



HAL
open science

Unsupervised clustering to differentiate rheumatoid arthritis patients based on proteomic signatures

Maria Betânia Ferreira, Masatake Kobayashi, Rita Quelhas Costa, Tomás Fonseca, Mariana Brandão, Jose Carlos Oliveira, António Marinho, Henrique Cyrne Carvalho, Patrícia Rodrigues, Faiez Zannad, et al.

► To cite this version:

Maria Betânia Ferreira, Masatake Kobayashi, Rita Quelhas Costa, Tomás Fonseca, Mariana Brandão, et al.. Unsupervised clustering to differentiate rheumatoid arthritis patients based on proteomic signatures. *Scandinavian Journal of Rheumatology*, 2023, pp.1-8. 10.1080/03009742.2023.2196781 . hal-04089213

HAL Id: hal-04089213

<https://hal.univ-lorraine.fr/hal-04089213v1>

Submitted on 4 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Unsupervised Clustering to Differentiate Rheumatoid Arthritis Patients based on Proteomic Signatures

Maria Betânia Ferreira ^{1,2}; Masatake Kobayashi ³; Rita Quelhas Costa ⁴; Tomás Fonseca ^{5,7}; Mariana Brandão ⁷; Jose Carlos Oliveira ⁸; António Marinho ^{1,7}; Henrique Cyrne Carvalho ^{1,6}; Patrícia Rodrigues ⁶; Faiez Zannad ³; Patrick Rossignol ³; António S. Barros ⁹; João Pedro Ferreira ^{3,9}

1. UMIB - Unidade Multidisciplinar de Investigação Biomédica, ICBAS - Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal;
2. Internal Medicine Department, Hospital da Luz Arrábida, Porto, Portugal
3. Université de Lorraine, INSERM, Centre d'Investigations Cliniques Plurithématique 1433, Inserm U1116, CHRU de Nancy and F-CRIN INI-CRCT, Nancy, France.
4. Internal Medicine Department, Centro Hospitalar de Entre o Douro e Vouga, Aveiro, Portugal
5. Internal Medicine Department, Centro Hospitalar Universitário do Porto, Porto, Portugal
6. Cardiology Department, Centro Hospitalar Universitário do Porto, Porto, Portugal
7. Unidade de Imunologia Clínica, Centro Hospitalar Universitário do Porto, Porto, Portugal
8. Clinical Chemistry Service, Centro Hospitalar Universitário do Porto, Porto, Portugal
9. UnIC@RISE, Department of Surgery and Physiology, Faculty of Medicine of the University of Porto, Porto, Portugal & Heart Failure Clinic, Centro Hospitalar de Vila Nova de Gaia/Espinho, Portugal.

Corresponding author:

Dr João Pedro Ferreira

CIM - Centro de Investigação Médica

Faculdade de Medicina UP - Piso 6 - DCF

Rua Plácido da Costa, s/n
4200-450 Porto, Portugal
Tel.: +351 22 551 3600
Mail: jpferreira@med.up.pt

ABSTRACT

Objectives: Patients with rheumatoid arthritis (RA) have different presentations and prognoses. Cluster analysis based on proteomic signatures creates independent phenogroups of patients with different pathophysiologic backgrounds. We aim to identify distinct pathophysiologic clusters of RA patients based on circulating proteomic biomarkers.

Methods: Cohort study including 399 RA patients. Clustering was performed on 94 circulating proteins (92 CVDII Olink® + high sensitivity troponin T + C-reactive protein). Unsupervised clustering was performed using a partitioning cluster algorithm.

Results: The clustering algorithm identified 2 distinct clusters: cluster 1 (n=223) and cluster 2 (n=176). Compared with cluster 1, cluster 2 included older patients with a higher burden of comorbidities (cardiovascular and RA-related), more erosive and longer RA duration, more dyspnea and fatigue, walking a shorter distance in the 6-minute walking test, with more severe diastolic dysfunction and a 4.5-fold higher risk of death or hospitalization for cardiovascular reasons. TNF-Receptor superfamily-related pathways were mainly responsible for the model's discriminative ability.

Conclusion: Using unsupervised cluster analysis based on proteomic phenotypes, we identified two clusters of RA patients with distinct biomarkers profiles, clinical characteristics and different outcomes that could reflect different pathophysiologic backgrounds. TNF-Receptor superfamily-related proteins may be used to distinguish subgroups.

ClinicalTrials.gov ID: NCT03960515

Key-words: rheumatoid arthritis; cluster analysis; prognosis; TNF-receptor superfamily; TNFRSF11A; TRAILR2

Introduction

Rheumatoid arthritis (RA) affects approximately 1% of the worldwide population¹⁻³. Systemic inflammation plays a pivotal role in the underlying mechanism of RA, which may also add to the high cardiovascular risk of some patients with RA. Therefore, identifying such patients with high risk is important to better understand the disease's pathophysiology, ascertain prognosis, and potentially develop targeted therapies.

Unsupervised machine-learning (ML) algorithms may help identify independent phenogroups of patients with different disease pathophysiology and prognoses^{4,5}. For example, in patients with heart failure with preserved ejection fraction (HFpEF), cluster analysis based on proteomics patterns enabled identifying distinct disease phenotypes such as those at a high risk of cardiovascular events or those who benefited from specific disease treatments⁶⁻⁸.

In the present study we aimed to identify mutually exclusive phenogroups of RA patients based on a wide range of inflammatory and cardiovascular biomarkers. These biomarker-based phenogroups are then compared for clinical characteristics, biomarker profiles and outcomes.

METHODS

Study population

This prospective single-center study includes consecutive patients with RA aged 18 years or older followed in the Autoimmune Disease Outpatient Unit of *Centro Hospitalar Universitário do Porto*, Portugal, from June 2016 to June 2018 (ClinicalTrials.gov identifier: NCT03960515)⁹. RA was diagnosed based on the 2010 ACR/EULAR Classification criteria¹⁰. Patients who presented active neoplasm or had

a short life expectancy (<6 months), severe dementia, or severe frailty (e.g., inability to walk or totally dependent on a third person) were excluded.

Cardiovascular hospitalizations and deaths were prospectively recorded by regularly calling patients and their families, and then cross-checking the given information with medical records and hospital registries. Events were adjudicated by MBF and TF using a standardized event definition form.

This study was conducted following the principles of the Declaration of Helsinki and approved by the hospital ethics committee under number 2016-023 (020-DEFI/020-CES). All patients signed written informed consent prior to entry into the study.

We systematically filled the clinical data of all patients under a prespecified case report form (CRF) explicitly built for this project. An independent external data cleaning, consolidation, and verification to ensure data accuracy was performed.

Patient evaluation, echocardiogram, and routine laboratory tests

At baseline/first study visit, we collected medical history, physical examination, treatments, and a RA-specific questionnaire. A six-minute walking test (6MWT) was performed according to the American Thoracic Society guidelines¹¹. Echocardiogram was performed by an experienced echocardiographer, blinded to clinical data, following international recommendations^{12,13}.

Routine blood laboratory tests (including complete blood count, c-reactive protein [CRP], glucose, blood electrolytes, lipid profile, creatinine, high sensitivity troponin T [hs-TnT], and N-Terminal-pro Brain Natriuretic Peptide [NT-pro BNP]), were collected at baseline/first study visit and analyzed at the central laboratory of the hospital. CRP was measured by the enzyme-linked immunosorbent assay (Olympus CRP Latex Calibrator Normal Set®), hs-TnT by the Elecsys (Roche Diagnostics®),

and NT-pro BNP by Gen 5 STAT test (Roche Diagnostics®). NT-pro BNP, hs-TnT and CRP showed an intra-assay and inter-assay coefficient of variation of 10% at a concentration lower than the 99th percentile cut-off. At baseline/first study visit, plasma was frozen at -80°C in the central laboratory of the hospital and then sent to Uppsala, Sweden for Olink® analysis (see below).

Plasma Olink® Biomarkers

We measured a large protein biomarker panel (Olink® CVDII panel) that comprised 92 biomarkers from a wide range of pathophysiological domains. The Olink® CVDII panel was purposefully selected because it contains several known human circulating proteins associated with cardiovascular and inflammatory diseases. The overview of the 92 circulating proteins is depicted in **Supplement Table 1**, including their full names and Uniprot ID. In **Supplement Table 2**, the proteins are described according to their main biological functions. The full description and technical details of the analytical assay are described in the **Supplemental Addenda**. The CVDII Olink® panel displayed a mean intra-assay and inter-assay coefficient of variation between 9.1% and 11.7%.

Statistical analysis

Cluster analysis was performed based on 94 biomarkers (92 Olink® plus hs-TnT and C-reactive protein), which were centered and unit-scaled to allow each biomarker to contribute equally to cluster discovery. K-means partitioning was used to unravel biomarker clustering latent structure (**Supplement Figure 1**). This procedure minimized the within-cluster variation. The average silhouette statistics were used to find the optimal number of clusters. As shown in **Figure 1**, two clusters were identified, almost perfectly separated. The starting centroid was set at random (with

50 randomization starts to improve cluster stability), k-means was run with 300 bootstrapped samples.

After identifying the patient clusters, logistic regression was used to assess each biomarker's influence on the separation of the clusters. The contribution of each protein biomarker toward patient clusters was evaluated by the impact on the biomarker added value to the classification by using the AIC (Akaike information criterion), **Supplement Figure 2**.

The primary outcome was the composite of cardiovascular mortality or hospitalization for cardiovascular reasons (i.e., heart failure, acute myocardial infarction, stroke or sudden death). Survival probabilities were estimated using Cox proportional hazards regression model and the Kaplan-Meier method stratified according to the patient clusters and adjusting on age, sex, diabetes, RA duration, and estimated glomerular filtration rate (eGFR) (due to their clinical and prognostic value and for consistency across reports from this cohort)⁹. No data imputation was performed.

All statistical analyses were performed using R version 4.0.3 (R Development Core Team, Vienna, Austria). A P-value <0.05 was considered statistically significant.

Results

A total of 399 patients with RA were included in the present analysis; the median age was 61 years, 77% were female, 46% had hypertension, and the median left ventricular ejection fraction (LVEF) was 61%. The median RA duration of these patients was 9 years, 39% had articular erosion and 45% took corticosteroids.

The unsupervised cluster analysis identified 2 groups with distinct proteomic profiles, labelled as cluster 1 (n=223) and cluster 2 (n=176). The comparison of the baseline characteristics between the 2 clusters is presented in **Table 1**. There were significant between-cluster differences in age, several cardiovascular risk factors and comorbidities (i.e., diabetes, hypertension, obesity, and heart failure history), RA duration and erosive disease, dyspnea, fatigue, walked distance, systolic and diastolic dysfunction, and renal function with all the worse clinical conditions observed in cluster 2.

Proteomic biomarkers independently expressed between clusters

Of the 94 measured proteins, 86 (91%) were differentially expressed between clusters with higher expression levels in cluster 2. **Supplement Table 3**. The adjusted logistic regression model showed that TRAILR2 (TNF-receptor superfamily member 10b) and TNFRSF11A (TNF-receptor superfamily member 11a) were the 2 biomarkers with a higher impact on the model's predictive ability (**Supplement Figures 3a & 3b**). Compared with all biomarkers, these two proteins showed a good accuracy for separating our clusters (accuracy =0.83). **Figure 2** and **Table 2**. This simpler model with similar accuracy is preferred over a more complex model, because it reduces the risk of overfitting and makes the predictions easier to implement in a clinical setting.

Outcome

During a median follow-up period of 1.5 (percentile₂₅₋₇₅ 0.7-2.3) years, 41 patients had the primary outcome (combined primary cardiovascular endpoint). After adjustment for age, sex, a history of diabetes, RA duration and eGFR, cluster 2 was independently and significantly associated with a higher risk of the primary outcome

compared with cluster 1 (adjusted-HR 4.5, 95%CI 2.2-9.1, $p < 0.0001$) (**Figure 3**). In fact, the cluster risk separation clearly identified two independent phenotypes with almost all events occurring in cluster 2 patients, while in cluster 1 almost no events were recorded within 2 years of follow-up.

Discussion

Unsupervised ML approach based on 94 proteomic data applied to 399 patients with RA, identified 2 distinct clusters: cluster 2 - the high-risk cluster and cluster 1 - the low-risk cluster. Cluster separation was unsupervised and agnostic to the cardiovascular status of the patients i.e., it was based on circulating proteomic markers only. These 2 mutually exclusive clusters were organized mainly due to a heightened expression of TNF-receptors superfamily-related proteins (TRAILR2 and TNFRSF11A), reinforcing TNF superfamily-related pathways' relevance identify subgroups of patients with RA with different characteristics, symptom severity and prognosis. Supporting these findings, a long-term anti-TNF- α therapy for the treatment of RA has been shown to reduce the risk of RA-related complications, improve cardiovascular outcomes, and quality of life¹⁴⁻¹⁶.

To the best of our knowledge, the present study is the first to use multiple circulating proteins to identify phenogroups of RA patients with distinct characteristics and prognosis. Other recent studies in RA patients performed cluster analysis based on clinical characteristics but not on biomarker profiles.^{17,18} For example, one study used 14 clinical features to identify 4 subgroups with different risk profiles and potential for initiation of biologic disease-modifying anti-rheumatic drugs.¹⁸ Another

study tried to distinguish patients with "true refractory RA" from "non-adherent dissatisfied patients" based on clinical data.¹⁷.

Proteomic analysis leveraging proteins involved in diverse disease pathways allows to identify subgroups of patients with the same disease but with a different molecular and pathophysiologic fingerprint, and different prognoses. Whether these proteomic footprints identified herein may help identifying treatment responders requires adequate testing in prospective trials.

The existence of two circulating proteins members of the TNF receptor superfamily (i.e., TNFRSF11A, TRAIL-R2) identified in our study could facilitate the clinical applicability of our results. These two proteins might help in patient screening, prognostic stratification and potentially identify patients' responses to therapeutic interventions. Thus, the TNF superfamily pathway was central to our cluster identification. The TNF superfamily system leads to a vast communication network among various cell types and tissues that mediates signaling that controls the survival, proliferation, and cell differentiation¹⁹.

Regarding the two "top" proteins identified in our study, their biological activity may explain their central in cluster separation. Specifically, TRAILR2 exacerbated autoimmune arthritis by enhancing both cellular and humoral immune responses, leading to the amplification of inflammation, hyperproliferation of synovial cells and arthritogenic lymphocytes and increased production of cytokines (IL-2 and INF-gamma). Furthermore, increased levels of TRAILR2 were associated with a poor cardiovascular prognosis²⁰⁻²⁴. TNFRSF11A has an important role in bone metabolism regulation and appears to be involved in immune response²⁵. Like TRAILR2, TNFRSF11A is also associated with poor cardiovascular outcomes^{9,26}.

This study is innovative and these biomarkers (particularly TRAILR2 and TNFRSF11A) could be used to differentiate RA patients according to pathophysiological types and have potential for tailoring treatment if adequately tested in prospective randomized trials.

The limitations of this study include that we lacked external validation of our presented proteomic phenotypes in other RA cohorts. This analysis was exploratory, and no power calculation was performed. Given the relatively small sample size, and short follow-up time with few events, this analysis should be regarded as exploratory and further replication in larger cohorts with longer follow-up is required. In addition, the observational nature of this study precludes from inferring any causality. Our RA population had a long history of disease (median of 9 years) and many were still treated with corticosteroids and methotrexate; therefore, these data may not be necessarily applied to more contemporary cohorts.

Conclusion

Using unsupervised cluster analysis based on proteomic phenotypes, we identified two clusters of RA patients with distinct biomarkers profiles, clinical characteristics and different outcomes that could reflect different pathophysiologic backgrounds. TNF-Receptor superfamily-related proteins may be used to distinguish subgroups.

Statements

Funding statement

No funding was received for this study. Olink® analyses were made within the HOMAGE analyses package that had independent FP7-HEALTH (Grant agreement ID: 305507) funding.

Conflicts of interests

The authors declare that they have no conflicts of interest regarding this work.

Ethics

This study was conducted following the principles of the Declaration of Helsinki and approved by the *Hospital de Santo António, Porto, Portugal* ethics committee under number #2016-023 (020-DEFI/020-CES). All participants provided the written informed consent before enrollment in the study.

Consent for publication

Not applicable.

Data Availability

The authors state they have full control of all primary data and agree to allow the journal to review their data after reasonable request.

Acknowledgements

This work was supported by national funds through FCT Fundação para a Ciência e a Tecnologia, I.P., within the scope of the Cardiovascular R&D Center (UIDB/00051/2020 and UIDP/00051/2020) and RISE (LA/P/0053/2020).

References

1. Smolen JS, Aletaha D, Barton A, Burmester GR, Emery P, Firestein GS, Kavanaugh A, McInnes IB, Solomon DH, Strand V, et al. Rheumatoid arthritis. *Nat Rev Dis Primers*. 2018;4:18001. doi: 10.1038/nrdp.2018.1
2. Choy E, Ganeshalingam K, Semb AG, Szekanecz Z, Nurmohamed M. Cardiovascular risk in rheumatoid arthritis: recent advances in the understanding of the pivotal role of inflammation, risk predictors and the impact of treatment. *Rheumatology (Oxford)*. 2014;53:2143-2154. doi: 10.1093/rheumatology/keu224
3. Solomon DH, Karlson EW, Rimm EB, Cannuscio CC, Mandl LA, Manson JE, Stampfer MJ, Curhan GC. Cardiovascular morbidity and mortality in women diagnosed with rheumatoid arthritis. *Circulation*. 2003;107:1303-1307. doi: 10.1161/01.cir.0000054612.26458.b2
4. Bayes-Genis A, Liu PP, Lanfear DE, de Boer RA, González A, Thum T, Emdin M, Januzzi JL. Omics phenotyping in heart failure: the next frontier. *Eur Heart J*. 2020;41:3477-3484. doi: 10.1093/eurheartj/ehaa270
5. Kobayashi M, Huttin O, Magnusson M, Ferreira JP, Bozec E, Huby AC, Preud'homme G, Duarte K, Lamiral Z, Dalleau K, et al. Machine Learning-Derived Echocardiographic Phenotypes Predict Heart Failure Incidence in Asymptomatic Individuals. *JACC Cardiovasc Imaging*. 2021. doi: 10.1016/j.jcmg.2021.07.004
6. Woolley RJ, Ceelen D, Ouwkerk W, Tromp J, Figarska SM, Anker SD, Dickstein K, Filippatos G, Zannad F, Metra M, et al. Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction. *Eur J Heart Fail*. 2021;23:983-991. doi: 10.1002/ejhf.2144
7. Stienen S, Ferreira JP, Kobayashi M, Preud'homme G, Dobre D, Machu JL, Duarte K, Bresso E, Devignes MD, López N, et al. Enhanced clinical phenotyping by mechanistic bioprofiling in heart failure with preserved ejection fraction: insights from the MEDIA-DHF study (The Metabolic Road to Diastolic Heart Failure). *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*. 2020;25:201-211. doi: 10.1080/1354750x.2020.1727015
8. Sabbah MS, Fayyaz AU, Denus Sd, Felker GM, Borlaug BA, Dasari S, Carter RE, Redfield MM. Obese-Inflammatory Phenotypes in Heart Failure With Preserved Ejection Fraction. *Circulation: Heart Failure*. 2020;13:e006414. doi: doi:10.1161/CIRCHEARTFAILURE.119.006414
9. Ferreira MB, Fonseca T, Costa R, Marinhoc A, Carvalho HC, Oliveira JC, Zannad F, Rossignol P, Gottenberg JE, Saraiva FA, et al. Prevalence, risk factors and proteomic bioprofiles associated with heart failure in rheumatoid arthritis: The RA-HF study. *Eur J Intern Med*. 2020. doi: 10.1016/j.ejim.2020.11.002
10. Villeneuve E, Nam J, Emery P. 2010 ACR-EULAR classification criteria for rheumatoid arthritis. *Rev Bras Reumatol*. 2010;50:481-483.
11. Laboratories ATSCoPSfCPF. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002;166:111-117. doi: 10.1164/ajrccm.166.1.at1102
12. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, et al.

- Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging*. 2015;16:233-270. doi: 10.1093/ehjci/jev014
13. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr*. 2015;28:1-39 e14. doi: 10.1016/j.echo.2014.10.003
 14. Roubille C, Richer V, Starnino T, McCourt C, McFarlane A, Fleming P, Siu S, Kraft J, Lynde C, Pope J, et al. The effects of tumour necrosis factor inhibitors, methotrexate, non-steroidal anti-inflammatory drugs and corticosteroids on cardiovascular events in rheumatoid arthritis, psoriasis and psoriatic arthritis: a systematic review and meta-analysis. *Ann Rheum Dis*. 2015;74:480-489. doi: 10.1136/annrheumdis-2014-206624
 15. Blyszczuk P, Szekanecz Z. Pathogenesis of ischaemic and non-ischaemic heart diseases in rheumatoid arthritis. *RMD Open*. 2020;6. doi: 10.1136/rmdopen-2019-001032
 16. Rolski F, Blyszczuk P. Complexity of TNF-alpha Signaling in Heart Disease. *J Clin Med*. 2020;9. doi: 10.3390/jcm9103267
 17. Roodenrijs NMT, van der Goes MC, Welsing PMJ, Tekstra J, Lafeber F, Jacobs JWG, van Laar JM. Difficult-to-treat rheumatoid arthritis: contributing factors and burden of disease. *Rheumatology (Oxford)*. 2021;60:3778-3788. doi: 10.1093/rheumatology/keaa860
 18. Jung SM, Park KS, Kim KJ. Clinical phenotype with high risk for initiation of biologic therapy in rheumatoid arthritis: a data-driven cluster analysis. *Clin Exp Rheumatol*. 2021;39:1282-1290.
 19. Croft M, Siegel RM. Beyond TNF: TNF superfamily cytokines as targets for the treatment of rheumatic diseases. *Nat Rev Rheumatol*. 2017;13:217-233. doi: 10.1038/nrrheum.2017.22
 20. Wallentin L, Eriksson N, Olszowka M, Grammer TB, Hagstrom E, Held C, Kleber ME, Koenig W, Marz W, Stewart RAH, et al. Plasma proteins associated with cardiovascular death in patients with chronic coronary heart disease: A retrospective study. *PLoS Med*. 2021;18:e1003513. doi: 10.1371/journal.pmed.1003513
 21. Eggers KM, Lindhagen L, Baron T, Erlinge D, Hjort M, Jernberg T, Marko-Varga G, Rezelj M, Spaak J, Lindahl B. Predicting outcome in acute myocardial infarction: an analysis investigating 175 circulating biomarkers. *Eur Heart J Acute Cardiovasc Care*. 2021. doi: 10.1093/ehjacc/zuaa014
 22. Skau E, Henriksen E, Wagner P, Hedberg P, Siegbahn A, Leppert J. GDF-15 and TRAIL-R2 are powerful predictors of long-term mortality in patients with acute myocardial infarction. *Eur J Prev Cardiol*. 2017;24:1576-1583. doi: 10.1177/2047487317725017
 23. Lind L, Arnlov J, Sundstrom J. Plasma Protein Profile of Incident Myocardial Infarction, Ischemic Stroke, and Heart Failure in 2 Cohorts. *J Am Heart Assoc*. 2021;10:e017900. doi: 10.1161/JAHA.120.017900
 24. Hong Y, Xie W, Chen CS. [Effect of shengmai injection on TRAIL death receptor of patients with congestive heart failure]. *Zhongguo Zhong Xi Yi Jie He Za Zhi*. 2005;25:1092-1095.

25. Kotake S, Nanke Y, Mogi M, Kawamoto M, Furuya T, Yago T, Kobashigawa T, Togari A, Kamatani N. IFN-gamma-producing human T cells directly induce osteoclastogenesis from human monocytes via the expression of RANKL. *Eur J Immunol*. 2005;35:3353-3363. doi: 10.1002/eji.200526141
26. Ferreira JP, Verdonschot J, Collier T, Wang P, Pizard A, Bar C, Bjorkman J, Boccanelli A, Butler J, Clark A, et al. Proteomic Bioprofiles and Mechanistic Pathways of Progression to Heart Failure. *Circ Heart Fail*. 2019;12:e005897. doi: 10.1161/CIRCHEARTFAILURE.118.005897

Table 1. Characteristics of the study population according to identified clusters.

	Overall (N=399)	Cluster-1 (N=223)	Cluster-2 (N = 176)	p-value
Demographic data				
Age, years	61 (50,68)	54 (46, 64)	66 (59, 72)	<0.001
Female, N (%)	306 (77%)	185 (83%)	121 (69%)	<0.001
BMI, kg/m ²	25.9 (23.6, 29.4)	25.4 (23.1, 28.5)	26.8 (24.1, 30.1)	0.006
Waist circumference, cm	91 (82, 100)	87 (79, 96)	96 (88, 103)	<0.001
Medical history, N (%)				
Diabetes mellitus	53 (13%)	20 (9.0%)	33 (19%)	0.004
Hypertension	184 (46%)	71 (32%)	113 (64%)	<0.001
Dyslipidemia	190 (48%)	86 (39%)	104 (59%)	<0.001
Heart failure	24 (6.0%)	4 (1.8%)	20 (11%)	<0.001
Atrial fibrillation	14 (3.5%)	4 (1.8%)	10 (5.7%)	0.036
Ischemic stroke or TIA	14 (3.5%)	2 (0.9%)	12 (6.8%)	0.001
COPD	18 (4.5%)	7 (3.1%)	11 (6.2%)	0.14
RA history				
Time since diagnosis of RA, years	9 (5, 17)	7 (4, 15)	10 (5, 21)	0.005
RF or anti-CCP positive, N (%)	316 (79%)	173 (78%)	143 (81%)	0.40
Articular erosions, N (%)	139 (39%)	68 (33%)	71 (46%)	0.01
DAS28 VS (ESR)	2.58 (1.98, 3.44)	2.50 (1.96, 3.40)	2.78 (2.13, 3.50)	0.13
Symptoms and physical exam				
Dyspnea, N (%)	113 (28%)	47 (21%)	66 (38%)	<0.001
Fatigue, N (%)	242 (61%)	123 (55%)	119 (68%)	0.011
Systolic blood pressure, mmHg	132 (121, 145)	130 (118, 142)	134 (123, 149)	0.011
Heart rate, bpm	78 (70, 89)	77 (70, 87)	80 (72, 90)	0.05
Echocardiogram				
LVEF, %	61 (56, 66)	61 (57, 66)	62 (55, 66)	0.50
LVEF <50%, N (%)	17 (4.8%)	5 (2.5%)	12 (7.8%)	0.023
Laboratorial data				
Total cholesterol, mg/dl	184 (162, 208)	186 (163, 208)	183 (161, 208)	0.70
LDL cholesterol, mg/dl	98 (80, 119)	99 (81, 120)	98 (80, 116)	0.60
Hb, g/dl	13.20 (12.30, 14.20)	13.20 (12.40, 14.20)	13.20 (12.15, 14.20)	0.70
eGFR, ml/min/1.73m ²	91 (75, 102)	96 (87, 107)	78 (64, 94)	<0.001
NT-pro BNP, pg/ml	85 (44, 174)	64 (39, 114)	135 (66, 312)	<0.001
Troponin T, ng/ml	0.005 (0.002,	0.003 (0.002, 0.005)	0.008 (0.005, 0.012)	<0.001

	0.008)			
CRP, mg/dl	3 (1, 8)	3 (1, 7)	4 (1, 9)	0.021
Medication, N (%)				
ACEi/ARB	152 (38%)	54 (24%)	98 (56%)	<0.001
Beta-blocker	43 (11%)	10 (4.5%)	33 (19%)	<0.001
Loop diuretic	10 (2.5%)	2 (0.9%)	8 (4.5%)	0.025
Statin	150 (38%)	69 (31%)	81 (46%)	0.002
Corticosteroids	181 (45%)	91 (41%)	90 (51%)	0.04
Methotrexate	248 (62%)	142 (64%)	106 (61%)	0.50
NSAIDs	101 (25%)	60 (27%)	41 (23%)	0.40
Biological DMARDs	75 (19%)	43 (19%)	32 (18%)	0.80
Anti-TNF	43 (57%)	27 (64%)	16 (48%)	
Rituximab	13 (17%)	7 (17%)	6 (18%)	0.31
Tocilizumab	19 (25%)	8 (19%)	11 (33%)	
Six-minute walk test (m)	375 (330, 420)	390 (360, 450)	360 (270, 400)	<0.001

Values are expressed as mean \pm SD, n (%) or median (25th to 75th percentile)
 Bold face values indicate statistical significance.

BMI, body mass index; COPD, chronic obstructive pulmonary disease; RA, rheumatoid arthritis; RF, rheumatoid factor; Anti-CCP, anti-citrulline antibody; DAS28 VS (ESR), Disease Activity Score for Rheumatoid Arthritis with Erythrocyte Sedimentation Rate; LVEF, left ventricular ejection fraction; Hb, hemoglobin; eGFR, estimated Glomerular Filtration Rate; LDL, low-density lipoprotein; HDL, high-density lipoprotein; NT-proBNP, N-terminal pro B-type Natriuretic Peptide; CRP, c-reactive protein; ACEi/ARB, angiotensin converting enzyme inhibitors or angiotensin-receptor blockers; MRA, mineralocorticoid receptor antagonists; MTX, methotrexate; NSAIDs, nonsteroidal anti-inflammatory drugs; DMARDs, disease-modifying anti-rheumatic drugs; m, meters.

Note: Chi-square tests were used to compare proportions and Mann-Whitney tests to compare continuous variables (because they do not meet normality assumptions).

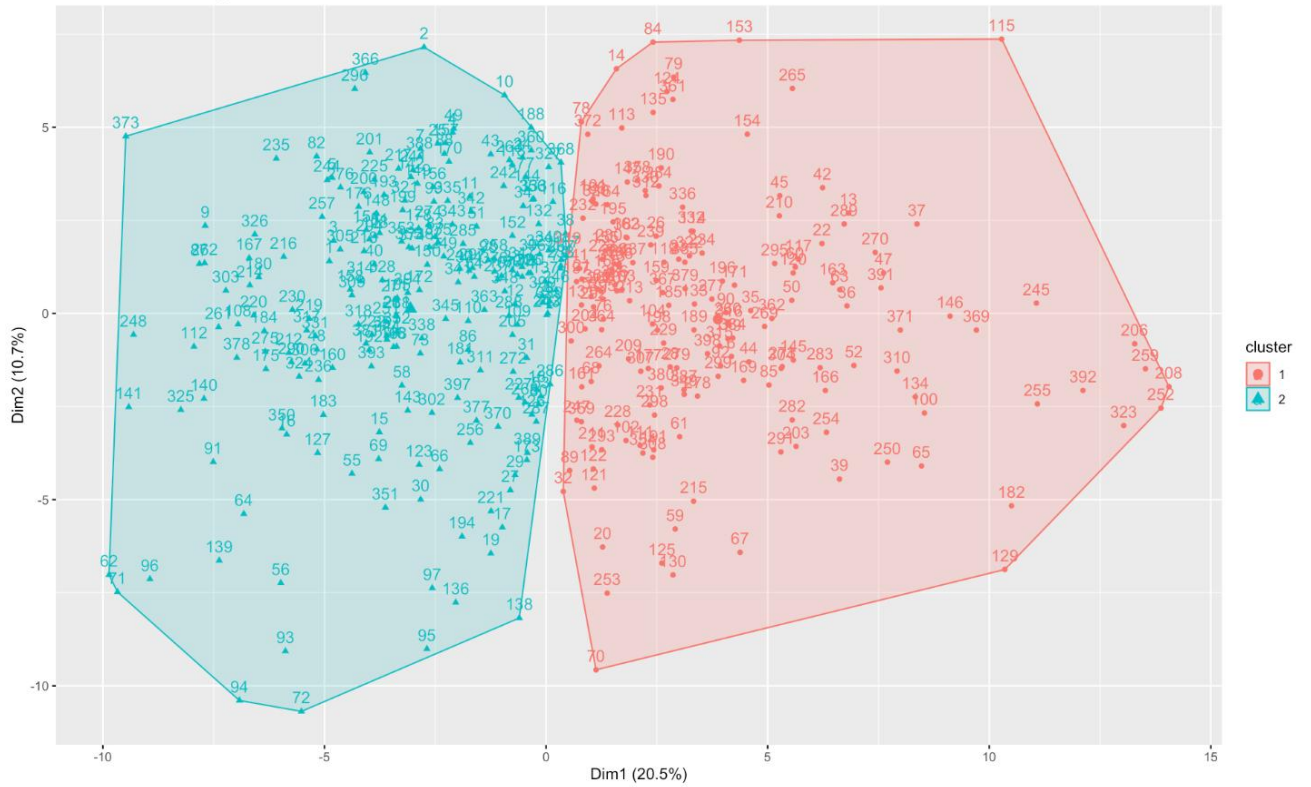
Table 2 -Confusion matrix comparing clustering all biomarkers (BM) and clustering with the top 2 most relevant biomarkers (2-BM): TRAILR2 and TNFRS11A.

	cluster 1 (2 BM)	cluster 2 (2 BM)
cluster 1 (94 BM)	214	1
cluster 2 (94 BM)	58	118

Accuracy = 0.83, 95% CI = [0.80, 0.87]

Figure 1. Factorial patients' distribution according to 2-clusters k-means partitioning.

KMEANS Clustering



Caption: Separation of patients into 2 independent and mutually exclusive groups based on K-means clustering, an unsupervised Machine Learning technique.

Legend: Dim, dimension.

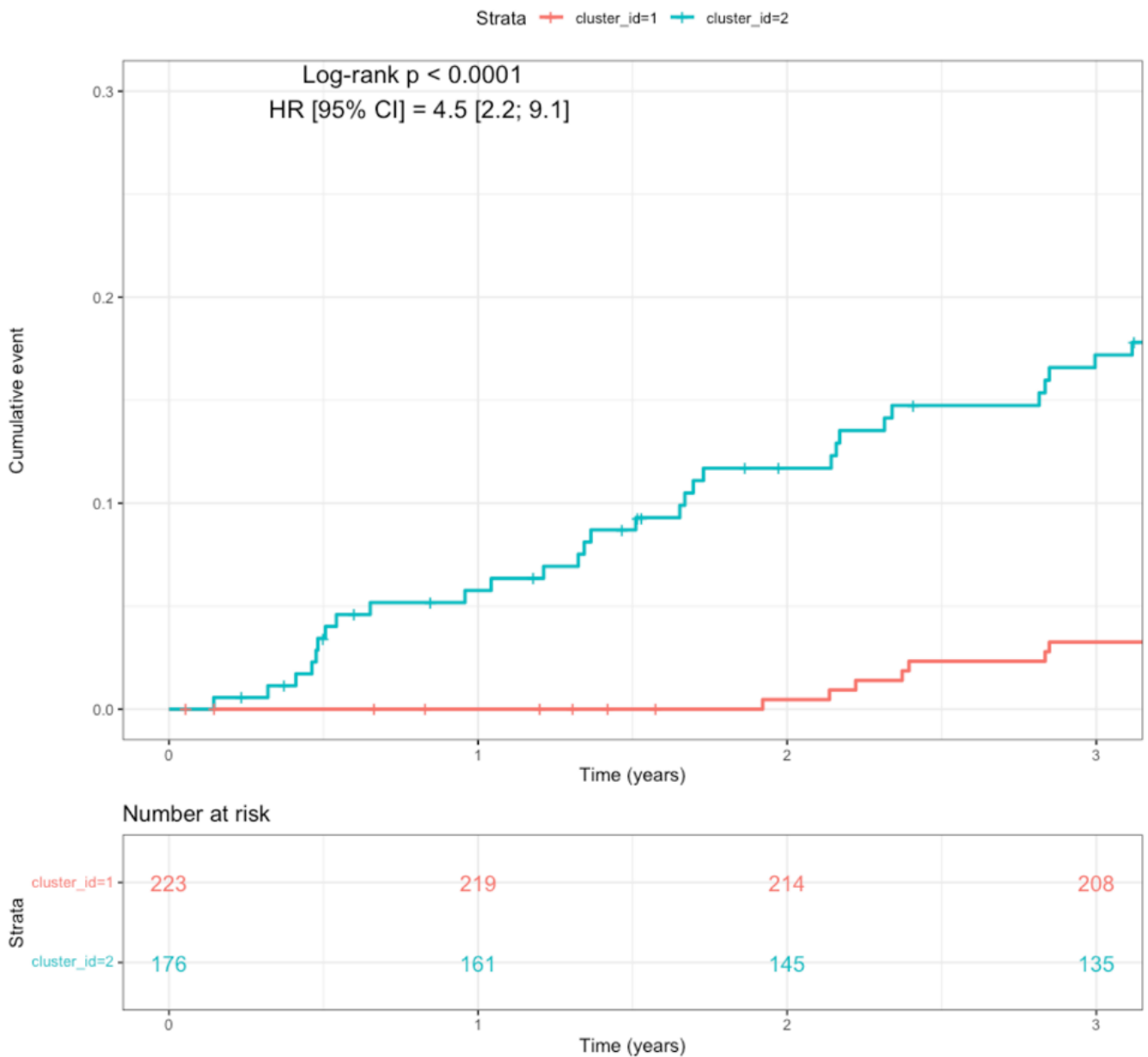
Figure 2. Factorial patients' distribution according to 2-clusters k-means partitioning



using TRAILR2 and TNFRSF11A protein biomarkers.

Caption: Separation of patients into 2 independent and mutually exclusive groups based on K-means clustering using only 2 biomarkers: TRAILR2 and TNFRSF11A. The results are superimposable to those shown in Figure 1 using 94 biomarkers.

Figure 3. Kaplan-Meier survival curve compares patients in cluster 1 and 2 for



cardiovascular death or cardiovascular hospitalization outcomes.

Caption: HR of 4.5, 95% CI [2.2; 9.1] for cluster 2.