppendorf
Martinistraße 52
20246 Hamburg

Phone: +49 (0)40 7410-52438
Fax: +49 (0)40 7410-55862
Email: p.kirchhof@uke.de
www.uhz.de

Word count: 4722 (introduction – end of disclosures)

Abstract

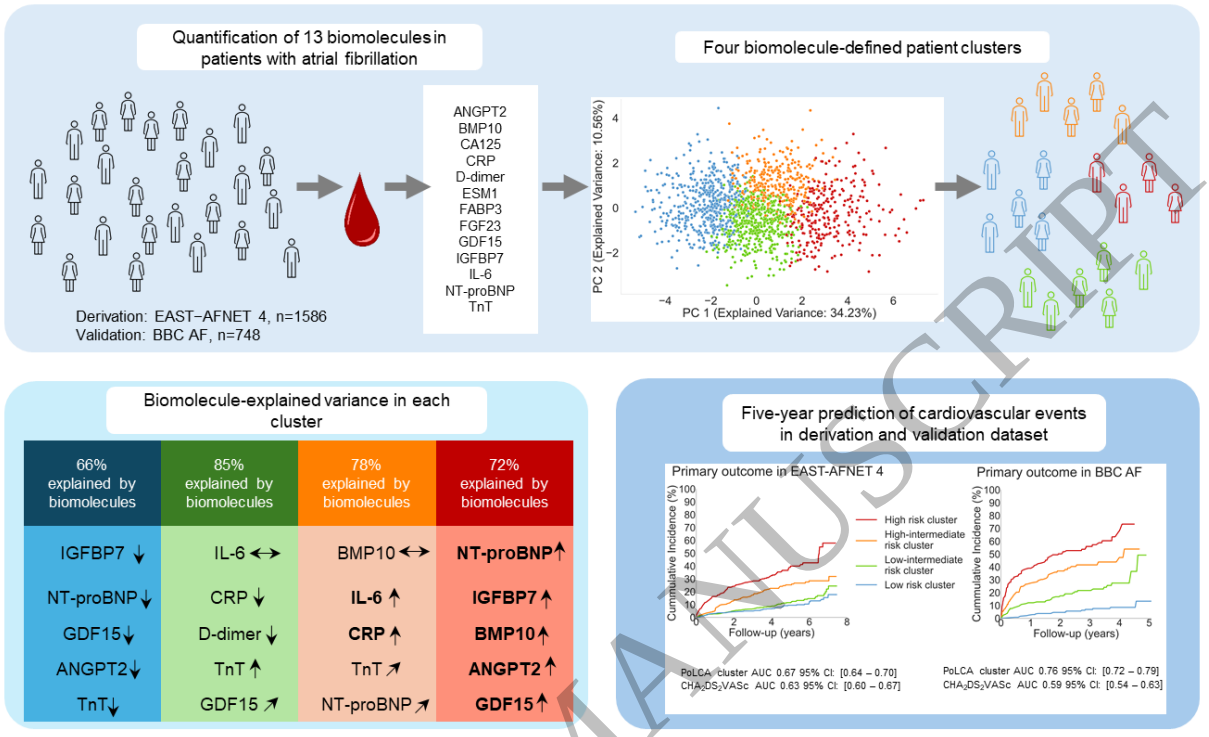
Background Atrial fibrillation (AF) and concomitant cardiometabolic disease processes interact and combine to lead to adverse events such as stroke, heart failure, myocardial infarction, and cardiovascular death. Circulating biomolecules provide quantifiable proxies for cardiometabolic disease processes. Their role in defining subphenotypes of AF is not known.

Methods and results This prespecified analysis of the EAST-AFNET4 biomolecule study assigned patients to clusters using polytomous variable latent class analysis (poLCA) based on baseline concentrations of thirteen precisely-quantified biomolecules potentially reflecting ageing, cardiac fibrosis, metabolic dysfunction, oxidative stress, cardiac load, endothelial dysfunction, and inflammation. In each cluster, rates of cardiovascular death, stroke, or hospitalization for heart failure or acute coronary syndrome, the primary outcome of EAST-AFNET 4, were calculated and compared between clusters over median 5.1 years follow-up. Findings were independently validated in a prospective cohort of 748 patients with AF (BBC-AF; median follow up 2.9 years).

Unsupervised biomolecule analysis assigned 1586 patients (71 years old, 46% women) into four clusters. The highest-risk cluster was dominated by elevated BMP10, IGFBP7, NT-proBNP, ANGPT2 and GDF15. Patients in the lowest-risk cluster showed low concentrations of these biomolecules. Two intermediate-risk clusters differed by high or low concentrations of hsCRP, IL-6, and D-dimer. Patients in the highest-risk cluster had a 5-fold higher cardiovascular event rate than patients in the low-risk cluster. Early rhythm control was effective across clusters ($p_{interaction}$=0.63). Sensitivity analyses and external validation in BBC-AF replicated clusters and risk gradients.

Conclusions Biomolecule concentrations identify cardiometabolic subphenotypes in patients with atrial fibrillation at high and low cardiovascular risk.

Graphical abstract



Quantification of thirteen biomolecules preselected for their ability to provide quantitative proxies for cardiovascular disease processes assigns patients with recently diagnosed AF and comorbidities to four clusters. The main biomolecules contributing to the clustering process are NT-proBNP, IGFBP7, BMP10, AngPT2, GDF-15, IL-6, and CRP. These biomolecules explain 66%-85% of the assignment to a cluster. The four clusters differ in their cardiovscular risk over a five-year time horizon. Early rhythm control therapy is effective in all four clusters ($p_{interaction}$=0.63).

Introduction

Several chronic and interacting disease processes[1,2] contribute to the development of atrial fibrillation (AF) and to cardiovascular events in patients with AF [3,4]. More than 80% of patients with atrial fibrillation (AF) suffer from comorbidities such as hypertension, heart failure, diabetes, or coronary, cerebral or peripheral artery disease at the time of diagnosing AF [4,5]. These include cardiometabolic dysfunction [6], systemic and atrial thrombo-inflammation [7,8], vascular and endothelial dysfunction [7,8] and cardiomyocyte dysfunction [9]. Quantification of underlying disease processes and their interactions may identify drivers of cardiovascular complications in patients with AF.

Circulating biomolecules provide quantitative proxies of cardiovascular disease processes including at early, subclinical stages [1,2]. For example, slight, chronic elevations of circulating troponin concentrations are associated with sub-clinical myocyte injury and increased cardiovascular risk [10], including in patients with atrial fibrillation [11]. Quantification of selected biomolecules in a single blood draw can furthermore refine prediction of stroke and bleeding risk in patients with AF [11-13]. Whether a selection of biomolecules aiming to represent different cardiovascular and inflammatory disease processes can be used to identify cardiometabolic subphenotypes of patients with AF has not been tested.

This prespecified secondary analysis of the EAST-AFNET 4 biomolecule study embedded into the **E**arly treatment of **A**trial fibrillation for **ST**roke prevention (EAST-AFNET 4) trial [14], quantified thirteen biomolecules reflecting different diseases processes in AF that were defined a priori [1]. Unsupervised clustering methods capturing interactions between biomolecules were applied to identify patients at risk of cardiovascular events based on biomolecule concentrations. Additionally, independent validation was performed in a prospective registry of patients with AF (BBC-AF [15]).

Methods

The main analyses described here report a prespecified analysis of the EAST-AFNET 4 biomolecule study (see protocol appendix) [16]. Post-hoc exploratory analyses were added to better understand the main findings.

Derivation cohort (EAST-AFNET 4). The Early treatment of atrial fibrillation for stroke prevention trial (EAST-AFNET 4) randomized patients with recently diagnosed AF and stroke risk factors to systematic early rhythm control or usual care including symptom-based rhythm control [14]. All patients were followed-up for a median of 5.1 years. The primary outcome of the trial was a composite of stroke, cardiovascular death, and unplanned hospitalization for heart failure or acute coronary syndrome [14]. Details of the EAST-AFNET 4 biosample study collecting a baseline blood sample in 1586 patients enrolled in the EAST-AFNET 4 trial have been published [17]. In brief, all consenting patients provided a blood sample at baseline. Samples were shipped to the core biostorage facility at UKE Hamburg, spun, shock-frozen and stored at -80C. EAST-AFNET 4 and its biomolecule substudy were approved at all study sites. Written informed consent was obtained from all patients. This study complied with the Declaration of Helsinki.

Validation cohort (BBC-AF atrial fibrillation sub-cohort). Details of the BBC-AF cohort have been described before [15]. In brief, consecutive patients eligible for recruitment had ECG-diagnosed AF or presented with at least two cardiovascular conditions (congestive heart failure, hypertension, diabetes, prior stroke, or vascular disease) to a large teaching hospital (Sandwell and West Birmingham NHS Trust). Patients who did not have a diagnosis of AF underwent 7-day ambulatory ECG monitoring to rule out undiagnosed ECG-documented AF. For this analysis, only patients with ECG-documented AF were included. All patients underwent a detailed interview, physical examination, 12-lead ECG, and a transthoracic echocardiography at time of recruitment. Follow-up data for events were collected by assessing local hospital records corroborated against Hospital Episode Statistics data, general practitioner (GP) records, and mortality data from NHS Digital, at 2.5 years after the final patient was recruited [18]. The follow-up duration was calculated as the time between the baseline assessment date to an event, or to the date of record review where no events were documented. This study complied with the Declaration of Helsinki, was approved by the National Research Ethics Service Committee (IRAS ID 97753) and was sponsored by the University of Birmingham. All patients provided written informed consent.

Selection of biomolecules and their quantification. The methodological approach taken for this work tried to combine the multifaceted information of several circulating biomolecules and the reliability of high precision assays. A group of scientists within the EU-funded CATCH-ME consortium searched literature and scouted meetings for disease mechanisms related to atrial fibrillation and to cardiovascular events in patients with atrial fibrillation. One summary of this initial effort has been published [1]. In a next step, biomolecules that could potentially reflect these disease processes (called "health modifiers" in [1]) were identified based on a literature and patent search enriched with knowledge available in the EU Horizon 2020 CATCH-ME consortium. A modified Delphi expert consensus process was conducted to identify biomolecules representing these disease mechanisms available forhigh precision quantification. Thirteen biomolecules were identified (Table 1): Angiopoietin 2 (ANGPT2), bone morphogenetic protein 10 (BMP10), cancer antigen 125 (CA125), C-reactive protein (CRP), D-dimer, endothelial specific molecule 1 (ESM1), fatty acid binding protein 3 (FABP3), fibroblast growth factor 23 (FGF23), growth differentiation factor 15 (GDF15), insulin-like growth factor binding protein 7 (IGFBP7), interleukin-6 (IL-6), N-terminal pro–B-type natriuretic peptide (NT-proBNP), and cardiac troponin (TnT).

Biomolecules were centrally quantified using pre-commercial and commercial high-throughput, high-precision platforms (Roche, Penzberg, Germany). The biomolecule quantification was provided as an in-kind contribution of Roche to the CATCH ME consortium. Absolute protein concentrations were centrally quantified in EDTA plasma. Run controls and calibrators were measured twice each run, and lab staff involved were blinded to clinical status and data. Blood samples were shipped to and quantifications were conducted at the Roche biomolecule research facility in Penzberg, Germany. This is the first analysis of the biomolecules in the EAST-AFNET 4 trial substudy.

Data preprocessing and clustering. Biomolecule concentrations were one-percent winsorized and Blom-transformed[19] and each patient was assigned into one quintile for each biomolecule. These quintiles were used to cluster patients using poLCA. K-means clustering was used as sensitivity analysis. Unsupervised models were developed using latent-class analysis (LCA), available in the package *poLCA* [20] in R (https://www.r-project.org/). Latent class analysis was performed on 13 biomarker variables. Patients with any missing biomolecule concentrations were excluded. The models were created in a bootstrapping fashion by repeating 100 times with data resampling with replacement. Each data resampling was performed with a fixed initialization seed to ensure reproducibility. Models between 2 and 10 clusters were assessed. The Bayesian information criterion (BIC) was used to assess the best number of clusters by penalizing models with too many parameters. The number of clusters with the lower BIC score was counted over all bootstraps.

This most frequent number of clusters was used to create the final model using the original data, without resampling. We also fitted k-means clustering models against the same Blom-transformed biomolecules without conversion of those into categorical variables. To assess the optimal number of cluster groups we followed the exact same approach as for poLCA models. As base model-instance, we made use of SciKit-learns algorithm(https://scikit-learn.org/stable/index.html) with Lloyd algorithm [21], k-means++ [22] as initial cluster centroid selection strategy and let the algorithm run for 10 iterations with different centroid seeds to fit the model against our dataset.

Phenotypic description. Clusters were formed agnostic to any clinical data. Patients' characteristics were summarized for each cluster. BMI was categorized into obese (BMI ≥30) and non-obese. The eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration equation[23, 24] and categorized into normal kidney function (eGFR≤60 ml/min) and chronic kidney disease. Left-ventricular ejection fraction was categorized into groups of ≥50% and <50%. Differences between categorical variables were calculated using generalized logistic mixed models with study Center as random effect. For continuous variables, linear mixed models with study Center as random effect were used for normally distributed and non-normally distributed variables. P-values resulted from Analysis of Deviance Table (Type II Wald chi^2 tests). The R packages lme4 and car were used for this analysis. In the non-multi-centric BBC AF cohort, t-test and Chi^2 test were applied for quantification of differences between continuous respectively categorical variables.

Communality analysis. Biomarkers are biologically secreted and reabsorbed reflecting different disease processes. They are excreted or shed based on common pathways (e.g. secretion via the kidney), creating collinearity in the data. The presence of multicollinearity, as can be demonstrated using a correlation matrix or calculating the variance inflation factor (VIF), complicates the interpretation of regression model outcomes since both unique and shared variances are contributing to an effect on the outcome. Communality analysis allows an exploration of relationships between biomarkers by decomposing the R^2 of the regression model or respectively the pseudo R^2 of the binomial regression model to quantify unique and shared variances of each biomarker in explaining the outcome. The analysis returns 2^k-1 communality coefficients (k = number of variables entered).

Dominance analysis. As communality analysis is one way to assess the relative importance of predictors (p) (the 13 biomolecules) on an outcome (the cluster group), dominance analysis (DA) is a different approach that makes two distinct contributions. Firstly, it measures the relative

importance of a predictor in a pairwise fashion and secondly, it does this in the context of all $2^{(p-2)}$models that contain any subset of the remaining predictors. In a refined version of DA by Azen & Budescu [25], the concept of complete, conditional and general dominance was introduced. We employed the Dominance-Analysis Python package [26] (Python 3: https://www.python.org/) that refines this concept further by introducing individual dominance, average partial dominance, interactional dominance and total dominance. Finally, the percentage relative importance can be calculated from the mean of those four dominance measures for each predictor variable.

Change in clustering after removal of biomolecules. To generate a more global understanding of the contribution of each biomolecule to the clustering, the patient clustering was repeated with reduced feature sets by removing one, two, three, four, or five of the biomolecules. For each possible biomolecule combination of the reduced feature sets we estimated the optimal number of clusters, predicted the clusters and used those to partition the patients again. We computed the adjusted rand index (ARI) for those new partitions in comparison with the original ones derived from the poLCA model fitted against 13 biomolecules and predicting four clusters. The biomolecule clusters were used as predictors in Cox proportional hazards (PH) model instances with the first primary composite outcome as event of interest to obtain Hazard ratios and c-index.

Risk of cardiovascular events. Cox proportional hazard (PH) models were fitted using cluster membership as the predictor to predict a composite outcome of cardiovascular death, heart failure hospitalization, stroke or systemic embolism, and acute coronary syndrome. To infer hazard ratios we used models with center as shared frailty term and the R package Survival. For sensitivity analyses we added age, sex and randomization group as confounding variables into the models. To compare the unsupervised cluster assignment with existing predictors, separate risk prediction models were built using a) *CHA_2DS_2-VASc score*, *b) ABC stroke* [12] *and bleeding* [13] *scores, c)* discretized [18] Troponin T and *d) NTpro-BNP* quartiles. For the ABC scores, published criteria [12,13] were computed.

In the BBC AF validation cohort, there are 68 missing values for the first primary composite outcome and 59 missing values for the first primary safety outcome for either the event-status information or the time-to-event information. We dropped those participants from the primary analysis, but added a best and worst-case scenario analysis. For the best-case scenario all missing event values have been imputed by one (occurrence of an event) and for the best-case scenario with zero (no occurrence of an event). We imputed all missing time-to-event data by the median censoring time.

To calculate the Area under the ROC curve we fitted unadjusted Cox PH models without center as frailty as this information is missing in the validation dataset and our aim was to measure discriminatory power of inter-cohort generalizable models. We used the Python lifelines[27] and Sklearn packages for this analysis. The ABC scores[12,13] and genomic risk scores were used as continuous variables and all other predictors discretized as proposed in the literature (details in legend to Figure 5).

Genomic risk scores. Genomic DNA extraction was performed from buffy coat samples derived from EDTA blood samples. DNA samples were sent to the Broad Institute of MIT and Harvard in Cambridge, MA, USA. After quality control of the DNA, array genotyping (Infinium PsychArray-24 v1.2 BeadChip) and imputation with the TOPMed Freeze 5 dataset as reference was performed. Previously published Polygenic risk scores (PRS) for the risk of AFAF (PRS-AF) and ischemic stroke risk (PRS-Stroke) were computed using PLINK2[16]. The sum scores were obtained, and PRS calculations were based on TOPMed imputed genotype dosages, ensuring an imputation quality measure exceeding 0.3 for each variant on every chromosome. Following quality control and imputation, a total of 6,363,335 single nucleotide variants (out of 6,730,541) were utilized to calculate PRS-AF, and 516,013 single nucleotide variants (out of 530,933) were employed for PRS-Stroke calculation.

Results

Biomolecule concentrations define four distinct clusters of patients. Clinical features of the EAST-AFNET 4 biomolecule study were similar to the patient population enrolled in the main trial (Table 2). Unsupervised clustering of patients based on concentrations of the 13 biomolecules without any clinical information identified four distinct patient clusters (Figure 2A) with overlapping clinical characteristics (Table 2).

Almost all patients were clearly assigned to a cluster (Figure 2A). In the validation data set BBC-AF, the classification criteria derived in the EAST-AFNET 4 data set sorted patients in the validation set into similar clusters with similar sizes (Figure 2B). The definition of four clusters was robust within poLCA and also when another method, K-means, was applied to the data set (Figure 2C).

The cluster later shown to be the high-risk cluster was dominated by elevated BMP10, IGFBP7, NT-proBNP, ANGPT2 and GDF15. Patients in the cluster with the lowest risk of cardiovascular events showed low concentrations of these biomolecules. Two intermediate-risk clusters differed by high or low concentrations of thrombo-inflammatory markers (hsCRP, IL-6, D-dimer, Figure 2D).

Clinical features differed between clusters, illustrated e.g. by ages between 68 years (low-risk cluster) and 72-74 years (low-intermediate risk cluster, high-intermediate risk cluster, high-risk cluster, Table 2) or differences in rates of obesity (the intermediate-high risk cluster had the highest rate of obese patients, Table 2). The estimated CHA_2-DS_2-VASc score was similar between three of the four clusters. Only the lowest risk cluster included younger patients with fewer clinical stroke risk factors (Table 2). Clinical features were slightly different in the validation data set, BBC AF (supplementary Table 1). The distribution of clinical features across clusters was similar in BBC-AF (supplemental Table 2). Biomolecule concentrations are shown for each cluster in Table 3.

Risk of outcome events in each cluster. Each cluster had a distinct risk of primary outcome and safety outcome (Figure 3A, Figure 3C). Patients in the highest risk cluster had a five-fold higher rate of primary outcomes than patients in the lowest risk cluster in the derivation (Figure 3A) and validation (Figure 3B) data sets. Each component of the composite outcome moved in the same direction as the composite for the primary outcomes (Table 4a, Figure 3C) and for the safety outcome (Table 4b, Figure 3D). The clustering using k-means resulted in a similar risk gradient across clusters (Supplementary Figure 3). Early rhythm control was effective across all biomolecule clusters ($p_{interaction}$=0.63, Supplementary Table 5). The clustering model outperformed other risk scores for most of the tested outcomes (supplementary Figure 5 – 8). For the first primary composite outcome the Po-LCA cluster model yields an AUC 0.76 [95% CI: 0.72 – 0.79], the next best predictive model uses ABC bleeding score and yields an AUC 0.74 [95% CI: 0.70 – 0.78] in the validation dataset. Hazard ratios for the clustering were higher or similar to hazard ratios obtained by applying established risk predictions models using clinical features, combinations of clinical features and biomolecules, or a single biomolecule (Table 5).

Biomolecule combinations are required to assign patients to cluster groups. To estimate the relevance of each biomolecule for the assignment to a patient cluster, we computed the adjusted rand index and the C statistic for each combination of biomolecules as further post-hoc analyses. Removal of five or more biomolecules consistently yielded adjusted rand indices of 0.55 or less, indicating that almost half, or more, of the patients were no longer assigned to their original cluster (Figure 4A). Figure 4B plots the adjusted rand index and the C statistic for each possible model with fewer biomolecules. While risk prediction remains reasonable with only 2-3 biomolecules (C statistic estimates 0.67 to 0.69), assignment to cluster requires more information.

To estimate the relevance of each biomolecule for the assignment of patients to the risk clusters, another in-silico exercise was performed: For each number of biomolecules, the five models with the lowest adjusted rand indices were selected. The missing biomolecules were listed and counted. Biomolecules whose removal often leads to a low adjusted rand index were considered relevant for the clustering process. Figure 4C provides a list of all biomolecules sorted by the number of relevant removals in this exercise.

Sensitivity analyses. Removing or adding biomolecules using forward and backward selection resulted in similar rankings of biomolecules (Supplementary Figure 4, Supplementary Table 6). For each cluster, the unique and common contribution to the clustering was calculated (Supplementary Tables 7-8). The number of biomolecule-based clusters remained constant at four clusters when one to five biomolecules were randomly removed from the data set (Supplementary Table 9). These analyses identify several key biomolecules relevant for patient clustering, including IGFBP7, NT-proBNP, BMP10, ANGPT2, and the thrombo-inflammatory biomolecules CRP, IL-6, and D-dimer (Figure 4C).

Discussion

Main findings. Integrating information contained in thirteen biomolecules that were selected as potential quantifiable proxies for different disease processes with relevance in atrial fibrillation defines four distinct clusters of patients with atrial fibrillation. Each cluster has a unique biomolecule pattern and cardiovascular risk profile. The findings were robust in sensitivity analyses and in an independent prospective cohort of patients with AF. They identify shared disease mechanisms in sub-phenotypes of AF.

Clustering patients based on biomolecules potentially reflecting overlapping disease processes, as done by the unsupervised analyses used here, suggests possible disease mechanisms related to cardiovascular complications in patients with AF. The clustering may enable development of stratified therapies that may differ in patients with similar clinical features by highlighting treatable underlying disease processes linked to cardiometabolic dysfunction and load (BMP10, NT-proBNP, IGFBP7), endothelial dysfunction and shear stress (ANGPT2, BMP10) and increased thrombo-inflammation (CRP, IL-6, D-dimer). Six biomolecules related to atrial cardiomyocyte dysfunction and vascular smooth muscle cell dysregulation (BMP10) [9, 28], endothelial cell dysfunction (IGFBP7 [29, 30], ANGPT2 [31]), atrial and ventricular volume load (NT-proBNP), myocardial metabolism (FABP3) and mitochondrial dysfunction (GDF15) contributed most to the

biomolecule-based clustering. Patients at high risk showed elevated biomolecule concentrations related to cardiomyocyte dysfunction, disturbed metabolism, and increased endothelial stress. Intermediate risk patients were further differentiated into intermediate-high and intermediate-low risk by elevated concentrations of thrombo-inflammatory biomolecules (CRP, IL-6, d-dimer). Patients with low concentrations of these biomolecules have a very low event rate on current therapy. The results were similar using different unsupervised clustering techniques (Figure 2C), in sensitivity analyses, and in an independent data set (BBC-AF, 748 patients, Figures 2+3). Pending further validation, the results highlight that patients with AF can be stratified using circulating biomolecules without added clinical information. The distinct signature of biomolecules in each cluster suggests that treatments beyond oral anticoagulation, treatment of concomitant conditions and early rhythm control may be needed to further reduce their risk.

This work was performed in two data sets of patients with atrial fibrillation and cardiovascular comorbidities (Table 2). Some have argued that patients with atrial fibrillation are a model population for elderly patients with multiple cardiovascular diseases. While the present results show that the biomolecule clusters identified here define groups of patients in AF with distinct biomolecule patterns and outcome risks, it is conceivable that similar biomolecule patterns and outcome associations can be found in cardiovascular patients without AF. Our findings call for future research into the effects of biomolecules on cardiovascular function. IGFBP7 (also called angiomedullin) is released following activation of TGF-beta in fibroblasts and in cardiomyocytes [29], including in heart failure [30]. Its elevation highlights cardiac fibroblasts as a potential target for risk-reducing therapies in atrial fibrillation. Further research into the reasons for ANGPT2 elevations in patients with atrial fibrillation may identify treatable disease mechanisms [31]. Its relevance for patient clustering, especially in context with IGFBP7 and BMP10, suggests relevant interactions between endothelial cells and cardiomyocytes. Some of the biomolecules used here are associated with systemic or general cardiovascular disease mechanisms. Future studies are needed to understand associations of the biomolecule clusters identified here with cardiac rhythm and with outcomes in patients without AF. Such work will determine to what extent the clusters identified here are specific in their application to patients with AF.

A growing array of medical [14,32], interventional [14,33], and surgical [34] treatment options in patients with AF illustrates the need to identify treatable, risk-modifying disease processes in these patients. The promising effects of SGLT2 inhibitors on AF and first results on PPAR1 inhibitors on preventing and reversing AF hold promise for metabolic interventions [35]. Our analysis suggests to first test such interventions in patients assigned to the high-risk cluster in this analysis.

This analysis was not designed to select patients for a specific therapy, but the patient clusters can potentially form a basis to select therapy responders: BMP10 is almost exclusively secreted from atrial cardiomyocytes [9, 36] and secreted BMP10 regulates vascular smooth muscle cells [28], rendering atrial-specific therapies such as rhythm control, but also antihypertensive therapy and metabolic interventions useful in patients with elevated BMP10 concentrations [9, 37, 38]. Reducing inflammation using specific interleukin-targeting antibodies such as canakinumab [39] or the general anti-inflammatory agent colchicine [40-42] may be most effective in patients in the intermediate-high risk cluster defined by inflammatory biomolecules.

Comparison to other risk estimation scores and limitations of the C statistic. As expected, the biomolecule-based clustering process evaluated here provides better risk estimation than the CHA_2DS_2VASc score [12]. Its C statistic was better than or comparable to other proposed risk scores, including the ABC stroke and bleeding scores [12, 13] (Figure 4B). In view of the summative nature of the C statistic, this may not come as a surprise. The UK Biobank provided first insights into the added value of multiple biological measurements for risk prediction [43]. Previous work on biomarkers tested their predictive value of a biomolecule when added to clinical characteristics [44], or for single biomolecules on their own [10, 45], prior to combining biomolecules into scores [12, 13]. Classical statistical methods, including forward and backward selection, did not identify these biomolecules, probably due to a different handling of shared and common information.

Comparison to proteomic methods. Proteomic technologies now enable quantification of thousands of proteins from a small sample of plasma or blood. Earlier iterations of these technologies contributed to the discovery of AF-related biomolecules quantified here, e.g. FGF23 [15], while RNA sequencing contributed to the discovery of BMP10 [9] as a biomolecule of interest in patients at risk of AF. Others used proteomic analyses in all-encompassing analyses of circulating proteins related to cardiovascular events in patients with atrial fibrillation [11]. Such proteomic analyses are hypothesis-free but necessarily highlight proteins with large concentration ranges that can be quantified at high precision using proteomic technologies. These proteomic methods will be extremely helpful in identifying additional proteins related to AF. Such work is likely to confirm, refine and extend the present findings. The present analysis preselected thirteen biomolecules hypothesized to reflect different modifiable disease processes that can be quantified at high precision. These biomolecules were used to identify groups of patients who share pathophysiological patterns based on these biomolecules. This method identifies groups of patients with different predominant disease mechanisms. It may be useful alongside continued hypothesis-free research aiming to identify additional disease mechanisms leading to age-related diseases [46] and to chronic cardiovascular diseases [11].

Strengths and limitations. The biomolecule-based clusters developed here are agnostic to clinical information. They can be used to identify disease processes and to estimate risk in anonymized samples without clinical information and in settings where clinical assessment is not available or feasible. Another feature of the present clustering is its ability to identify patients with AF at risk of cardiovascular events beyond stroke. This broadens the potential therapeutic benefits for patients. A novel methodology chosen here is the preselection and simultaneous quantification of thirteen distinct biomolecules chosen for their potential relevance in atrial fibrillation [1]. Biomolecules were identified in a semi-formalized a priori process and centrally quantified using high-precision assays. While this can be viewed as a strength in view of the selection process, the precision of the measurements, and the disease processes reflected by these biomolecules, it is also a weakness as it limits the analyses to these biomolecules. Another strength of the analysis is the collection of samples in a broad range of care setting in a cohort of adequately treated patients with AF in the context of a clinical trial with centrally adjudicated outcomes and externally monitored data collection, and external validation in a cohort of patients with atrial fibrillation enrolled in a routine care setting.

The study has important limitations: One, assessment of smaller number of biomolecules is limited to in-silico calculations. Two, almost all patients received guideline-recommended anticoagulation, rate and rhythm control, and often effective treatment of concomitant conditions. The clusters presented here will require independent assessment in patients not receiving these therapies, which might be difficult given the ethical need to treat patients according to evidence. Three, a limitation is the lack of follow-up samples that would enable assessment of treatment effects, the lack of an untreated population of patients with AF, and lack of validation in data sets of patients without AF. Four, while BMP10, NT-proBNP, FABP3 and troponin are proteins released by the heart, the other markers are more systemic in nature, and cannot differentiate between cardiac, vascular, and other origins of the measured biomolecules. On the other hand, especially the vascular and inflammatory molecules might reflect ongoing systemic changes associated with cardiovascular outcomes unrelated to cardiac defects. Five, quantification of plasma biomolecules in a single sample may have missed smaller, but pathophysiologically relevant changes in the heart or atria diluted by systemic production and elimination of circulating biomolecules. Six, this study is limited to the 13 biomolecules quantified. Sequencing of cardiac tissue [47], quantification of circulating RNAs and advanced proteomics [48,49] enable hypothesis-free quantification of many molecules at once. These methods will discover additional molecules and may help to refine the disease processes suggested in this analysis. Seven, this study cannot evaluate whether the biomolecule combinations identified here

truly identify patients who are likely to respond to cardiometabolic, anti-inflammatory or other disease-process-modifying therapies. This will need prospective testing, e.g. by using the biomolecule clusters identified here as inclusion criteria in interventional trials. Eight, the EAST-AFNET 4 cohort is predominantly of Caucasian ethnicity. Validation in other ethnicities is needed. Nine, while some of the biomolecules can be measured in clinical routine as in-vitro diagnostic devices (IVDs) with regulatory approval, some other assays are not approved for use in clinical routine and available for research use only. Ten, clinical features were not used for clustering. This enables application to samples without precise information on clinical features but limits cause-specific interpretation.

In conclusion, these findings open the possibility that preselected plasma biomolecules as studied here and unbiased plasma multiomics can define distinct AF subphenotypes and thereby advance the management of this condition. Future studies are needed to determine whether such sub-phenotypes can be used to select therapies and to identify therapy responders.

Funding: The EAST-AFNET 4 trial and its biomolecule study were funded by BMBF, DZHK, AFNET, EHRA, St. Jude Medical–Abbott, Sanofi, and the German Heart Foundation. The biomolecule analysis was funded by EU 633196 [CATCH ME] and EU 965286 [MAESTRIA]. Further support came from EU IMI 116074 [BigData@Heart], British Heart Foundation (PG/20/22/35093; AA/18/2/34218), German Center for Cardiovascular Research supported by the German Ministry of Education and Research (DZHK, grant numbers DZHK (FKZ 81X2800182, 81Z0710116, and 81Z0710110), German Research Foundation (Ki 509167694), and Leducq Foundation.

LF received institutional research grants by EU 633196 [CATCH ME] and EU 965286 [MAESTRIA]. British Heart Foundation (AA/18/2/34218), German Center for Cardiovascular Research supported by the German Ministry of Education and Research (DZHK). PE is supported by grants from the National Institutes of Health (RO1HL092577, RO1HL157635) and from the American Heart Association (18SFRN34230127, 961045), US by grants of the Dutch Heart Foundation (CVON2014-09, RACE V Reappraisal of Atrial Fibrillation: Interaction between hyperCoagulability, Electrical remodeling, and Vascular Destabilisation in the Progression of AF, and Grant number 01-002-2022-0118, EmbRACE: Electro-Molecular Basis and the theRapeutic management of Atrial Cardiomyopathy, fibrillation and associated outcomEs), SK by a Walter Benjamin Fellowship from the Deutsche Forschungsgemeinschaft (521832260). CM receives study-specific funding from the German Center for Cardiovascular Research (DZHK; Promotion of women scientists programme; *FKZ 81X3710112*), the *Deutsche Stiftung für Herzforschung*, the *Dr. Rolf M. Schwiete Stiftung*, NDD, and Loewenstein *Medical* unrelated to the current work. RBS has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under the grant agreement No 648131, from the European Union's Horizon 2020 research and innovation programme under the grant agreement No 847770 (AFFECT-EU), from the European Union's Horizon Europe research and innovation programme under the grant agreement ID: 101095480 and German Center for Cardiovascular Research (DZHK e.V.) (81Z1710103 and 81Z0710114); German Ministry of Research and Education (BMBF 01ZX1408A) and ERACoSysMed3 (031L0239). Wolfgang Seefried project funding German Heart Foundation.

Conflict of interest: WC is an employee of ICF (www.icf.com). VRC is an employee of Citibank. Both WC and VRC performed the main work described here while employed by University of Birmingham. AJC receives personal funds from Acesion, Incarda, Menarini, Milestone, Sanofi, Bayer, Anthos, Daiichi Sankyo, Pfizer, Abbott, Biosense Webster, Biotronik, Boston Scientific, Medtronic, Johnson and Johnson. PE receives sponsored research support from Bayer AG, IBM Research, Bristol Myers Squibb, Pfizer and Novo Nordisk; PE has also served on advisory boards or consulted for Bayer AG. HJGMC discloses advisory board fees from InCarda Therapeutics, Roche Diagnostics, Daiichi Sankyo, Sanofi, Acesion and Atricure. Speaker fee from Medtronic. US received consultancy fees or honoraria from Università della Svizzera Italiana (USI, Switzerland) and Roche Diagnostics (Switzerland). US was supported by a grant from EP Solutions Inc. (Switzerland) and is co-founder and shareholder of YourRhythmics BV, a spin-off company of the University Maastricht. CM has received speaker fees from AstraZeneca, Novartis, Boehringer Ingelheim/Lilly, Bayer, Pfizer, Sanofi, Aventis, Apontis, Abbott outside this work. CM has participated in advisory boards for Boehringer Ingelheim and Novo Nordisk. RBS has received lecture fees and advisory board fees from BMS/Pfizer and Bayer outside this work. LF received institutional research grants by EU 633196 [CATCH ME] and EU 965286 [MAESTRIA]. British Heart Foundation (AA/18/2/34218), German Center for Cardiovascular Research supported by the German Ministry of Education and Research (DZHK) and several biomedical companies active in the field of research. LF is listed as inventor on two issued patents held by the employing institution (Atrial Fibrillation Therapy WO 2015140571, Markers for Atrial Fibrillation WO 2016012783). PK received research support for basic, translational, and clinical research projects from European Union, British Heart Foundation, Leducq Foundation, Medical Research Council (UK), and German Center for Cardiovascular Research, from several drug and device companies active in atrial fibrillation, and has received honoraria from several such companies in the past, but not in the last three years. PK is listed as inventor on two issued patents held by the institution (Atrial Fibrillation Therapy WO 2015140571, Markers for Atrial Fibrillation WO 2016012783).

References

1. Fabritz L, Guasch E, Antoniades C, Bardinet I, Benninger G, Betts TR, Brand E, Breithardt G, Bucklar-Suchankova G, Camm AJ, Cartlidge D, Casadei B, Chua WW, Crijns HJ, Deeks J, Hatem S, Hidden-Lucet F, Kaab S, Maniadakis N, Martin S, Mont L, Reinecke H, Sinner MF, Schotten U, Southwood T, Stoll M, Vardas P, Wakili R, West A, Ziegler A, Kirchhof P. Expert consensus document: Defining the major health modifiers causing atrial fibrillation: a roadmap to underpin personalized prevention and treatment. *Nat Rev Cardiol* 2016;**13**:230-237.
2. Nielsen JB, Thorolfsdottir RB, Fritsche LG, Zhou W, Skov MW, Graham SE, Herron TJ, McCarthy S, Schmidt EM, Sveinbjornsson G, Surakka I, Mathis MR, Yamazaki M, Crawford RD, Gabrielsen ME, Skogholt AH, Holmen OL, Lin M, Wolford BN, Dey R, Dalen H, Sulem P, Chung JH, Backman JD, Arnar DO, Thorsteinsdottir U, Baras A, O'Dushlaine C, Holst AG, Wen X, Hornsby W, Dewey FE, Boehnke M, Kheterpal S, Mukherjee B, Lee S, Kang HM, Holm H, Kitzman J, Shavit JA, Jalife J, Brummett CM, Teslovich TM, Carey DJ, Gudbjartsson DF, Stefansson K, Abecasis GR, Hveem K, Willer CJ. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet* 2018;**50**:1234-1239.
3. Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD, Newton-Cheh C, Lubitz SA, Magnani JW, Ellinor PT, Seshadri S, Wolf PA, Vasan RS, Benjamin EJ, Levy D. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. *Lancet* 2015;**386**:154-162.
4. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomstrom-Lundqvist C, Boriani G, Castella M, Dan GA, Dilaveris PE, Fauchier L, Filippatos G, Kalman JM, La Meir M, Lane DA, Lebeau JP, Lettino M, Lip GYH, Pinto FJ, Thomas GN, Valgimigli M, Van Gelder IC, Van Putte BP, Watkins CL, Group ESCSD. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J* 2021;**42**:373-498.
5. Kany S, Cardoso VR, Bravo L, Williams JA, Schnabel R, Fabritz L, Gkoutos GV, Kirchhof P. Eligibility for early rhythm control in patients with atrial fibrillation in the UK Biobank. *Heart* 2022;**108**:1873-1880.
6. Billing AM, Kim YC, Gullaksen S, Schrage B, Raabe J, Hutzfeldt A, Demir F, Kovalenko E, Lasse M, Dugourd A, Fallegger R, Klampe B, Jaegers J, Li Q, Kravtsova O, Crespo-Masip M, Palermo A, Fenton RA, Hoxha E, Blankenberg S, Kirchhof P, Huber TB, Laugesen E, Zeller T, Chrysopolou M, Saez-Rodriguez J, Magnussen C, Eschenhagen T, Staruschenko A, Siuzdak G, Poulsen PL, Schwab C, Cuello F, Vallon V, Rinschen MM. Metabolic Communication by SGLT2 Inhibition. *Circulation* 2023.
7. Hu YF, Chen YJ, Lin YJ, Chen SA. Inflammation and the pathogenesis of atrial fibrillation. *Nat Rev Cardiol* 2015;**12**:230-243.
8. Spronk HM, De Jong AM, Verheule S, De Boer HC, Maass AH, Lau DH, Rienstra M, van Hunnik A, Kuiper M, Lumeij S, Zeemering S, Linz D, Kamphuisen PW, Ten Cate H, Crijns HJ, Van Gelder IC, van Zonneveld AJ, Schotten U. Hypercoagulability causes atrial fibrosis and promotes atrial fibrillation. *Eur Heart J* 2017;**38**:38-50.
9. Reyat JS, Chua W, Cardoso VR, Witten A, Kastner PM, Kabir SN, Sinner MF, Wesselink R, Holmes AP, Pavlovic D, Stoll M, Kaab S, Gkoutos GV, de Groot JR, Kirchhof P, Fabritz

L. Reduced left atrial cardiomyocyte PITX2 and elevated circulating BMP10 predict atrial fibrillation after ablation. *JCI Insight* 2020;**5**.

10. Zeller T, Tunstall-Pedoe H, Saarela O, Ojeda F, Schnabel RB, Tuovinen T, Woodward M, Struthers A, Hughes M, Kee F, Salomaa V, Kuulasmaa K, Blankenberg S, Investigators M. High population prevalence of cardiac troponin I measured by a high-sensitivity assay and cardiovascular risk estimation: the MORGAM Biomarker Project Scottish Cohort. *Eur Heart J* 2014;**35**:271-281.
11. Pol T, Hijazi Z, Lindback J, Oldgren J, Alexander JH, Connolly SJ, Eikelboom JW, Ezekowitz MD, Granger CB, Lopes RD, Yusuf S, Siegbahn A, Wallentin L. Using multimarker screening to identify biomarkers associated with cardiovascular death in patients with atrial fibrillation. *Cardiovasc Res* 2022;**118**:2112-2123.
12. Hijazi Z, Lindback J, Alexander JH, Hanna M, Held C, Hylek EM, Lopes RD, Oldgren J, Siegbahn A, Stewart RA, White HD, Granger CB, Wallentin L, Aristotle, Investigators S. The ABC (age, biomarkers, clinical history) stroke risk score: a biomarker-based risk score for predicting stroke in atrial fibrillation. *Eur Heart J* 2016;**37**:1582-1590.
13. Hijazi Z, Oldgren J, Lindback J, Alexander JH, Connolly SJ, Eikelboom JW, Ezekowitz MD, Held C, Hylek EM, Lopes RD, Siegbahn A, Yusuf S, Granger CB, Wallentin L, Aristotle, Investigators R-L. The novel biomarker-based ABC (age, biomarkers, clinical history)-bleeding risk score for patients with atrial fibrillation: a derivation and validation study. *Lancet* 2016;**387**:2302-2311.
14. Kirchhof P, Camm AJ, Goette A, Brandes A, Eckardt L, Elvan A, Fetsch T, van Gelder IC, Haase D, Haegeli LM, Hamann F, Heidbuchel H, Hindricks G, Kautzner J, Kuck KH, Mont L, Ng GA, Rekosz J, Schoen N, Schotten U, Suling A, Taggeselle J, Themistoclakis S, Vettorazzi E, Vardas P, Wegscheider K, Willems S, Crijns H, Breithardt G, Investigators E-AT. Early Rhythm-Control Therapy in Patients with Atrial Fibrillation. *N Engl J Med* 2020;**383**:1305-1316.
15. Chua W, Purmah Y, Cardoso VR, Gkoutos GV, Tull SP, Neculau G, Thomas MR, Kotecha D, Lip GYH, Kirchhof P, Fabritz L. Data-driven discovery and validation of circulating blood-based biomarkers associated with prevalent atrial fibrillation. *Eur Heart J* 2019;**40**:1268-1276.
16. Kany S, Al-Taie C, Roselli C, Pirruccello JP, Borof K, Reinbold C, Suling A, Krause L, Reissmann B, Schnabel RB, Zeller T, Zapf A, Wegscheider K, Fabritz L, Ellinor PT, Kirchhof P. Association of genetic risk and outcomes in patients with atrial fibrillation: interactions with early rhythm control in the EAST-AFNET4 trial. *Cardiovasc Res* 2023;**119**:1799-1810.
17. Kany S, Al-Taie C, Roselli C, Pirruccello JP, Borof K, Reinbold C, Suling A, Krause L, Reissmann B, Schnabel RB, Zeller T, Zapf A, Wegscheider K, Fabritz L, Ellinor PT, Kirchhof P. Association of genetic risk and outcomes in patients with atrial fibrillation: interactions with early rhythm control in the EAST-AFNET4 trial. *Cardiovasc Res* 2023.
18. Brady PF, Chua W, Nehaj F, Connolly DL, Khashaba A, Purmah YJV, Ul-Qamar MJ, Thomas MR, Varma C, Schnabel RB, Zeller T, Fabritz L, Kirchhof PF. Interactions Between Atrial Fibrillation and Natriuretic Peptide in Predicting Heart Failure Hospitalization or Cardiovascular Death. *J Am Heart Assoc* 2022;**11**:e022833.
19. BLOM G. 'Transformations of the binomial, negative binomial, Poisson and χ2 distributions'. *Biometrika* 1956;**43**:235-235.
20. Linzer DA, Lewis JB. poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software* 2011;**42**:1 - 29.
21. Lloyd S. Least squares quantization in pcm. technical note. *Bell Laboratories* 1957.

22. Arthur D, Vassilvitskii S. K-Means++: The Advantages of Careful Seeding. 2007.
23. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;**150**:604-612.
24. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, Crews DC, Doria A, Estrella MM, Froissart M, Grams ME, Greene T, Grubb A, Gudnason V, Gutiérrez OM, Kalil R, Karger AB, Mauer M, Navis G, Nelson RG, Poggio ED, Rodby R, Rossing P, Rule AD, Selvin E, Seegmiller JC, Shlipak MG, Torres VE, Yang W, Ballew SH, Couture SJ, Powe NR, Levey AS. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. *N Engl J Med* 2021;**385**:1737-1749.
25. Azen R, Budescu DV. The dominance analysis approach for comparing predictors in multiple regression. *Psychol Methods* 2003;**8**:129-148.
26. Shekhar S. Dominance Analysis: A Python Library for Accurate and Intuitive Relative Importance of Predictors. In: Sajan Bhagat KSaBKK, ed., 2023.
27. Davidson-Pilon C. lifelines: survival analysis in Python. *Journal of Open Source Software* 2019;**4**:1317.
28. Wang L, Rice M, Swist S, Kubin T, Wu F, Wang S, Kraut S, Weissmann N, Bottger T, Wheeler M, Schneider A, Braun T. BMP9 and BMP10 Act Directly on Vascular Smooth Muscle Cells for Generation and Maintenance of the Contractile State. *Circulation* 2021;**143**:1394-1410.
29. Ko T, Nomura S, Yamada S, Fujita K, Fujita T, Satoh M, Oka C, Katoh M, Ito M, Katagiri M, Sassa T, Zhang B, Hatsuse S, Yamada T, Harada M, Toko H, Amiya E, Hatano M, Kinoshita O, Nawata K, Abe H, Ushiku T, Ono M, Ikeuchi M, Morita H, Aburatani H, Komuro I. Cardiac fibroblasts regulate the development of heart failure via Htra3-TGF-beta-IGFBP7 axis. *Nat Commun* 2022;**13**:3275.
30. Vaduganathan M, Sattar N, Xu J, Butler J, Mahaffey KW, Neal B, Shaw W, Rosenthal N, Pfeifer M, Hansen MK, Januzzi JL, Jr. Stress Cardiac Biomarkers, Cardiovascular and Renal Outcomes, and Response to Canagliflozin. *J Am Coll Cardiol* 2022;**79**:432-444.
31. Fiedler U, Scharpfenecker M, Koidl S, Hegen A, Grunow V, Schmidt JM, Kriz W, Thurston G, Augustin HG. The Tie-2 ligand angiopoietin-2 is stored in and rapidly released upon stimulation from endothelial cell Weibel-Palade bodies. *Blood* 2004;**103**:4150-4156.
32. Ruff CT, Giugliano RP, Braunwald E, Hoffman EB, Deenadayalu N, Ezekowitz MD, Camm AJ, Weitz JI, Lewis BS, Parkhomenko A, Yamashita T, Antman EM. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. *Lancet* 2014;**383**:955-962.
33. Sohns C, Fox H, Marrouche NF, Crijns H, Costard-Jaeckle A, Bergau L, Hindricks G, Dagres N, Sossalla S, Schramm R, Fink T, El Hamriti M, Moersdorf M, Sciacca V, Konietschke F, Rudolph V, Gummert J, Tijssen JGP, Sommer P, Investigators CH. Catheter Ablation in End-Stage Heart Failure with Atrial Fibrillation. *N Engl J Med* 2023;**389**:1380-1389.
34. Whitlock RP, Belley-Cote EP, Paparella D, Healey JS, Brady K, Sharma M, Reents W, Budera P, Baddour AJ, Fila P, Devereaux PJ, Bogachev-Prokophiev A, Boening A, Teoh KHT, Tagarakis GI, Slaughter MS, Royse AG, McGuinness S, Alings M, Punjabi PP, Mazer CD, Folkeringa RJ, Colli A, Avezum A, Nakamya J, Balasubramanian K, Vincent J, Voisine P, Lamy A, Yusuf S, Connolly SJ, Investigators LI. Left Atrial Appendage Occlusion during Cardiac Surgery to Prevent Stroke. *N Engl J Med* 2021;**384**:2081-2091.

35. Zhang HD, Ding L, Mi LJ, Zhang AK, Zhang K, Jiang ZH, Yu FY, Yan XX, Shen YJ, Tang M. SGLT2 inhibitors for the prevention of atrial fibrillation: a systemic review and meta-analysis. *Eur J Prev Cardiol* 2023.

36. Litvinukova M, Talavera-Lopez C, Maatz H, Reichart D, Worth CL, Lindberg EL, Kanda M, Polanski K, Heinig M, Lee M, Nadelmann ER, Roberts K, Tuck L, Fasouli ES, DeLaughter DM, McDonough B, Wakimoto H, Gorham JM, Samari S, Mahbubani KT, Saeb-Parsy K, Patone G, Boyle JJ, Zhang H, Zhang H, Viveiros A, Oudit GY, Bayraktar OA, Seidman JG, Seidman CE, Noseda M, Hubner N, Teichmann SA. Cells of the adult human heart. *Nature* 2020;**588**:466-472.

37. Hennings E, Blum S, Aeschbacher S, Coslovsky M, Knecht S, Eken C, Lischer M, Paladini RE, Krisai P, Reichlin T, Rodondi N, Beer JH, Ammann P, Conte G, De Perna ML, Kobza R, Blum MR, Bossard M, Kastner P, Ziegler A, Muller C, Bonati LH, Pfister O, Zuern CS, Conen D, Kuhne M, Osswald S, Swiss AFI. Bone Morphogenetic Protein 10-A Novel Biomarker to Predict Adverse Outcomes in Patients With Atrial Fibrillation. *J Am Heart Assoc* 2023;**12**:e028255.

38. Hijazi Z, Benz AP, Lindback J, Alexander JH, Connolly SJ, Eikelboom JW, Granger CB, Kastner P, Lopes RD, Ziegler A, Oldgren J, Siegbahn A, Wallentin L. Bone morphogenetic protein 10: a novel risk marker of ischaemic stroke in patients with atrial fibrillation. *Eur Heart J* 2023;**44**:208-218.

39. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, Fonseca F, Nicolau J, Koenig W, Anker SD, Kastelein JJP, Cornel JH, Pais P, Pella D, Genest J, Cifkova R, Lorenzatti A, Forster T, Kobalava Z, Vida-Simiti L, Flather M, Shimokawa H, Ogawa H, Dellborg M, Rossi PRF, Troquay RPT, Libby P, Glynn RJ, Group CT. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med* 2017;**377**:1119-1131.

40. Nidorf SM, Fiolet ATL, Mosterd A, Eikelboom JW, Schut A, Opstal TSJ, The SHK, Xu XF, Ireland MA, Lenderink T, Latchem D, Hoogslag P, Jerzewski A, Nierop P, Whelan A, Hendriks R, Swart H, Schaap J, Kuijper AFM, van Hessen MWJ, Saklani P, Tan I, Thompson AG, Morton A, Judkins C, Bax WA, Dirksen M, Alings M, Hankey GJ, Budgeon CA, Tijssen JGP, Cornel JH, Thompson PL, LoDoCo2 Trial I. Colchicine in Patients with Chronic Coronary Disease. *N Engl J Med* 2020;**383**:1838-1847.

41. Tardif JC, Kouz S, Waters DD, Bertrand OF, Diaz R, Maggioni AP, Pinto FJ, Ibrahim R, Gamra H, Kiwan GS, Berry C, Lopez-Sendon J, Ostadal P, Koenig W, Angoulvant D, Gregoire JC, Lavoie MA, Dube MP, Rhainds D, Provencher M, Blondeau L, Orfanos A, L'Allier PL, Guertin MC, Roubille F. Efficacy and Safety of Low-Dose Colchicine after Myocardial Infarction. *N Engl J Med* 2019;**381**:2497-2505.

42. Conen D, Ke Wang M, Popova E, Chan MTV, Landoni G, Cata JP, Reimer C, McLean SR, Srinathan SK, Reyes JCT, Grande AM, Tallada AG, Sessler DI, Fleischmann E, Kabon B, Voltolini L, Cruz P, Maziak DE, Gutierrez-Soriano L, McIntyre WF, Tandon V, Martinez-Tellez E, Guerra-Londono JJ, DuMerton D, Wong RHL, McGuire AL, Kidane B, Roux DP, Shargall Y, Wells JR, Ofori SN, Vincent J, Xu L, Li Z, Eikelboom JW, Jolly SS, Healey JS, Devereaux PJ, Investigators C-A. Effect of colchicine on perioperative atrial fibrillation and myocardial injury after non-cardiac surgery in patients undergoing major thoracic surgery (COP-AF): an international randomised trial. *Lancet* 2023;**402**:1627-1635.

43. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet* 2015;**386**:533-540.

44. Morrow DA, Cannon CP, Rifai N, Frey MJ, Vicari R, Lakkis N, Robertson DH, Hille DA, DeLucca PT, DiBattiste PM, Demopoulos LA, Weintraub WS, Braunwald E, Investigators T-T. Ability of minor elevations of troponins I and T to predict benefit from an early invasive strategy in patients with unstable angina and non-ST elevation myocardial infarction: results from a randomized trial. *JAMA* 2001;**286**:2405-2412.

45. Mueller C, Scholer A, Laule-Kilian K, Martina B, Schindler C, Buser P, Pfisterer M, Perruchoud AP. Use of B-type natriuretic peptide in the evaluation and management of acute dyspnea. *N Engl J Med* 2004;**350**:647-654.

46. Moqri M, Herzog C, Poganik JR, Ying K, Justice JN, Belsky DW, Higgins-Chen AT, Chen BH, Cohen AA, Fuellen G, Hagg S, Marioni RE, Widschwendter M, Fortney K, Fedichev PO, Zhavoronkov A, Barzilai N, Lasky-Su J, Kiel DP, Kennedy BK, Cummings S, Slagboom PE, Verdin E, Maier AB, Sebastiano V, Snyder MP, Gladyshev VN, Horvath S, Ferrucci L. Validation of biomarkers of aging. *Nat Med* 2024;**30**:360-372.

47. Winters J, Isaacs A, Zeemering S, Kawczynski M, Maesen B, Maessen J, Bidar E, Boukens B, Hermans B, van Hunnik A, Casadei B, Fabritz L, Chua W, Sommerfeld L, Guasch E, Mont L, Batlle M, Hatem S, Kirchhof P, Wakili R, Sinner M, Stoll M, Goette A, Verheule S, Schotten U. Heart Failure, Female Sex, and Atrial Fibrillation Are the Main Drivers of Human Atrial Cardiomyopathy: Results From the CATCH ME Consortium. *J Am Heart Assoc* 2023;**12**:e031220.

48. Barallobre-Barreiro J, Gupta SK, Zoccarato A, Kitazume-Taneike R, Fava M, Yin X, Werner T, Hirt MN, Zampetaki A, Viviano A, Chong M, Bern M, Kourliouros A, Domenech N, Willeit P, Shah AM, Jahangiri M, Schaefer L, Fischer JW, Iozzo RV, Viner R, Thum T, Heineke J, Kichler A, Otsu K, Mayr M. Glycoproteomics Reveals Decorin Peptides With Anti-Myostatin Activity in Human Atrial Fibrillation. *Circulation* 2016;**134**:817-832.

49. Ko D, Benson MD, Ngo D, Yang Q, Larson MG, Wang TJ, Trinquart L, McManus DD, Lubitz SA, Ellinor PT, Vasan RS, Gerszten RE, Benjamin EJ, Lin H. Proteomics Profiling and Risk of New-Onset Atrial Fibrillation: Framingham Heart Study. *J Am Heart Assoc* 2019;**8**:e010976.

Tables

Table 1: Disease mechanisms hypothesized to be related to atrial fibrillation and atrial fibrillation-related complications and corresponding circulating biomolecules selected for quantification in this study. The list of candidates for disease mechanisms ("health modifiers") is copied from Box 2 in [1] with one added mechanism (systemic inflammation, last line). The biomolecules selected as quantifiable proxies of each disease process this study are shown in the right column. The selection process is detailed in the methods of this paper and in [1]. This table aligns biomolecules with selected key processes. The assignment of biomolecules to one mechanism is an oversimplification as the concentrations of biomolecules will be influenced by several disease mechanisms (see Figure 1 in [1])

Disease process ("health modifier")	**Biomolecule selected as quantifiable proxy of this disease process**
Ageing	growth differentiation factor 15 (GDF15) cancer antigen 125 (CA125)
Loss of cardiomyocytes	cardiac troponin (TnT)
Replacement of cardiomyocytes with extracellular matrix	fibroblast growth factor 23 (FGF23) insulin-like growth factor binding protein 7 (IGFBP7)
Adaptive changes to increased work load	N-terminal pro–B-type natriuretic peptide (NT-proBNP)
Delayed left atrial activation	see two rows above: replacement of cardiomyocytes with extracellular matrix
Spontaneous electrical activity	possibly bone morphogenetic protein 10 (BMP10)
Genetic predisposition to AF	tested using genomic analysis (see supplement and [16])
Infiltration of fat cells in the atria and activation of atrial fat tissue	fatty acid binding protein 3 (FABP3)
Elevated atrial oxidative stress	bone morphogenetic protein 10 (BMP10)
Renal dysfunction	creatinine fibroblast growth factor 23 (FGF23)
Prothrombotic dysregulation - Endothelial	Angiopoietin 2 (ANGPT2) endothelial specific molecule 1 (ESM1)
Prothrombotic dysregulation - humoral	D-dimer
Additional processes: Systemic inflammation	interleukin-6 (IL-6) C-reactive protein (CRP)

Table 2: Distribution of patient characteristics in the four biomolecule-derived patient clusters in the EAST-AFNET 4 biomolecule data set. Continuous and discrete numeric parameters are shown as median (interquartile range), nominal features as number of patients (%). Chronic kidney disease was classified based on estimated creatinine clearance calculated using the CKD-Epi formula.

Characteristic	Cluster in EAST AFNET4 by Po-LCA				p-value
	Blue cluster N = 502 (32%)	Green cluster N = 512 (32%)	Orange cluster N = 302 (19%)	Red cluster N = 270 (17%)	
Randomised to early rhythm control	257 (51.2)	263 (51.4)	142 (47.0)	138 (51.1)	0.8
CHA_2DS_2VASc score	3.0 [2.0, 3.0]	3.0 [2.0, 4.0]	3.0 [3.0, 4.0]	4.0 [3.0, 5.0]	<0.001
Female sex	233 (46%)	238 (46%)	119 (39%)	123 (46%)	0.2
Body mass index (BMI)	28.6 (25.6, 31.6)	28.0 (25.2, 31.4)	30.0 (26.7, 33.4)	29.1 (25.7, 33.3)	0.026
Obese, defined as BMI ≥ 30	190 (38%)	177 (35%)	151 (50%)	114 (42%)	0.026
Arterial hypertension	435 (87%)	446 (87%)	270 (89%)	249 (92%)	<0.001
Diabetes mellitus	98 (20%)	112 (22%)	100 (33%)	86 (32%)	0.7
Stable heart failure NYHA stage II-IV or left ventricular ejection fraction <50%	108 (22%)	134 (26%)	86 (28%)	147 (54%)	<0.001
Prior stroke or TIA	53 (11%)	63 (12%)	46 (15%)	33 (12%)	0.8
History of myocardial infarction or revascularisation by stenting or bypass surgery	58 (12%)	65 (13%)	68 (23%)	61 (23%)	0.8
Chronic kidney disease	38 (7.6)	99 (19.3)	73 (24.2)	118 (43.7)	<0.001
Chronic obstructive lung disease	27 (5.4%)	37 (7.2%)	29 (9.6%)	31 (11%)	0.3
Peripheral artery disease	12 (2.4%)	19 (3.7%)	18 (6.0%)	21 (7.8%)	0.3

Table 3: Biomolecule concentrations in the clusters. All biomolecule concentrations are given as median and interquartile range. Note that some of the most relevant biomolecules by explained variance show a relatively small range of concentrations, e.g. BMP10 and IGFBP7 compared to other biomolecules with known predictive effects in patients with cardiovascular diseases and a high range of values, e.g. TnT or NT-pro-BNP.

	Blue cluster N = 502 (32%)	**Green cluster** N = 512 (32%)	**Orange cluster** N = 302 (19%)	**Red cluster** N = 270 (17%)	**Overall range [min. – max.]**
IL-6 (pg/ml)	1.8 [1.5 - 2.5]	2.1 [1.5 - 2.9]	4.6 [3.6 - 7.1]	4.7 [3.2 - 7.7]	[1.50 – 38.83]
NT-proBNP (pg/ml)	154.8 [84.3 - 303.7]	560.2 [302.0 - 1024.5]	461.3 [223.2 - 844.5]	1527.5 [919.8 - 541.5]	[27.49 – 5409]
TnT (ng/l)	8.0 [6.3 - 10.4]	11.0 [8.6 - 14.5]	15.1 [11.1 - 22.6]	19.1 [13.9 - 29.5]	[3.53 – 79.27]
GDF15 (pg/ml)	937.0 [760.6 - 226.8]	1295.5 [1006.0 - 699.8]	1716.5 [1329.0 - 333.5]	2499.0 [1910.2 - 409.5]	[507.7 – 8007]
CRP (mg/l)	1.4 [0.6 - 3.0]	1.6 [0.7 - 2.7]	4.7 [2.5 - 8.9]	4.5 [2.1 - 11.7]	[0.08 – 88.63]
D-dimer (µg/ml)	0.1 [0.1 - 0.2]	0.1 [0.1 - 0.2]	0.3 [0.1 - 0.5]	0.3 [0.2 - 0.7]	[0 – 3.54]
CA125 (U/ml)	9.9 [7.3 - 12.9]	11.3 [8.0 - 14.8]	11.8 [8.2 - 17.1]	15.8 [11.0 - 25.5]	[3.46 – 97.98]
ANGPT2 (ng/ml)	1.8 [1.5 - 2.3]	2.8 [2.1 - 3.9]	2.6 [2.0 - 3.4]	4.9 [3.3 - 7.5]	[0.95 – 12.58]
BMP10 (ng/ml)	1.9 [1.7 - 2.1]	2.3 [2.0 - 2.5]	2.0 [1.7 - 2.2]	2.8 [2.4 - 3.1]	[1.30 – 3.89]
ESM1 (ng/ml)	1.8 [1.5 - 2.2]	2.1 [1.8 - 2.5]	2.0 [1.5 - 2.8]	2.6 [2.1 - 4.0]	[0.98 – 10.58]
FABP3 (ng/ml)	26.4 [23.0 - 31.1]	32.2 [27.2 - 37.3]	35.2 [30.6 - 45.4]	45.4 [34.9 - 55.4]	[15.52 – 92.53]
FGF23 (pg/ml)	116.0 [97.3 - 150.0]	153.6 [126.5 - 192.0]	176.4 [129.8 - 236.5]	254.4 [189.4 - 400.6]	[68.11 – 1352.70]
IGFBP7 (ng/ml)	87.8 [81.3 - 95.0]	106.1 [98.4 - 114.6]	103.6 [92.7 - 114.7]	135.0 [121.6 - 157.4]	[68.52 – 208.46]

Table 4A: Efficacy outcomes in the EAST-AFNET 4 biomolecule data set for each biomolecule cluster. Given are patients with event per observation-years and the annualized event rate in percent (in brackets).

EAST-AFNET 4 (Derivation)				
	Low risk (blue) cluster N = 502 (32%)	**Low-intermediate risk (green) cluster** N = 512 (32%)	**High-intermediate risk (orange) cluster** N = 302 (19%)	**High risk (red) cluster** N = 270 (17%)
Stroke	10/2587 (0.4)	16/2634 (0.6)	11/1450 (0.8)	14/1160 (1.2)
Cardiovascular death	10/2624 (0.4)	17/2667 (0.6)	21/1474 (1.4)	42/1190 (3.5)
Unplanned heart failure hospitalization	20/2585 (0.8)	37/2598 (1.4)	43/1367 (3.1)	65/1024 (6.3)
Unplanned hospitalization for acute coronary syndrome	17/2581 (0.7)	14/2627 (0.5)	18/1429 (1.3)	14/1151 (1.2)
BBC-AF (Validation)				
	Low risk (blue) cluster N = 268 (36%)	**Low-intermediate risk (green) cluster** N = 185 (25%)	**High-intermediate risk (orange) cluster** N = 123 (16%)	**High risk (red) cluster** N = 172 (23%)
Stroke	2/843 (0.2)	5/516 (1.0)	5/288 (1.7)	7/360 (1.9)
Cardiovascular death	5/814 (0.6)	19/509 (3.7)	30/292 (10.3)	72/366 (19.7)
Unplanned heart failure hospitalization	15/826 (1.8)	32/469 (6.8)	29/248 (11.7)	56/269 (20.8)
Unplanned hospitalization for acute coronary syndrome	0/870 (0.0)	3/546 (0.5)	0/297 (0.0)	0/378 (0.0)

Table 4B. Safety outcomes in the EAST-AFNET 4 biomolecule data set and in the BBC-AF data set. Numbers show patients with events and annualized event rates in percent. The safety outcome component "major adverse events related to rhythm control therapy" was not exactly defined in BBC-AF. Therefore, the clinically relevant outcome "major bleeding" was used.

EAST-AFNET 4 (Derivation)				
	Low risk (blue) cluster N = 502 (32%)	**Low-intermediate risk (green) cluster** N = 512 (32%)	**High-intermediate risk (orange) cluster** N = 302 (19%)	**High risk (red) cluster** N = 270 (17%)
Death	18/2624 (0.7)	32/2667 (1.2)	44/1474 (3.0)	64/1190 (5.4)
Stroke	10/2587 (0.4)	16/2634 (0.6)	11/1450 (0.8)	14/1160 (1.2)
Major adverse events related to rhythm control	41/2583 (1.6)	61/2597 (2.3)	62/1436 (4.3)	81/1150 (7.0)
BBC-AF (Validation)				
	Low risk (blue) cluster N = 268 (36%)	**Low-intermediate risk (green) cluster** N = 185 (25%)	**High-intermediate risk (orange) cluster** N = 123 (16%)	**High risk (red) cluster** N = 172 (23%)
Death	23/814 (2.8)	39/509 (7.7)	55/292 (18.8)	104/366 (28.4)
Stroke	2/843 (0.2)	5/516 (1.0)	5/288 (1.7)	7/360 (1.9)
Major bleeding	6/1098 (0.5)	18/744 (2.4)	17/464 (3.7)	22/680 (3.2)

	Low risk (blue) cluster / reference group	Low-intermediate risk (green) cluster / other risk group	High-intermediate risk (orange) cluster / other risk group	High risk (red) cluster / other risk group
EAST-AFNET 4 (derivation data set)				
Biomolecule clusters	1 (reference)	1.3 [0.9, 1.9]	2.7 [1.9, 3.6]	5.2 [3.7, 7.2]
CHA_2DS_2VASc	1 (reference)	1.5 [1.0 – 2.2]	2.4 [1.6 – 3.34]	3.8 [2.8 – 5.3]
ABC-Stroke	No events	1 (reference)	2.7 [2.2 – 3.4]	4.7 [3.1 – 6.9]
ABC Bleeding	1 (reference)	1.9 [1.1 – 3.3]		4.8 [2.7 – 8.3]
NT-proBNP (quartiles)	1 (reference)	1.5 [1.2 - 2.1]	2.1 [1.5 – 2.8]	4.7 [3.2 – 6.9]
TnT (discretized)	1 (reference)	1.2 [0.9 - 1.7]	2.3 [1.7 - 3.1]	4.4 [3.1 - 6.2]
PRS AF	1 (reference)	0.9 [0.7 - 1.3]		0.9 [0.7 – 1.3]
PRS Stroke	1 (reference)	1.1 [0.8 – 1.4]		1.3 [0.9 – 1.9]
BBC-AF (validation data set)				
Biomolecule clusters BBC-AF (validation)	1 (reference)	4.0 [2.3 – 7.0]	8.3 [4.80 – 14.4]	14.1 [8.4 – 23.7]
CHA_2DS_2VASc	1 (reference)	1.9 [1.2 – 3.0]	1.6 [0.9 – 2.5]	2.3 [1.5 – 3.4]
ABC-Stroke	No events	1 (reference)	4.3 [2.6 – 6.9]	8.2 [5.0 – 13.4]
ABC Bleeding	1 (reference)	3.4 [2.4 – 4.7]		4.8 [3.4 – 6.9]
NT-proBNP (quartiles)	1 (reference)	2.3 [1.2 – 4.4]	6.3 [3.4 – 11.7]	10.2 [5.7 – 18.3]
TnT (discretized)	1 (reference)	2.7 [1.5 – 4.8]	5.5 [3.1 – 9.6]	7.2 [4.2 – 12.5]

Figure legends

Figure 1.

CONSORT flow chart of the derivation data set, the EAST-AFNET 4 biomolecule study.

Figure 2.

Figure 2A: Unsupervised clustering based on concentrations of 13 biomolecules reflecting different cardiovascular disease mechanisms assigns patients to four clusters. Projection of patients into two-dimensional space spanned by first two principal components derived by applying principal component analysis on the 13 biomolecules in EAST-AFNET 4. Clustering was performed without any clinical information, relying on biomolecule concentrations only.

Figure 2B: Application of the assignment rules derived in the EAST-AFNET 4 trial assigns patients in the validation cohort to four clusters with similar frequencies. The validation cohort consisted of all patients with atrial fibrillation enrolled into the BBC-AF cohort, a prospective cohort study enrolling patients with cardiovascular conditions presenting to a large NHS teaching trust.

Figure 2C: Sankey plot comparing the partitioning of the EAST participants to cluster groups based on thirteen biomolecules resulting from K-means and Po-LCA clustering. Both methods create comparable clusters.

Figure 2D: List of the top five biomolecules explaining variance in each cluster. The highest-risk cluster was dominated by high concentrations of NT-pro-BNP, IGFBP7, BMP10, ANGPT2, and GDF15. The lowest-risk cluster showed low concentrations of these biomolecules. Elevated concentrations of IL-6, CRP, and low concentrations of D-dimer contribute additional information to the variance in the two intermediate-risk clusters.

Figure 3.

Figure 3A: Aalen-Johansen curves for cluster groups from poLCA clustering model for the first primary outcome in EAST-AFNET 4, a composite of all-cause mortality, stroke, or unplanned hospitalization for heart failure or acute coronary syndrome.

Figure 3B: Aalen-Johansen curves for cluster groups from poLCA clustering model for the first primary outcome in BBC-AF, a composite of all-cause mortality, stroke, or unplanned hospitalization for heart failure or acute coronary syndrome. Administrative censoring has been applied to events occurring after number of at-risk patients dropped below five.

Figure 3C: Aalen-Johansen curves for cluster groups from poLCA clustering model for the safety outcome in EAST-AFNET 4.

Figure 3D: Aalen-Johansen curves for cluster groups from poLCA clustering model for the safety outcome in BBC-AF. Administrative censoring has been applied to events occurring after number of at-risk patients dropped below five.

Figure 4.

Figure 4A: Plot of adjusted rand indices for each simulated clustering process when using less biomolecules. Each colour indicates one number of biomolecules (2-12) used for clustering. Each dot represents one simulated set of clusters. Adjusted rand indices are shown for each model in ascending order from right to left.

Figure 4B: Plot of the adjusted rand index, a measure of the assignment of a patient to a biomolecule-derived cluster, and of the corresponding c index, for each virtual model using POL-CA clustering with a reduced number of biomolecules (2 – 12). Each dot represents one clustering model. The colour indicates the number of biomolecules used. Models relying on 8 or less biomolecules (yellow, orange, and red colours) consistently yield adjusted rand indices below 0.55. Only models using 7-12 biomolecules achieve correct assignment of patients to biomolecule clusters. The c index, a summarized measure of the accuracy of risk prediction, changes only marginally (x axis).

Figure 4C: Importance of each biomolecule included in this study based on the effect of its removal from the clustering process. For each number of biomolecules (2-12), the five clusters with the lowest rand indices were selected. The missing biomolecules were listed and ranked by number of clusters lacking that biomolecule. This list provides an estimate of the importance of each biomolecule in the clustering process.

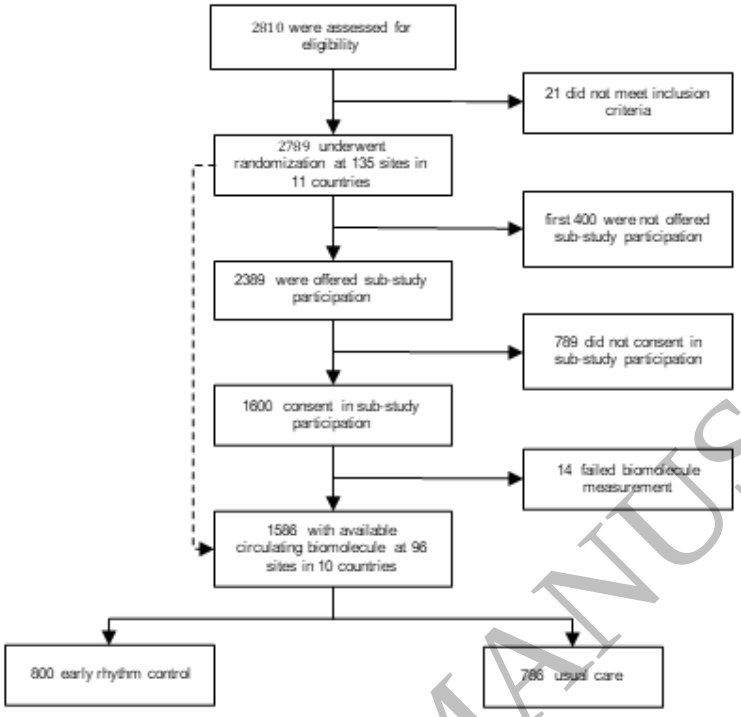
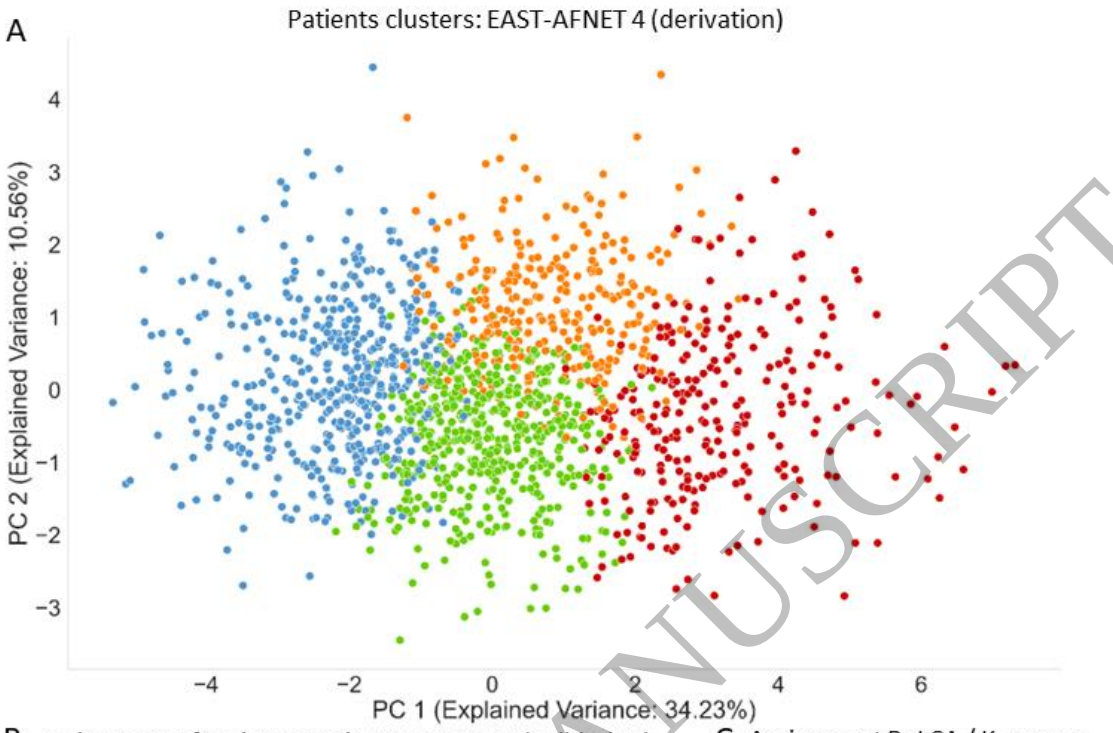


Figure 1

Figure 1
159x188 mm (x DPI)



D Biomolecules defining assignment to clusters

66% explained by biomolecules	85% explained by biomolecules	78% explained by biomolecules	72% explained by biomolecules
IGFBP7 ↓	IL-6 ↔	BMP10 ↔	**NT-proBNP ↑**
NT-proBNP ↓	CRP ↓	**IL-6 ↑**	**IGFBP7 ↑**
GDF15 ↓	D-dimer ↓	**CRP ↑**	**BMP10 ↑**
ANGPT2 ↓	TnT ↑	TnT ↗	**ANGPT2 ↑**
TnT ↓	GDF15 ↗	NT-proBNP ↗	**GDF15 ↑**

Figure 2

Figure 2
159x220 mm (x DPI)

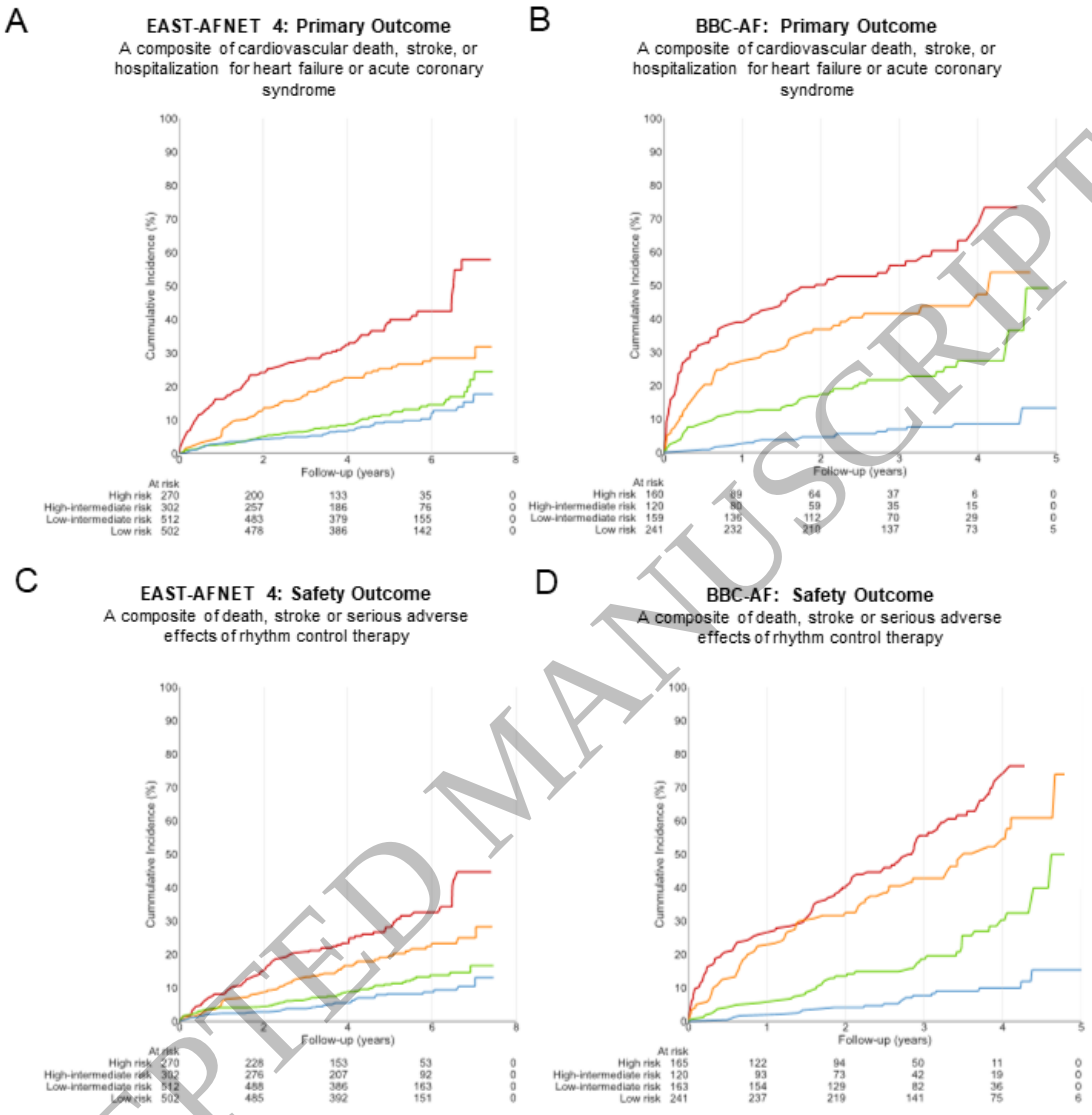


Figure 3

Figure 3
159x196 mm (x DPI)

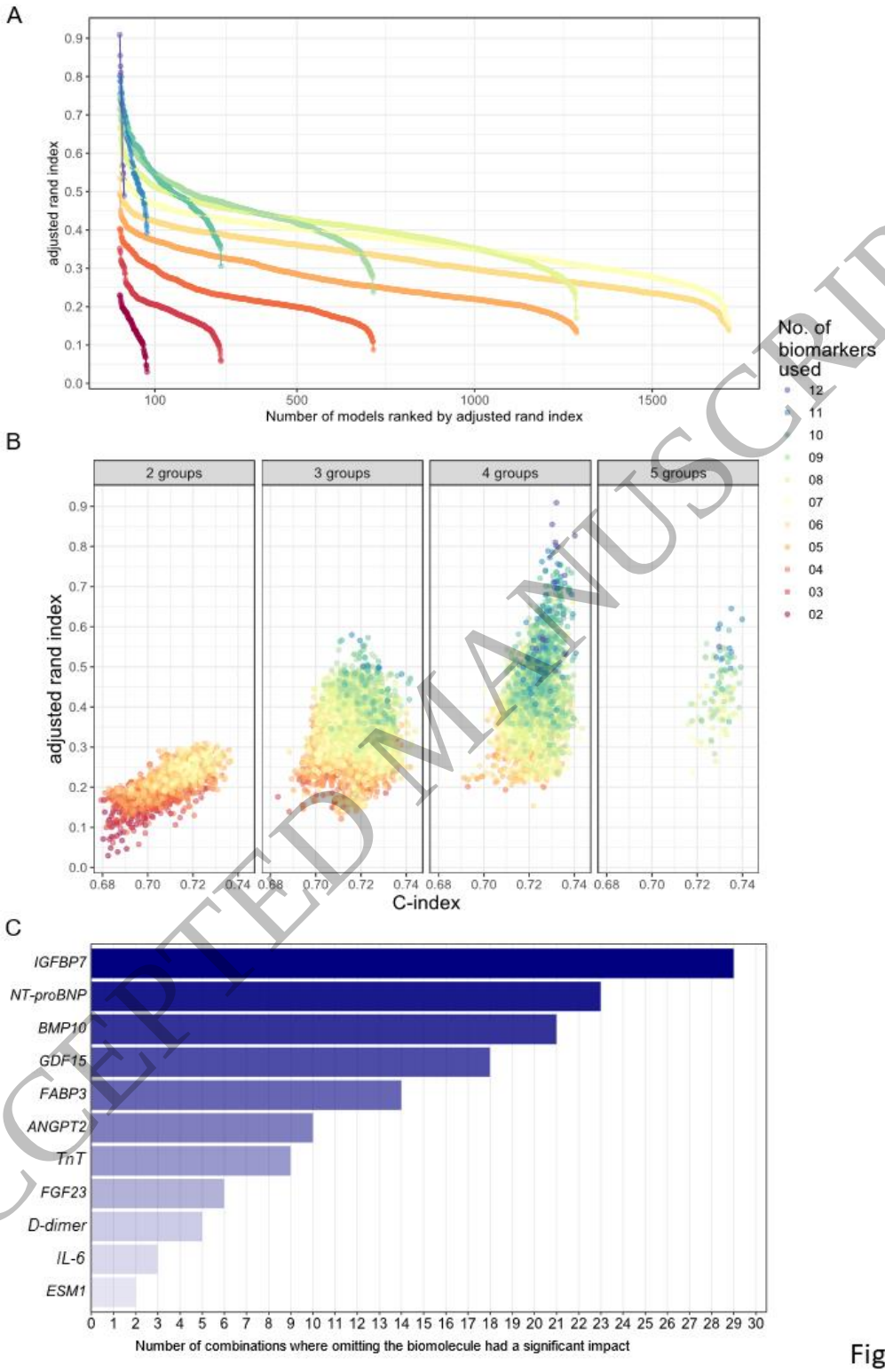


Figure 4

Figure 4
159x220 mm (x DPI)