Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction

Authors:

Rebecca J. Woolley, M.D.[1], Daan Ceelen[1], Wouter Ouwerkerk, Ph.D.[2,3], Jasper Tromp, M.D., Ph.D[1,2,4] , Sylwia M. Figarska, Ph.D.[1], Stefan D. Anker, M.D., Ph.D[5,6], Kenneth Dickstein M.D., Ph.D.[7,8], Gerasimos Filippatos, M.D., Ph.D.[9,10], Faiez Zannad, M.D., Ph.D.[11], Metra Marco, M.D.[12], Leong Ng M.D.[13], Nilesh Samani, M.D.[13], Dirk van Veldhuisen M.D., Ph.D.[1], Chim Lang, M.D.[14], Carolyn S. Lam M.D., Ph.D.[1,2,4], Adriaan Voors, M.D., Ph.D.[1].

Affiliations:

1. Department of Cardiology, University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands
2. National Heart Centre Singapore, Singapore.
3. Department of Dermatology and Netherlands Institute for Pigment Disorders, Amsterdam, University Medical Centers, University of Amsterdam, Amsterdam Infection & Immunity Institute, Cancer Center Amsterdam, Amsterdam, The Netherlands.
4. Duke-NUS Medical School, Singapore
5. Department of Cardiology (CVK) and Berlin Institute of Health Center for Regenerative Therapies (BCRT), Germany.
6. German Centre for Cardiovascular Research (DZHK) partner site Berlin, Charité Universitätsmedizin Berlin, Germany.
7. University of Bergen, Bergen, Norway.
8. Stavanger University Hospital, Stavanger, Norway.
9. National and Kapodistrian University of Athens, School of Medicine, Athens, Greece.
10. University of Cyprus, School of Medicine, Nicosia, Cyprus.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ejhf.2144

11. Universite de Lorraine, Inserm, Centre d'Investigations Cliniques-1433 and F-CRIN INI CRCT, Nancy, France (F.Z.).

12. Institute of Cardiology, ASST Spedali Civili di Brescia and Department of Medical and Surgical Specialties, Radiological Sciences, and Public Health, University of Brescia, Brescia, Italy.

13. Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, and NIHR Leicester Biomedical Research Centre, Leicester, UK.

14. School of Medicine Centre for Cardiovascular and Lung Biology, Division of Medical Sciences, University of Dundee, Ninewells Hospital & Medical School, Dundee, UK.

Address for Correspondence

Prof. Dr. A. A. Voors, Department of Cardiology, University Medical Center Groningen, Hanzeplein 1, 9713 GZ, Groningen, The Netherlands.

Email: a.a.voors@umcg.nl

Abstract:

Aims: The lack of effective therapies for patients with heart failure with preserved ejection fraction (HFpEF) is often ascribed to the heterogeneity of patients with HFpEF. We aimed to identify distinct pathophysiologic clusters of HFpEF based on circulating biomarkers.

Methods and Results: We performed an unsupervised cluster analysis using 363 biomarkers from 429 patient with HFpEF. Relative differences in expression profiles of the biomarkers between clusters were assessed and used for pathway over-representation analyses.

We identified four distinct patients subgroups based on their biomarker profiles : cluster 1 with the highest prevalence of diabetes mellitus and renal disease; cluster 2 with oldest age and frequent age-related comorbidities; cluster 3 with youngest age, largest body size, least symptoms and lowest NT-proBNP levels; and cluster 4 with highest prevalence of ischemic aetiology, smoking and chronic lung disease, most symptoms, as well as highest NT-proBNP and troponin levels. Over a median follow-up of 21 months, the occurrence of death or HF hospitalization was highest in clusters 1 and 4 (62.1% and 62.8% respectively) and lowest in cluster 3 (25.6%). Pathway over-representation analyses revealed that the biomarker profile of patients in cluster 1 was associated with activation of inflammatory pathways while the biomarker profile of patients in cluster 4 was specifically associated with pathways implicated in cell proliferation regulation and cell survival.

Conclusion: Unsupervised cluster analysis based on biomarker profiles identified mutually exclusive subgroups of patients with HFpEF with distinct biomarker profiles, clinical characteristics and outcomes, suggesting different underlying pathophysiological pathways.

Keywords: Heart Failure, Machine Learning, Heart Failure with Preserved Ejection Fraction, Cluster Analysis

Introduction:

Heart failure with preserved ejection fraction (HFpEF) is labelled as the "single largest unmet need in cardiovascular medicine"(1). Despite accounting for approximately half of heart failure (HF) cases, an effective therapy for HFpEF is yet to be found. Several pathological mechanisms have been proposed as underlying mechanisms in HFpEF including systemic microvascular inflammation, cardiometabolic abnormalities and cellular and extracellular structural changes. However, studies targeting these mechanisms have yielded mixed results and have yet to show improved prognosis in these patients (2).

The 'heterogeneity' of HFpEF has been cited as a reason for clinical trials not being effective in HFpEF patients, suggesting that a "one size fits all" approach may not work in HFpEF (1,3). The pathophysiology of HFpEF is highly complex (4). The identification of mutually exclusive subgroups of patients with HFpEF based on their underlying pathophysiology may allow for the development of targeted treatment options. Several studies have sought to identify subgroups of patients with HFpEF using machine learning techniques, classifying these patients into clinical phenotypes and advocated phenotype specific treatment of these subgroups (5–9). Such techniques have managed to identify subgroups of patients with similar phenotypic characteristics with differences in outcome. However, subgroups based on clinical characteristics do not necessarily represent differences in pathological mechanisms. Similar to HFpEF, heart failure with reduced ejection fraction (HFrEF) represents a heterogeneous group comprising of patients with a multitude of heart failure (HF) aetiologies, yet a common pathway of systemic neuroendocrine activation (2). Therefore, there are limitations in using this approach.

We hypothesised that unsupervised machine learning techniques applied to protein biomarkers in HFpEF patients would allow the identification of biological HFpEF subgroups, representing different pathological mechanisms in HFpEF.

Methods:

Patient population

This study utilized patients from the Scottish cohort of BIOSTAT-CHF, which is described elsewhere (10). In brief, the Scottish cohort of BIOSTAT-CHF includes 1738 patients from six centres in Scotland, United Kingdom. Patients were required to be ≥ 18 years of age, diagnosed with HF and previously admitted with HF requiring diuretic treatment. Patients were also sub-optimally treated with angiotensin-converting enzyme inhibitors (ACEi) and beta-blockers with an anticipated up-titration over the following 3 months.

Of the 1738 patients included, only patients with a left ejection fraction of ≥ 50% were included. Of the remaining 441 patients, there were 12 patients with missing biomarker values. Subsequent analyses were done with the remaining 429 patients.

Clinical and biomarker measurements

Medical history, physical examinations, echocardiography and laboratory tests were recorded at baseline as previously described (10).

An overview of biomarkers and their pathophysiological functions are presented in the supplementary material, *Figure S1 and S2*. Four biomarker panels comprising each of 92 protein biomarkers provided by the Olink Bioscience analysis service (Uppsala, Sweden) were measured. These respective panels were Cardiovascular II (CVII), CVD III, Immune response and Oncology II panels. The proteins were profiled using Olink Proseek® Multiplex Inflammatory^{96x96} platform. The Proseek® kit uses proximity extension assay (PEA) technology, whereby oligonucleotide-labelled antibody probe pairs bind to their respective targets. Quantification was achieved using a Fluidigm BioMark™ real-time PCR platform. The platform provides normalised protein expression (NPX, log2-normalised), rather than an absolute quantification.

Across the panels there were several duplicates therefore the mean of normalized protein expression duplicates was used (*supplementary materials*). Further analyses were completed using the set of 363 non-redundant protein biomarkers.

Statistical analysis

A more comprehensive description of statistical methods used is provided in the supplementary materials. In short, the primary aim of this study was to identify mutually exclusive subgroups of patients based on their biomarker profile using 363 biomarkers, which are referred to as clusters. Principle component analysis (PCA) was performed in order to reduce biomarker dimensions and collinearity. Clustering was performed on principle components (PCs) with an eigenvalue of one or above using a hierarchical clustering algorithm. The NbClust package in R, which utilises an array of different indexes, was used in order to determine the optimal number of clusters. The number most often selected by these indexes is then selected as the optimal number of clusters (11,12).

Differences between clinical characteristics of the clusters were compared using one-way analysis of variance (ANOVA), the Kruskal-Wallis test or chi-squared, χ^2, test where appropriate. Differences between clusters biomarker means were plotted using a heatmap after *z*-standardization.

The association of cluster membership with all-cause mortality and HF hospitalisation was investigated using Kaplan-Meier curves and the log-rank test. For multivariable analyses, cox-regression models were performed, correcting for age, sex and previous HF hospitalisation and New York Heart Association (NYHA) class.

Relative differences in protein biomarker levels between clusters was assessed using the *Limma* package in R (13). Proteins were considered differentially expressed at a $\log_2$ fold-change cut-off of 0.8 and false discovery rate corrected p-value of <0.05. Protein biomarkers identified to be differentially expressed were further investigated for pathway overrepresentation.

Over-representation was assessed using ClueGo (in Gene Ontology in biological processes, KEGG and Reactome pathways)(14). This was performed using the hypergeometric test and the default Bonferroni step down method for multiple testing corrections. The whole annotation option

was used as a reference set and only biological processes with a corrected p-value ≤ 0.05 were considered significant.

Results:

Clustering outcomes

The optimal number of clusters was six (*Figure S3*). Due to the small size, clusters 5 and 6 (n=3 and n=2, respectively, *table S1*) were excluded from the downstream analyses which focused on the remaining four patient clusters. Expression patterns of biomarkers across the four clusters are depicted in the heatmap in *Figure* 1. Lower levels of biomarker means are depicted in lighter colours, while darker colours represent higher biomarker levels. Cluster 1 biomarkers are markedly higher compared to other clusters, while cluster 3 shows low levels of almost all biomarkers.

Clinical characteristics

Baseline characteristics of the 4 clusters are presented in *Table 1*. Patients in cluster 1 had the highest prevalence of chronic kidney disease (73.7%) and diabetes mellitus (53.4%), and had the highest plasma concentrations of creatinine, glucose, gamma-glutamyl transferase (GGT) and growth differentiation factor-15 (GDF-15). Patients in cluster 2 were the eldest (mean age 80 years), with a high frequency of age-related comorbidities such as atrial fibrillation (46.5%) and hypertension (72.3%), however these did not reach significance. Patients in cluster 3 were youngest (mean age 74 years), had the lowest prevalence of most comorbidities, except obesity (mean BMI 30.4 kg/m^2, mean BSA 2.02 m^2), were the least symptomatic and had the lowest plasma NT-proBNP concentrations (median 591 pg/L). Patients in cluster 4 had the highest prevalence of COPD (41.9%), smoking (65.1%) and ischemic aetiology (72.1%); were the most symptomatic and had the highest levels of NT-proBNP (median 2175 pg/L) and troponin (median 46.2 ng/l). *Table 2* shows the echocardiographic characteristics of the four clusters. In terms of cardiac structure and function, clusters were remarkably similar, except for a lower estimated right ventricular systolic pressure and tricuspid regurgitation velocity in patients from cluster 3.

Differential expression analysis and pathway overrepresentation analysis

In cluster 1 there were 29 proteins that were significantly upregulated compared to the rest of the clusters (*supplementary table 2*). At the fold-change cut-off, no proteins were found to be significantly up or down regulated in cluster 2 compared to the other clusters. A total of 26 proteins were discovered to be significantly downregulated in cluster 3 (*supplementary table 3*). In cluster 4, 1 protein, was significantly downregulated, while 34 proteins were found to be significantly upregulated (*supplementary table 4*).

Pathway over-representation analysis of the differentially expressed proteins per cluster yielded several significant pathways. The 29 differentially expressed proteins in cluster 1 yielded four significant biological processes ($p < 0.001$): 'tumour necrosis factor-activated receptor activity'; 'TNFs bind to their physiological receptors'; 'regulation of natural killer cell mediated immunity' and 'regulation of regulatory T cell differentiation'. The 26 downregulated proteins in cluster 3 were significantly associated with the following biological processes ($p < 0.001$): 'tumour necrosis factor - activated receptor activity', 'viral protein interaction with cytokine and cytokine receptor' and 'regulation of cardiac muscle hypertrophy'. The 34 upregulated proteins in Cluster 4 were significantly associated with 6 biological processes ($p < 0.001$): 'protein serine/threonine kinase inhibitor activity'; 'regulation of receptor internalisation'; 'viral myocarditis'; 'Kaposi sarcoma-associated herpes virus infection'; 'PI3K/AKT signalling in cancer' and 'positive regulation of phosphatidylinositol 3-kinase activity'.

Clinical Outcome

After a median follow-up of 21 months, approximately 44% of patients either had been hospitalised for HF or died. The occurrence of death or HF hospitalization was highest in clusters 1 and 4 (62.1% and 62.8% respectively) and lowest in cluster 3 (25.6%). Rate of HF hospitalisation alone was highest in cluster 1 (36.2%), compared with 23.3% in cluster 2, 17.7% in cluster 3 and 20.9% in cluster 4 (*Figure 2*). After correction for age, sex, previous HF hospitalisation and NYHA class, compared to cluster 1, patients in cluster 2 and 3 had a lower risk of death or HF hospitalization

([hazard ratio (HR) 0.58; 95% confidence interval (CI) 0.39-0.87] and HR=0.30 (95% CI 0.19 – 0.48), respectively, *supplementary table 5*).

Discussion:

In this study, unsupervised machine learning identified distinct HFpEF clusters based on circulating biomarker profiles. The identified HFpEF clusters were remarkably different in their clinical characteristics and outcomes. Using a novel approach of employing differential expression analysis and pathways analysis on HFpEF clusters, we were able to identify dysregulated biological pathways in each cluster. This is the first study to provide a pathophysiological basis on a proteomic level of clinical phenotypes of HFpEF observed in previous studies.

Previous studies that have identified HFpEF subgroups via cluster analyses have typically done so based on clinical characteristics, echocardiographic and laboratory data (5–9). One such study conducted by *Shah et al.* demonstrated that clustering based on clinical data, or "phenomapping" provided a better discrimination of risk than NT-proBNP and risk scores alone (5). Nevertheless, studies employing these techniques have not yet assessed the pathophysiology underlying these clinical phenotypes. Our results therefore extend prior studies by providing insights into potential biological mechanisms at a proteomic level, that may underpin the observed clinical phenotypes, thus paving the way to mechanism-based pathophysiologic interventions in specific HFpEF subgroups. The added value of our biomarker-based approach is evident in that patients in different biomarker clusters could have similar echocardiographic and clinical profiles yet differ drastically in clinical outcome compared to other subgroups. One such example of this is clusters 1 and 2 – both these clusters were elderly patients with multiple age-related comorbidities who had indistinguishable echocardiographic features; yet cluster 2 had a 40% lower adjusted risk of death or HF hospitalization compared to cluster 1.

Clinical correlates of the clusters

The main discriminating clinical parameters identifying each of the biomarker clusters were (1) age; (2) diabetes and chronic kidney disease (CKD); and (3) smoking/COPD, and ischemic aetiology.

Age: Mean age was only 74 years in cluster 3, compared with 79-80 years in the other clusters. This was associated with lower co-morbidity burden in cluster 3 patients, except for obesity where there was significantly greater BSA and a trend towards higher BMI compared to other clusters. Obokata et al. describes the obese HFpEF phenotype which is characterised by increased concentric left ventricular remodelling, greater LV filling pressures and increased plasma volume despite lower NT-proBNP levels compared to non-obese HFpEF patients (15). A prospective study from 11 Asian regions sought to identify differences between 'young' (<65 years) and 'elderly' (≥ 75 years) HFpEF patients. This study found that younger age HFpEF was associated with a male majority, higher prevalence of obesity and lower NT-proBNP levels, while left ventricular filling pressures and left ventricular hypertrophy were comparative with the elderly HFpEF patients (10). Similar results were observed in 3 large HFpEF trials (TOPCAT [Aldosterone Antagonist Therapy for Adults With Heart Failure and Preserved Systolic Function], I-PRESERVE [Irbesartan in Heart Failure With Preserved Systolic Function], and CHARM Preserved [Candesartan Cilexetil in Heart Failure Assessment of Reduction in Mortality and Morbidity]), where younger patients with HFpEF were more often obese men, whereas older patients with HFpEF were more often women with a higher prevalence of atrial fibrillation, hypertension, and CKD (16). Of note, *Tromp et al*. also found a separate 'elderly/AF HFpEF' cluster (17), similar to cluster 2 in the present study; while *Shah et al*. identified a distinct elderly pheno-group with highest serum creatinine/ lowest GFR, and highest natriuretic peptides and MAGGIC risk score values, thus resembling a combination of clusters 1 and 2 in the current study (5).

Diabetes and CKD: Patients in cluster 1 had a substantially higher prevalence of diabetes than cluster 2,3, and 4. Similarly, *Tromp et al*. identified a 'lean diabetic HFpEF' cluster with a high prevalence of diabetes and CKD while clustering on comorbidities in a large unselected population of

Asian patients with HF (17). *Shah et al.* also identified a diabetes-predominant pheno-group among 3 HFpEF clusters based on clinical characteristics, although in this US-based cohort the diabetic pheno-group also had the highest prevalence of obesity and obstructive sleep apnea.(2) High prevalence of chronic kidney disease, observed in cluster 1 and to a lesser extent in cluster 2, have often been observed in HFpEF populations and represent a high risk phenotype with a poor prognosis (6–9,18). A prospective analysis by *Unger et al.* in HFpEF patients found that CKD was independently associated with several echocardiographic parameters, including left atrial (LA) reservoir strain and LV longitudinal strain, after adjusting for potential cofounders (18). Development of cardiac abnormalities prior to onset of clinical symptoms in patients with renal dysfunction underpins the hypothesis that the pathogenesis of CKD HFpEF subgroups, such as that identified by cluster 1, may lie with renal dysfunction and its downstream effects.

Smoking/ COPD and ischemic aetiology: Cluster 4 was predominantly characterized by a high prevalence of COPD (>40%), double that of the other "high-risk" cluster 1 (<20%), associated with smoking history, ischemic aetiology in >70% and highest troponin and NT-proBNP levels. Previous studies have not previously described subgroups of HFpEF patients with pulmonary disease. However, several studies have reported a close association between COPD and HFpEF. The Framingham Heart Study reported that several subclinical noncardiac organ dysfunctions were associated with increased risk of HF. In particular, ratio of forced expiratory volume in 1 second to forced vital capacity (FEV1:FVC ratio) was associated with incident HFpEF (19). The importance of the link with smoking history lies in the potential to modify the risk of both COPD and ischemic heart disease by smoking cessation. Indeed, current or prior smoking was associated with higher prevalence of coronary microvascular dysfunction among patients with HFpEF in the PROMIS-HFpEF (20). Both micro- and macro- vascular coronary artery disease may contribute to myocardial ischemia and worsening HFpEF (21). Importantly, revascularization in those with macrovascular coronary artery disease may be associated with preservation of cardiac function and improved outcomes in HFpEF (21).

Echocardiographic findings

Echocardiography was performed in all patients. We found remarkable similarities in cardiac phenotype between the clusters. Left ventricular size and function and left atrial dimensions were similar between the clusters. The only discriminating parameter was right ventricular systolic pressure, which was significantly lower in cluster 3, consistent with less severe symptoms and better outcomes in these patients.

Up- and down-regulated biological pathways

Pathways identified by upregulated biomarkers in cluster 1 were related to cells of both innate and adaptive immunity. Natural killer (NK) cells are important in repairing tissue damage and appear to be preventative against development of fibrosis (22). Chronic decreases in NK cells has been reported to be correlated with low-grade inflammation in the heart (14). In addition, the T cell-mediated immune response is implicated in cardiac remodelling and the progression of heart failure (23,24). In mice with cardiac hypertrophy, depletion of T cells led to reduced myocardial fibrosis and decrease infiltration of macrophages (24). In a previous cluster analysis *Cohen et al.* identified a CKD/DM phenogroup in HFpEF, which also exhibited biomarkers for tumour necrosis factor alpha mediated inflammation (9).

Differential expression analysis of cluster 2 did not result in any up- or down-regulated biomarkers relative to the other clusters at a fold-change of 0.8. Considering the overlap in clinical phenotype of cluster 1 and 2, we postulate that this may have overshadowed any differential expression of biomarkers in cluster 2.

Several significant pathways identified in cluster 4 were related to phosphoinositide 3-kinases (PI3K) and their downstream effects. These effects include signalling pathways involved in protein synthesis, cell proliferation, metabolism and cell survival and has been implicated in the pathogenesis of various diseases (25,26). Interestingly, increased PI3K/AKT axis activity is postulated to play an important role in cell senescence, which is considered as a key mechanism in COPD pathogenesis

(27). A study observed that increased PI3K/AKT axis activity was found in lung tissue and peripheral blood mononuclear cells of COPD patients when compared to controls (28). PI3K isoforms are found in both cardiomyocytes and leukocytes and have been implemented in pathways influencing hypertrophy, contractility, vascular and myocardial inflammation (29). In addition, AKT effects on monocytes/macrophages are postulated to have an effect on atherosclerosis formulation, with increased AKT signalling postulated to accelerate atherosclerosis (26). We postulate that increased PI3K/AKT axis in airway obstruction could have adverse effects on the heart and endothelium, leading to the development of heart failure.

Strengths and limitations

The strengths of this study are the use of large panel of biomarkers from several biological domains. This is especially important for HFpEF for which it is postulated to be a disease highly influenced by cardiac and non-cardiac comorbidities. In addition, by clustering on biomarkers rather on clinical variables, this allows for the identification of potential biological phenotypes that may represent underlying biological heterogeneity in HFpEF and in turn different pathophysiological mechanisms. We acknowledge there are several limitations to this study including the small number of patients and the lack of external validation. BIOSTAT is also primarily a Caucasian cohort and the extent to which the results of this study represent the general HFpEF population is unclear. Despite a lack of external validation, there does appear to be overlap between the results of cluster analyses and subgroups previously identified in other studies, suggesting that there is commonality across different methods and cohorts. We aim to validate our findings in further studies.

Conclusion:

Using unsupervised cluster analysis based on a broad range of circulating biomarkers, we identified four distinct clusters of HFpEF with remarkable differences in clinical characteristics and outcomes, potentially reflecting differences in underlying pathophysiology. Cluster 1 patients can be identified as those with diabetic nephropathy, high event rates and a specific activation of inflammatory pathways; cluster 2 are the elderly patients with frequent age-related comorbidities;

cluster 3 are young with low burden of comorbidities except obesity, lowest NT-proBNP levels and lowest risk of adverse outcomes; and cluster 4 are those with ischemic aetiology, smoking and chronic lung disease, most symptoms, as well as highest NT-proBNP and troponin levels, characterized by increased activity of the PI3K/AKT pathway and with the highest risk of death and/or heart failure hospitalization. Left ventricular and atrial size and function were not different among the groups. These data provide a pathophysiological basis on a proteomic level of clinical phenotypes of HFpEF observed in previous studies, and thus open the door to mechanism-based pathophysiologic interventions in specific HFpEF subgroups.

References:

1. Lewis GA, Schelbert EB, Williams SG, Cunnington C, Ahmed F, McDonagh TA, et al. Biological Phenotypes of Heart Failure With Preserved Ejection Fraction. J Am Coll Cardiol. 2017 Oct 24;70(17):2186–200.
2. Lam CSP, Voors AA, de Boer RA, Solomon SD, van Veldhuisen DJ. Heart failure with preserved ejection fraction: from mechanisms to therapies. Eur Heart J. 2018 07;39(30):2780–92.
3. Senni M, Paulus WJ, Gavazzi A, Fraser AG, Díez J, Solomon SD, et al. New strategies for heart failure with preserved ejection fraction: the importance of targeted therapies for heart failure phenotypes. Eur Heart J. 2014 Oct 21;35(40):2797–815.
4. Paulus WJ, Tschöpe C. A novel paradigm for heart failure with preserved ejection fraction: comorbidities drive myocardial dysfunction and remodeling through coronary microvascular endothelial inflammation. J Am Coll Cardiol. 2013 Jul 23;62(4):263–71.
5. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation. 2015 Jan 20;131(3):269–79.

6. Kao DP, Lewsey JD, Anand IS, Massie BM, Zile MR, Carson PE, et al. Characterization of subgroups of heart failure patients with preserved ejection fraction with possible implications for prognosis and treatment response. Eur J Heart Fail. 2015 Sep;17(9):925–35.

7. Segar MW, Patel KV, Ayers C, Basit M, Tang WHW, Willett D, et al. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. Eur J Heart Fail. 2020 Jan;22(1):148–58.

8. Hedman ÅK, Hage C, Sharma A, Brosnan MJ, Buckbinder L, Gan L-M, et al. Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning. Heart Br Card Soc. 2020 Mar;106(5):342–9.

9. Cohen JB, Schrauben SJ, Zhao L, Basso MD, Cvijic ME, Li Z, et al. Clinical Phenogroups in Heart Failure With Preserved Ejection Fraction: Detailed Phenotypes, Prognosis, and Response to Spironolactone. JACC Heart Fail. 2020 Mar;8(3):172–84.

10. Voors AA, Anker SD, Cleland JG, Dickstein K, Filippatos G, van der Harst P, et al. A systems BIOlogy Study to TAilored Treatment in Chronic Heart Failure: rationale, design, and baseline characteristics of BIOSTAT-CHF. Eur J Heart Fail. 2016;18(6):716–26.

11. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. J Stat Softw. 2014 Nov 3;61(1):1–36.

12. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; YEAR. Available from: https://www.R-project.org

13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

14. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinforma Oxf Engl. 2009 Apr 15;25(8):1091–3.

15. Obokata M, Reddy YNV, Pislaru SV, Melenovsky V, Borlaug BA. Evidence Supporting the Existence of a Distinct Obese Phenotype of Heart Failure With Preserved Ejection Fraction. Circulation. 2017 Jul 4;136(1):6–19.

16. Tromp J, Shen L, Jhund PS, Anand IS, Carson PE, Desai AS, et al. Age-Related Characteristics and Outcomes of Patients With Heart Failure With Preserved Ejection Fraction. J Am Coll Cardiol. 2019 06;74(5):601–12.

17. Tromp J, Tay WT, Ouwerkerk W, Teng T-HK, Yap J, MacDonald MR, et al. Multimorbidity in patients with heart failure from 11 Asian regions: A prospective cohort study using the ASIAN-HF registry. PLOS Med. 2018 Mar 27;15(3):e1002541.

18. Unger ED, Dubin RF, Deo R, Daruwalla V, Friedman JL, Medina C, et al. Association of chronic kidney disease with abnormal cardiac mechanics and adverse outcomes in patients with heart failure and preserved ejection fraction. Eur J Heart Fail. 2016 Jan;18(1):103–12.

19. Lam CSP, Lyass A, Kraigher-Krainer E, Massaro JM, Lee DS, Ho JE, et al. Cardiac dysfunction and noncardiac dysfunction as precursors of heart failure with reduced and preserved ejection fraction in the community. Circulation. 2011 Jul 5;124(1):24–30.

20. Shah SJ, Lam CSP, Svedlund S, Saraste A, Hage C, Tan R-S, et al. Prevalence and correlates of coronary microvascular dysfunction in heart failure with preserved ejection fraction: PROMIS-HFpEF. Eur Heart J. 2018 01;39(37):3439–50.

21. Hwang S-J, Melenovsky V, Borlaug BA. Implications of coronary artery disease in heart failure with preserved ejection fraction. J Am Coll Cardiol. 2014 Jul 1;63(25 Pt A):2817–27.

22. Ong S, Rose NR, Čiháková D. Natural killer cells in inflammatory heart disease. Clin Immunol Orlando Fla. 2017;175:26–33.

23. Strassheim D, Dempsey EC, Gerasimovskaya E, Stenmark K, Karoor V. Role of Inflammatory Cell Subtypes in Heart Failure. J Immunol Res. 2019;2019:2164017.

24. Frieler RA, Mortensen RM. Immune cell and other noncardiomyocyte regulation of cardiac hypertrophy and remodeling. Circulation. 2015 Mar 17;131(11):1019–30.

25. Sala V, Margaria JP, Murabito A, Morello F, Ghigo A, Hirsch E. Therapeutic Targeting of PDEs and PI3K in Heart Failure with Preserved Ejection Fraction (HFpEF). Curr Heart Fail Rep. 2017;14(3):187–96.

26. Linton MF, Moslehi JJ, Babaev VR. Akt Signaling in Macrophage Polarization, Survival, and Atherosclerosis. Int J Mol Sci. 2019 Jun 1;20(11).

27. Mercado N, Ito K, Barnes PJ. Accelerated ageing of the lung in COPD: new concepts. Thorax. 2015 May;70(5):482–9.

28. To Y, Ito K, Kizawa Y, Failla M, Ito M, Kusama T, et al. Targeting phosphoinositide-3-kinase-delta with theophylline reverses corticosteroid insensitivity in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2010 Oct 1;182(7):897–904.

29. Ghigo A, Li M. Phosphoinositide 3-kinase: friend and foe in cardiovascular disease. Front Pharmacol. 2015;6:169.

Figures and Tables:

Figure 1. Heatmap displaying biomarkers across HFpEF clusters

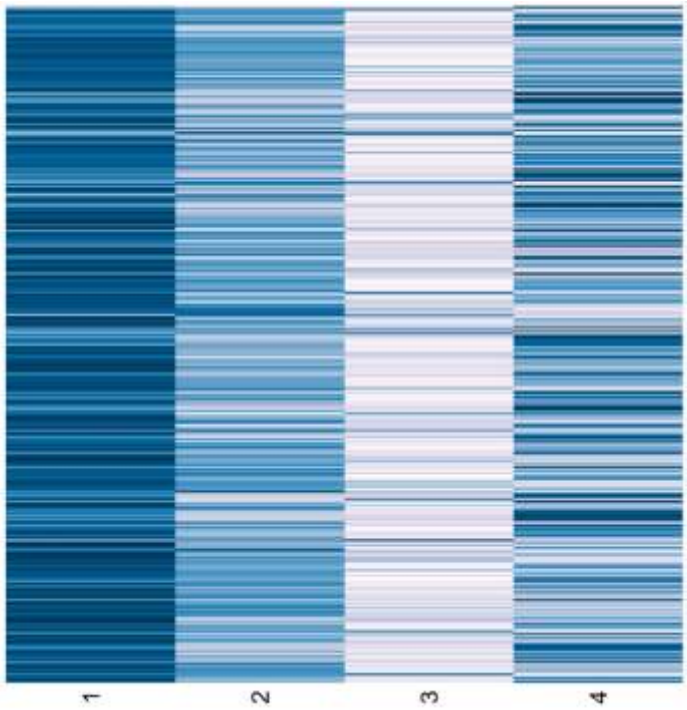


Table 1. Baseline characteristics stratified by HFpEF cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	p-value
N	58	159	164	43	
Demographics					
Age (years)	79.0 [73.0;82.0]	80.0 [72.5;86.5]	74.0 [66.8;81.0]	79.0 [71.5;83.0]	<0.001
Male (%)	35 (60.3%)	84 (52.8%)	93 (56.7%)	26 (60.5%)	0.689
BMI (kg/m^2)	29.1 [24.8;33.2]	28.4 [24.2;34.2]	30.4 [25.4;34.7]	29.4 [25.4;33.4]	0.408
BSA (m^2)	1.97 [1.80;2.11]	1.85 [1.68;2.01]	2.02 [1.79;2.21]	2.01 [1.79;2.13]	0.032
Ischaemic aetiology (%)	31 (54.4%)	99 (65.1%)	101 (66.4%)	31 (72.1%)	0.272
Cardiomyopathy (%)	4 (7.02%)	4 (2.63%)	3 (1.97%)	1 (2.33%)	0.293
NYHA, *n* (%)					<0.001
I	0(0.0%)	0(0.0%)	4(2.44%)	0(0.0%)	

II	15(25.9%)	37(23.3%)	72(43.9%)	8(18.6%)	
III	32(55.2%)	83(52.2%)	65(39.6%)	24(55.8%)	
IV	11(19.0%)	39(24.5%)	23(14.0%)	11(25.6%)	
Medical History (%)					
Anaemia	27(46.6%)	72(45.9%)	50(30.9%)	22(51.2%)	0.012
Atrial fibrillation	31(54.4%)	74(46.5%)	67(41.1%)	24(55.8%)	0.185
Diabetes	31(53.4%)	53(33.3%)	57(35.2%)	15(35.7%)	0.048
COPD	11(19.0%)	44(27.7%)	36(22.1%)	18(41.9%)	0.034
CKD	42(73.7%)	107(67.7%)	34(21.5%)	23(53.5%)	<0.001
Hypertension	42(72.4%)	115(72.3%)	102(62.2%)	31(72.1%)	0.190
Malignancy	6(10.5%)	5(3.14%)	6(3.66%)	2(4.65%)	0.141
Obesity	25(43.9%)	69(43.7%)	86(53.8%)	19(44.2%)	0.270
Stroke	13(22.4%)	31(19.5%)	30(18.4%)	9(21.4%)	0.913
Past/current smokers	25(43.9%)	72(45.9%)	69(42.3%)	28(65.1%)	0.063
Signs and symptoms					
Extent of peripheral oedema, *n* (%)					0.121
Not present	7(13.0%)	46(31.7%)	54(37.0%)	9(22.5%)	
Ankle	22(40.7%)	52(35.9%)	42(28.8%)	15(37.5%)	
Below knee	18(33.3%)	34(23.4%)	40(27.4%)	12(30.0%)	
Above knee	7(13.0%)	13(8.97%)	10(6.85%)	4(10.0%)	
JVP elevated (%)	24(49.0%)	41(29.1%)	45(34.1%)	9(25.7%)	0.035
Pulmonary congestion with rales	36(64.3%)	75(48.4%)	70(45.2%)	32(74.4%)	0.001
Laboratory					
Haemoglobin, (g/dl)	12.2 [11.2;13.5]	12.5 [10.9;13.5]	13.3 [12.3;14.4]	12.5 [10.8;13.8]	<0.001
Leukocytes, (10^9/L)	7.45 [5.55;10.0]	7.55 [5.93;9.00]	7.25 [5.90;8.80]	9.90 [7.50;12.8]	<0.001
Creatinine, (µmol/L)	126 [96.0;148]	112 [88.0;142]	81.5 [66.8;94.2]	96.0 [79.0;121]	<0.001
Urea, (mmol/L)	11.1 [7.90;15.3]	9.80 [7.93;13.5]	6.60 [5.47;8.43]	10.2 [7.90;13.9]	<0.001
e-GFR, (mL/min/1.73m2)	46.0 [36.0;59.5]	51.0 [38.0;60.0]	60.0 [60.0;60.0]	58.0 [48.0;60.0]	<0.001
Gamma-GT, (U/L)	61.0 [39.5;138]	44.0 [27.0;80.0]	35.0 [25.0;62.0]	47.0 [32.0;94.8]	0.001
Glucose, (mmol/L)	7.90 [5.60;10.6]	6.60 [5.60;9.10]	6.05 [5.10;8.40]	7.80 [6.00;10.2]	0.002

Iron, (micromol/L)	8.00 [6.00;12.0]	9.00 [6.00;13.0]	13.0 [8.00;16.0]	8.00 [4.25;11.8]	<0.001
Ferritin, (ng/mL)	154 [58.0;270]	93.0 [43.0;209]	94.5 [35.5;202]	98.0 [52.0;262]	0.162
NT-proBNP, (pg/L)	1720 [544;4831]	1304 [526;2938]	591 [234;1621]	2175 [898;4542]	<0.001
GDF-15, (pg/mL)	5877 [3412;8555]	3510 [2507;5228]	2174 [1532;2982]	3777 [2815;5970]	<0.001
Troponin T (ng/l)	45.0 [26.9;79.6]	32.6 [18.9;63.1]	19.0 [12.4;31.3]	46.2 [30.4;271]	<0.001

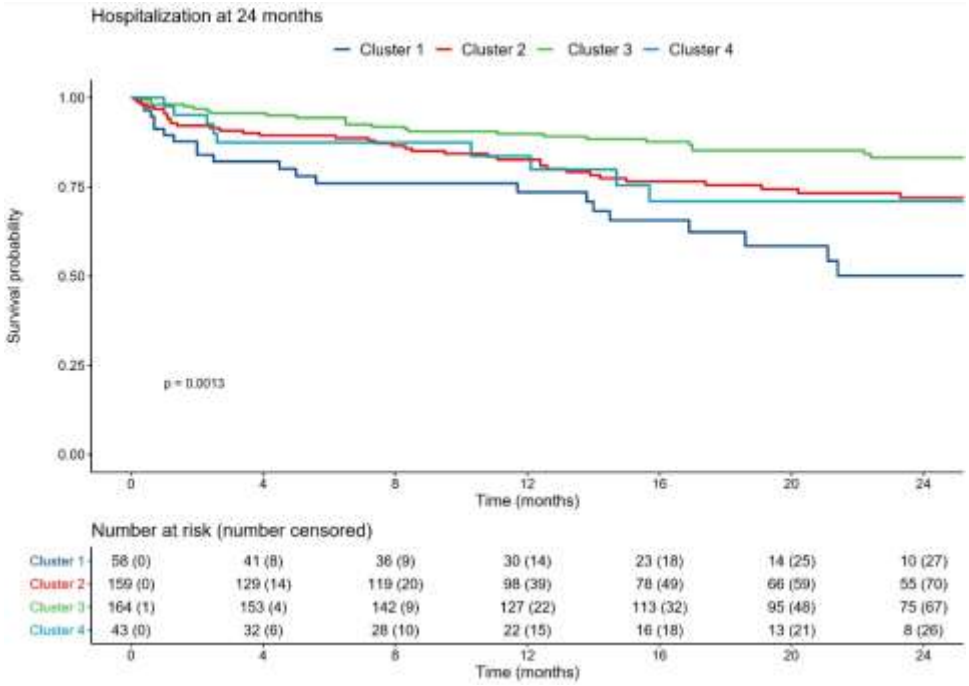
Table 2. Echocardiography stratified by cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	p-value
N	58	159	164	43	
LVEF (%)	54.0 [50.0;60.0]	55.0 [50.0;60.0]	54.5 [50.0;60.0]	55.0 [50.0;60.0]	0.953
LVEDD (mm)	46.2(6.52)	47.9(6.96)	48.6(7.71)	48.5(7.25)	0.359
LVESD (mm)	30.0 [26.0;33.5]	31.0 [24.0;37.0]	33.0 [26.0;37.5]	30.0 [24.5;38.5]	0.844
IVSd (mm)	13.0 [12.0;15.0]	13.0 [11.8;15.0]	13.0 [10.0;15.0]	13.0 [12.8;15.0]	0.502
PWd (mm)	12.0 [11.0;13.0]	11.0 [9.00;14.0]	10.0 [9.00;13.0]	12.0 [10.0;13.0]	0.477
Left atrial diameter (mm)	43.9(6.30)	45.0(7.15)	43.2(7.27)	44.6(6.87)	0.318
E/A ratio	0.80 [0.70;1.10]	1.00 [0.70;1.40]	0.90 [0.70;1.20]	0.90 [0.80;1.10]	0.671
IVC dilated	10(17.2%)	26(16.4%)	26(15.9%)	5(11.6%)	0.876
Right atrial pressure (mmHg)	10.0 [7.25;17.2]	10.0 [7.00;13.0]	9.00 [7.00;10.0]	9.00 [8.00;10.0]	0.440
RVSP (mmHg)	49.0 [39.0;63.5]	43.0 [35.0;55.0]	37.0 [30.0;49.0]	47.0 [37.0;52.0]	0.004
Right ventricular diameter ≥ 44mm	11(21.6%)	26(20.5%)	27(19.4%)	8(22.2%)	0.978
Tricuspid regurgitation gradient	36.0 [28.0;50.0]	34.0 [27.0;42.5]	29.0 [24.0;38.0]	40.0 [32.0;45.5]	0.004

LVEDD = left ventricular end diastolic dimension; LVESD = left ventricular end systolic dimension;
IVSd = septal wall thickness; PWd = posterior wall thickness ; ICV = ; RVSP = right ventricular systolic pressure.

Figure 2. Kaplan-Meier curves for A) heart failure hospitalisation and B) combined outcome of all-cause mortality and /or heart failure hospitalisation at 24 months stratified according to clusters.

(A)



(B)

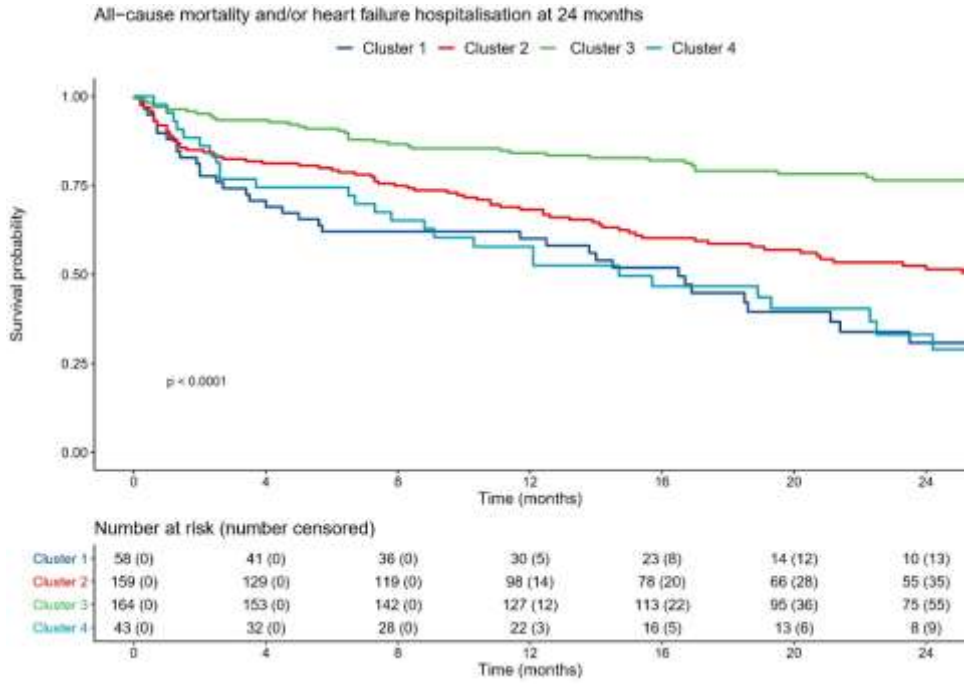
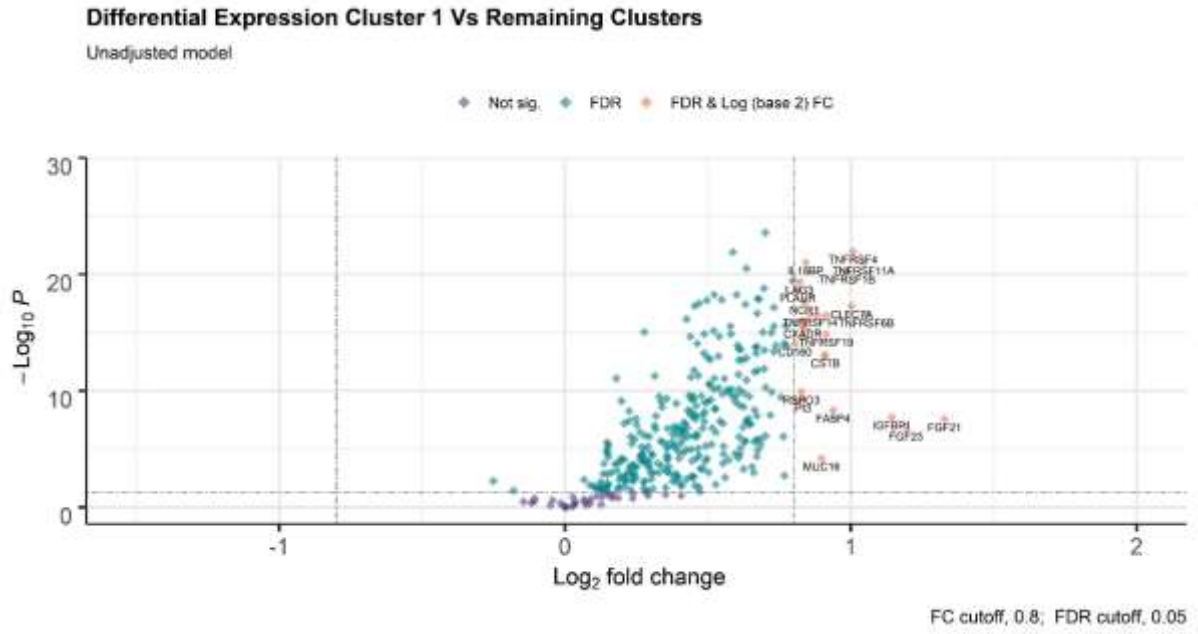
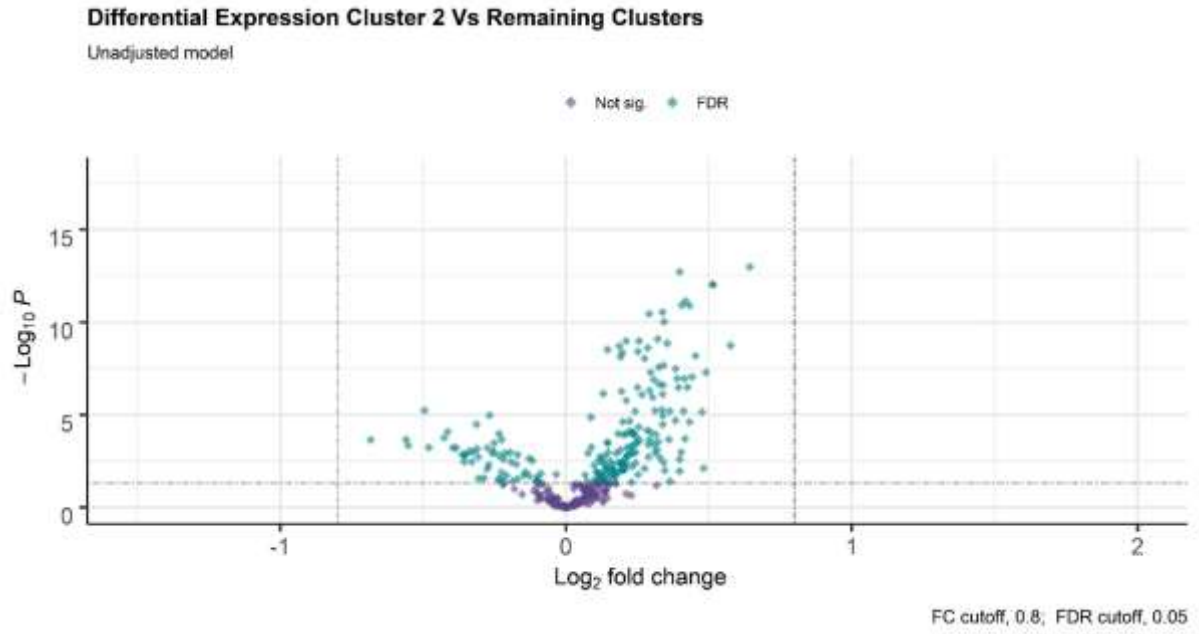


Figure 3. Differentially expressed proteins relative to other HFpEF clusters. A) Cluster 1 vs. Clusters 2-4, B) Cluster 2 vs Clusters 1,3 and 4, C) Cluster 3 vs Clusters 1,2 and 4 and D) Cluster 4 vs Clusters 1-3.

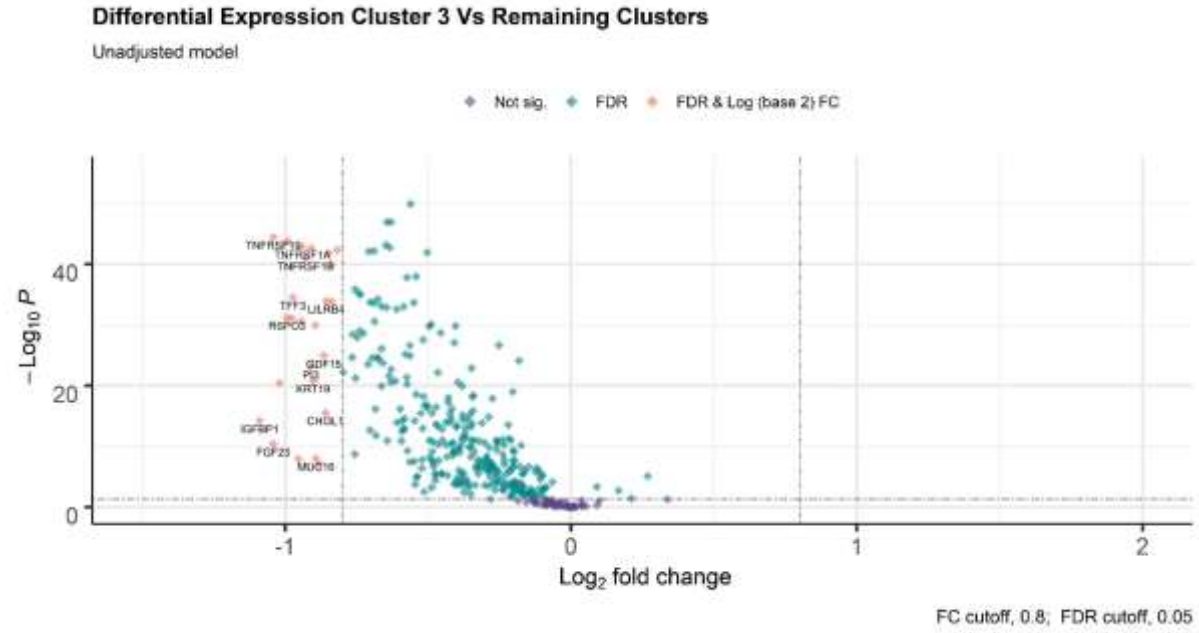
(A)



(B)



(C)



(D)

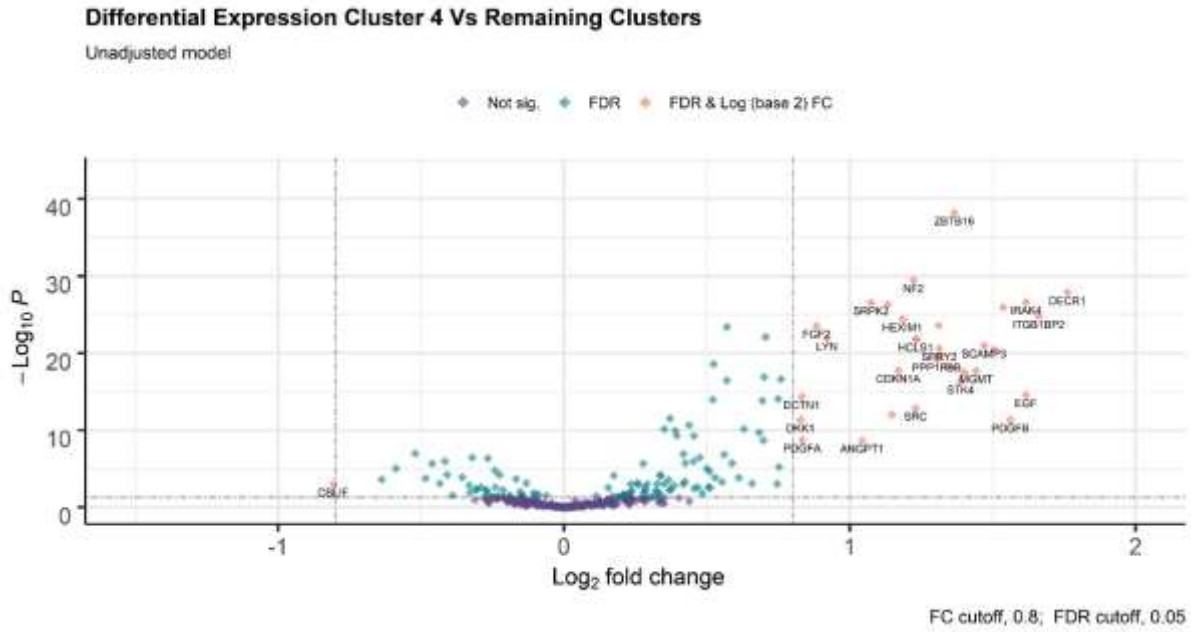
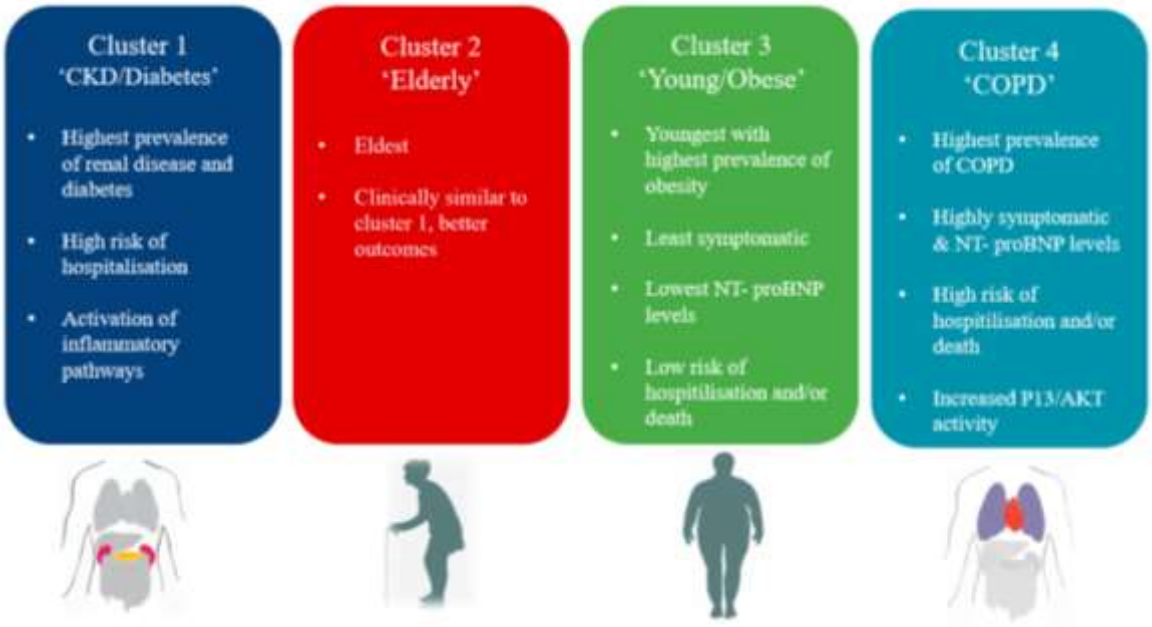


Figure 4. Take-home figure



Acknowledgement: This work was supported by the Netherlands Cardiovascular Research Initiative: an initiative with support of the Dutch Heart Foundation [CVON2014-11 RECONNECT].

Disclosures:

SDA reports receiving grant support from Abbott and Vifor Pharma, and fees from Abbott, Bayer, Boehringer Ingelheim, Cardiac Dimension, Impulse Dynamics, Novartis, Servier, and Vifor Pharma

G.F. reports being a Committee Member in trials sponsored by Medtronic, Vifor, Servier, Novartis, and BI, outside the submitted work.

MM received consulting honoraria from Abbott vascular, Actelion, Amgen, Astra Zeneca, Bayer, Edwards Therapeutics, Servier, Vifor pharma, WindTree for participation to trials' committees or speeches at sponsored meetings.

C.S.L. is supported by a Clinician Scientist Award from the National Medical Research Council of Singapore; has received research support from Boston Scientific, Bayer, Roche Diagnostics, AstraZeneca, Medtronic, and Vifor Pharma; has served as consultant or on the Advisory Board/ Steering Committee/ Executive Committee for Abbott Diagnostics, Amgen, Applied Therapeutics, AstraZeneca, Bayer, Biofourmis, Boehringer Ingelheim, Boston Scientific, Corvia Medical, Cytokinetics, Darma Inc., Eko.ai Pte Ltd, JanaCare, Janssen Research & Development LLC, Medtronic, Menarini Group, Merck, MyoKardia, Novartis, Novo Nordisk, Radcliffe Group Ltd., Roche Diagnostics, Sanofi, Stealth BioTherapeutics, The Corpus, Vifor Pharma and WebMD Global LLC; and serves as co-founder & non-executive director of EKo.ai Pte Ltd.