



Proteomic Approaches to Defining Remission and the Risk of Relapse in Rheumatoid Arthritis

Liam J. O'Neil^{1,2*}, Pingzhao Hu^{3,4}, Qian Liu^{3,4}, Md. Mohaiminul Islam^{3,4}, Victor Spicer², Juergen Rech⁵, Axel Hueber⁵, Vidyanand Anaparti², Irene Smolik¹, Hani S. El-Gabalawy^{1,2}, Georg Schett^{5†} and John A. Wilkins^{1,2†}

OPEN ACCESS

Edited by:

Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:

Loredana Frasca,
National Institute of Health (ISS), Italy
Gillian Inara Rice,
The University of Manchester,
United Kingdom
Feiyang Ma,
University of California, Los Angeles,
United States

*Correspondence:

Liam J. O'Neil
liam.oneil@umanitoba.ca

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Autoimmune and Autoinflammatory
Disorders,
a section of the journal
Frontiers in Immunology

Received: 23 June 2021

Accepted: 20 October 2021

Published: 18 November 2021

Citation:

O'Neil LJ, Hu P, Liu Q, Islam MM,
Spicer V, Rech J, Hueber A,
Anaparti V, Smolik I, El-Gabalawy HS,
Schett G and Wilkins JA (2021)
Proteomic Approaches to Defining
Remission and the Risk of
Relapse in Rheumatoid Arthritis.
Front. Immunol. 12:729681.
doi: 10.3389/fimmu.2021.729681

¹ Section of Rheumatology, Department of Internal Medicine, University of Manitoba, Winnipeg, MB, Canada, ² Manitoba Centre for Proteomics and Systems Biology, University of Manitoba and Health Sciences Centre, Winnipeg, MB, Canada, ³ Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada, ⁴ Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada, ⁵ Department of Medicine, Friedrich-Alexander University Erlangen-Nuernberg and Universitaetsklinikum Erlangen, Erlangen, Germany

Objectives: Patients with Rheumatoid Arthritis (RA) are increasingly achieving stable disease remission, yet the mechanisms that govern ongoing clinical disease and subsequent risk of future flare are not well understood. We sought to identify serum proteomic alterations that dictate clinically important features of stable RA, and couple broad-based proteomics with machine learning to predict future flare.

Methods: We studied baseline serum samples from a cohort of stable RA patients (RETRO, n = 130) in clinical remission (DAS28 < 2.6) and quantified 1307 serum proteins using the SOMAscan platform. Unsupervised hierarchical clustering and supervised classification were applied to identify proteomic-driven clusters and model biomarkers that were associated with future disease flare after 12 months of follow-up and RA medication withdrawal. Network analysis was used to define pathways that were enriched in proteomic datasets.

Results: We defined 4 proteomic clusters, with one cluster (Cluster 4) displaying a lower mean DAS28 score ($p = 0.03$), with DAS28 associating with humoral immune responses and complement activation. Clustering did not clearly predict future risk of flare, however an XGboost machine learning algorithm classified patients who relapsed with an AUC (area under the receiver operating characteristic curve) of 0.80 using only baseline serum proteomics.

Conclusions: The serum proteome provides a rich dataset to understand stable RA and its clinical heterogeneity. Combining proteomics and machine learning may enable prediction of future RA disease flare in patients with RA who aim to withdrawal therapy.

Keywords: rheumatoid arthritis, disease activity, outcomes research, treatment, proteomics

HIGHLIGHTS

- Serum proteomics defines clinically relevant clusters within a cohort of stable RA patients
- Machine learning and proteomics may identify individuals at highest risk for future disease flare
- Despite meeting criteria for remission, clinically detectable disease is associated with a serum proteomic signature in stable RA

INTRODUCTION

Rheumatoid Arthritis (RA) is a systemic autoimmune disease that is characterized by inflammation of synovial joints (1). Modern RA therapy is initiated early and escalated aggressively using a treat-to-target approach to try to obtain disease remission (2). The development of both targeted treatments and combination regimens continues to improve expected outcomes for patients. Encouragingly, clinical remission, defined by multiple measures of disease activity (3), has become a realistic expectation for most patients with RA. Recent registry data of RA cohorts consistently show that DAS28 (Disease Activity Score) remission is achieved in about 50% of patients (4), a number that may be increasing over time (5).

Patients with RA who are able to achieve disease remission using standard therapy are not well studied, given their lack of disease activity and need for treatment changes. The main issue facing these patients is whether or not to remain on their treatment, or risk withdrawal and the potential for disease flare. There are many prospective studies that have demonstrated successful Disease Modifying Anti-Rheumatic Drugs (DMARD) withdrawal in patients in clinical remission (6–9) but the determinants of maintaining remission status after medication withdrawal are poorly defined (10). Unfortunately, given the limited understanding of the pathological mechanisms that drive subclinical disease, clinicians are left to guess which of their patients might sustain remission using less aggressive therapy.

Technological advances in high-throughput proteomics have allowed for an improved understanding of disease processes and biomarker discovery (11). Although mass spectrometry tends to dominate this evolving field, broad-based targeted proteomics has its own advantages, including simplified sample preparation and user-friendly output data (12). Our group has previously defined protein sets that are associated with future disease flare from pre-clinical RA by coupling machine learning with proteomic approaches (13). Indeed, leveraging omics approaches to resolve heterogeneity in common diseases remains a distinct challenge in clinical medicine (14), though this has not been systematically undertaken in a stable RA cohort.

The RETRO (15) (Reducing therapy in rheumatoid arthritis patients in ongoing remission) study is a prospective randomized trial which enrolled patients who had achieved disease remission

with conventional RA therapy. One of the aims of this study is to define disease recurrence in patients with RA when either continuing or reducing their medications. It was previously shown in this trial that positive anti-citrullinated antibody (ACPA), and other biomarkers (16, 17) are associated with an increased likelihood of disease relapse. In spite of these studies, there is little understanding of the underlying biological mechanisms that are active in stable RA. If differences within RA patients in remission can be more clearly defined, there may be an enhanced understanding of the spectrum of RA pathogenesis, along with improved personalized clinical approaches surrounding the withdrawal of therapy. We hypothesized that high-throughput proteomics (18) would help identify underlying biological heterogeneity that might provide insights into mechanisms underpinning future disease flare. Our aim was to explore how the serum proteome shapes the underlying clinical experiences of stable RA patients.

METHODS

Patients and Inclusion Criteria

RETRO is a multicentre, randomized, open, prospective, controlled parallel-group study. Details of the study are described in the original publication (15). The objective of the study is to evaluate tapering or discontinuation of DMARDs in patients with RA. All enrolled patients fulfilled the 2010 American College of Rheumatology (ACR) criteria for RA (19). Patients had to have sustained clinical remission defined by the Disease Activity score (DAS28 < 2.6) criteria for at least 6 months (20). Ethics committee of the Friedrich-Alexander-University of Erlangen-Nuremberg approval was granted.

Treatment and Follow-up

Patients were randomized to one of three arms: Arm 1 continued with existing DMARD regimen at full dose for 12 months, arm 2 reduced the dose of all DMARDs by 50%, while arm 3 reduced the dose of all DMARDs by 50% in the first 6 months, then discontinued all medications. Relapse of disease was defined as a DAS28-ESR score greater than 2.6. Participants were assessed for clinical disease activity every 3 months until month 12.

Assessment of Demographic and Disease-Specific Parameters

Age and sex were recorded in all patients. Disease duration, tender joint count (68), swollen joint count (66), patient visual analogue scale (VAS) for pain and patient global were assessed and recorded. C-reactive protein (CRP), ESR, Rheumatoid Factor (RF), ACPA, DAS28-ESR and Health Assessment Questionnaire (HAQ-DI) were recorded.

SOMAscan

SOMAscan is a proteomics assay that measures 1307 proteins using an aptamer library. This high-throughput proteomics assay has been used in recent publications to study the aging proteome (21, 22) along with other human diseases (23). 130 baseline

serum samples were available from the RETRO study. Briefly, a library of aptamers were incubated with serum, and those that bind are isolated and hybridized to DNA microarray for detection. The identity and relative concentration of the detected proteins are revealed by localization and fluorescence intensity. Protein quantification is reported as relative fluorescence units (RFU), an arbitrary value. In general, agreement between aptamer and antibody-based assays is high (24). Further details regarding the SOMAscan assay are available (18).

Statistical Analysis

Descriptive results (Table 1) are stated in means and standard deviation. SOMAmer protein expression RFU values for the study patients were transformed into a log₂ scale for differential analysis (Supplemental File 2). Batch effect was removed in our SOMAmer data using internal controls within each plate to adjust proteomic intensity as per standard SOMAscan protocols. Batch effect was assessed between plates and determined to require no further correction. Data was loaded and analyzed in the R (v3.5.3) environment unless otherwise stated. Missing clinical data was imputed using multiple imputation by chained equations (MICE) (25). Differential analysis between groups was undertaken using linear modeling with the package *LIMMA* (26). GO pathways analysis was performed using *clusterprofiler* (27). Graphs were generated using the *ggplot2* package. Correlation analyses were performed for select proteins using Pearson correlation. Multi-dimensional scaling (MDS) was used for dimension reduction on all SOMAscan proteins.

The 200 most variable proteins measured by coefficient of variation were used to determine optimal number of clusters ranging from $k = 2$ to 10 and identify sample clusters using the R package *Consensusclusterplus*. We used 80% protein resampling and 80% patient resampling and selected Pearson as our distance function. Multinomial logistic regression implemented in R package *glmnet* was used to identify clinical variables that are independently associated with cluster assignment. Sliding window analysis of DAS28 scores and protein expression was

performed using a previously published algorithm, DE-SWAN (28). Briefly, this algorithm analyzed serum protein expression across quintiles of DAS28 scores using linear modeling, while adjusting for baseline demographic factors, in this case age and sex. A protein expression score was developed on proteins that correlated with DAS28 which were identified by DE-SWAN. We filtered the proteomic data on the 34 score members, scaled the data by the mean and standard deviation, and multiplied by 1 (positively associated with DAS28) or -1 (negatively associated with DAS28) for each protein. The final score was the mean expression of all 34 proteins for each patient. We randomly generated 5000 data sets with 34 randomly selected proteins in each set to evaluate the significance of the association score.

Machine Learning Classification Algorithm

We applied two supervised machine learning (ML) techniques to develop algorithms to classify flare or remission based on serum proteomics. The first approach we used is XGBoost (Extreme Gradient Boosting), which employs a regularization term to overcome the overfitting (29). The second approach is the LASSO model (30), which was used as a baseline to compare its performance with that of XGBoost. Data was loaded into Python, and samples were randomly split into a training ($n = 104$, 80% of the samples) and test ($n = 26$) set. The training set was used to train and tune the parameters in the two models and the test set was used evaluate the models' performance, which were measured by the area under of the curve (AUC) of receiver operating characteristic (ROC), accuracy, sensitivity and specificity. To increase the interpretability of the XGBoost model to predict the flare status of a given sample, we used SHAP values (Shapley Additive Explanation) (31). A higher SHAP value of a given feature in the model represents its strong influence on the model output. The final model parameters we used in the XGBoost are as follows: `learning_rate = 0.01`, `max_depth = 3`, `subsample = 0.6`, `colsample_bytree = 0.7`, `n_estimators = 100`, `gamma = 0.0`, `reg_alpha = 0.5`, the parameters used in the LASSO model is as follow: `cost=1.17` and `max_iterations = 5000`. We used 5-fold cross-validation to get the optimal hyperparameters.

TABLE 1 | Baseline characteristics of the patients, split by proteomic cluster.

Characteristics	Total (n = 130)	Cluster 1 (n = 34)	Cluster 2 (n = 12)	Cluster 3 (n = 46)	Cluster 4 (n = 38)
Age	55.2 (13.1)	52.7 (14.6)	54.1 (13.7)	54.7 (11.7)	58.6 (13.1)
Females, %	56.2%	67.6%	66.7%	56.5%	42.1%
Disease Duration (years)	6.8 (7.0)	7.9 (6.5)	8.6 (9.3)	6.9 (6.3)	4.9 (7.3)
DAS-28 (ESR)	1.7 (0.68)	1.93 (0.60)	1.73 (0.74)	1.71 (0.65)	1.51 (0.71)
ACR/EULAR remission, %	76.6%	67.7%	66.6%	88.9%	72.9%
HAQ, units	0.12 (0.32)	0.11 (0.17)	0.08 (0.12)	0.16 (0.46)	0.09 (0.26)
Positive RF, %	56.2%	73.5%	50.0%	52.2%	47.3%
Positive ACPA, %	57.7%	67.6%	66.7%	55.6%	50.0%
*Biological DMARD use, % (N)	40.0%	38.2%	25.0%	47.8%	36.8%
Flare, %	37.7%	38.2%	16.7%	43.5%	36.8%

ACPA, anticitrullinated protein antibody; ACR, American College of Rheumatology; CRP, C-Reactive protein; DAS-28, disease activity score-28 (based on ESR); DMARDs, disease modifying antirheumatic drugs; ESR, erythrocyte sedimentation rate; EULAR, European League Against Rheumatism; HAQ, Health Assessment Questionnaire; RF, Rheumatoid Factor; VAS, Visual analogue scale.

*Tumor necrosis factor inhibitors and tocilizumab.

Study Cohort

Baseline characteristics for 130 patients enrolled in the RETRO study are found in **Table 1**. Overall, the group had maintained clinical remission for 16.6 (16.2) months and mean disease duration of over 6 years. 57.7% of the patients were ACPA positive, while 40.0% required biologics to achieve remission. 76.6% of patients had achieved the most stringent definition of remission [ACR/EULAR remission (32)]. After 12 months of follow-up, 62.3% of the overall population remained in clinical remission (50% in those undergoing withdrawal).

RESULTS

Hierarchical Clustering on Serum Proteins Identifies Heterogeneity Amongst Stable RA Patients

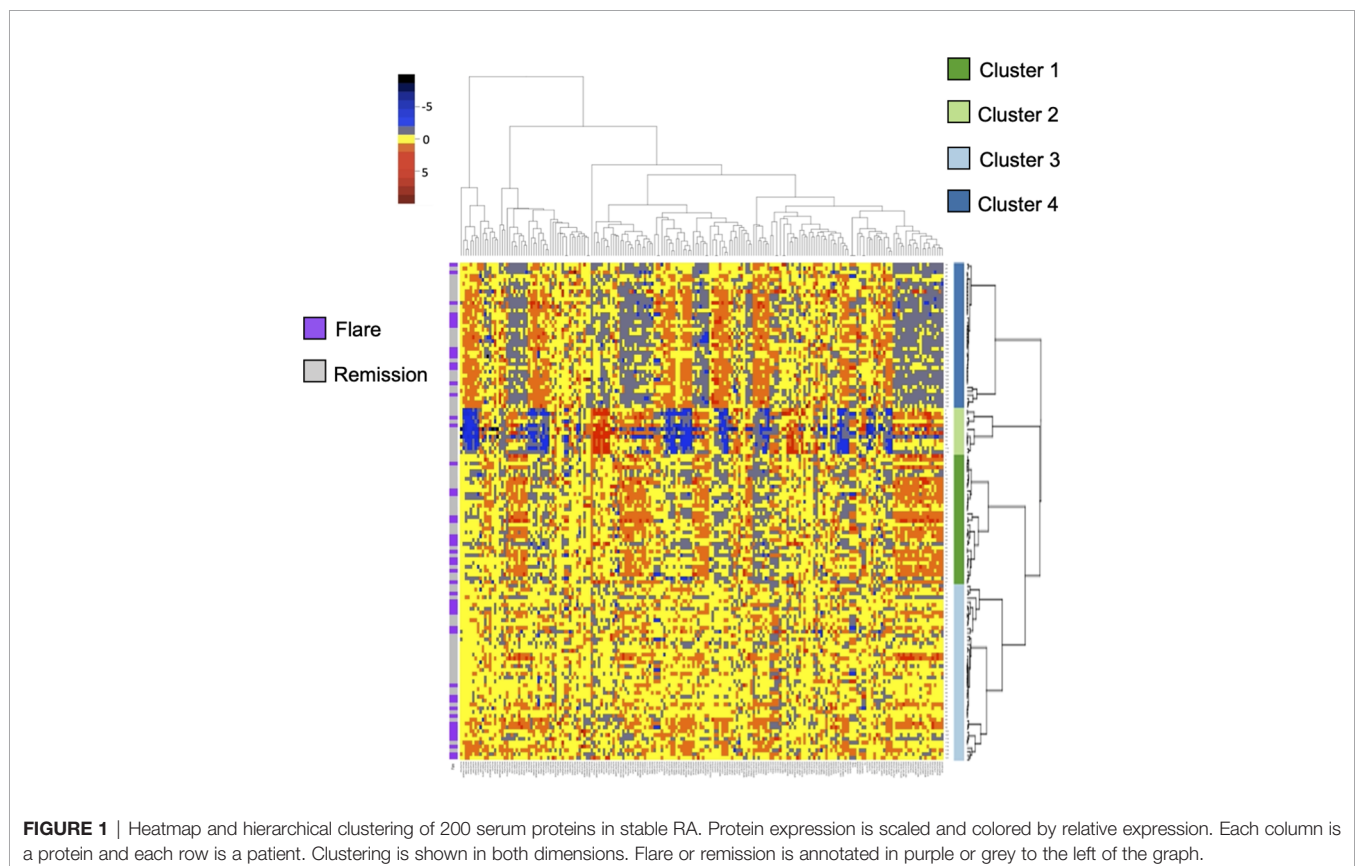
Given the paucity of data aimed at understanding subclinical disease activity in RA patients who achieve remission, we sought to explore underlying heterogeneity using serum proteomics in this established cohort. We quantified over 1300 serum proteins from 130 RETRO patients at their baseline visit, all of whom were in stable clinical remission ($\text{DAS28} < 2.6$). We hypothesized that despite the clinical similarities amongst individuals within this cohort, proteomic differences may provide important insights by identifying sub-clusters of patients. We applied consensus clustering to assign individuals to one of the

4 clusters (**Figure S1**) and clustered scaled protein expression by hierarchical clustering, which can be seen in **Figure 1**. MDS analysis revealed separation of the hierarchical clusters (**Figure S2**).

Baseline characteristics split by cluster are listed in **Table 1**. We found no differences in sex, age, biologic use, or serological status across our 4 proteomic clusters. Cluster 4 had significantly lower DAS28 scores compared to the remaining clusters. With respects to future flare, Cluster 2 trended toward lower rates relative to the remaining clusters, however this did not reach statistical significance (16.7% vs 39.8%, $p = 0.21$). To assess this association by multinomial regression, we assigned Cluster 2 as the reference cluster and found that that Cluster 3 had higher odds of flare (OR 5.6, 0.97 to 33.06, $p = 0.05$), relative to Cluster 2 with similar trends observed for Cluster 1 and Cluster 4 (**Table S1**). Indeed, no clear distinction between individuals who developed future flare was observed in the MDS plot (**Figure S2**). Overall, these results suggest that global proteomic clusters within a clinically homogenous cohort can be identified but are associated with current clinical status rather than future outcomes.

Machine Learning Classifies Future Flare Using Baseline Serum Proteomics

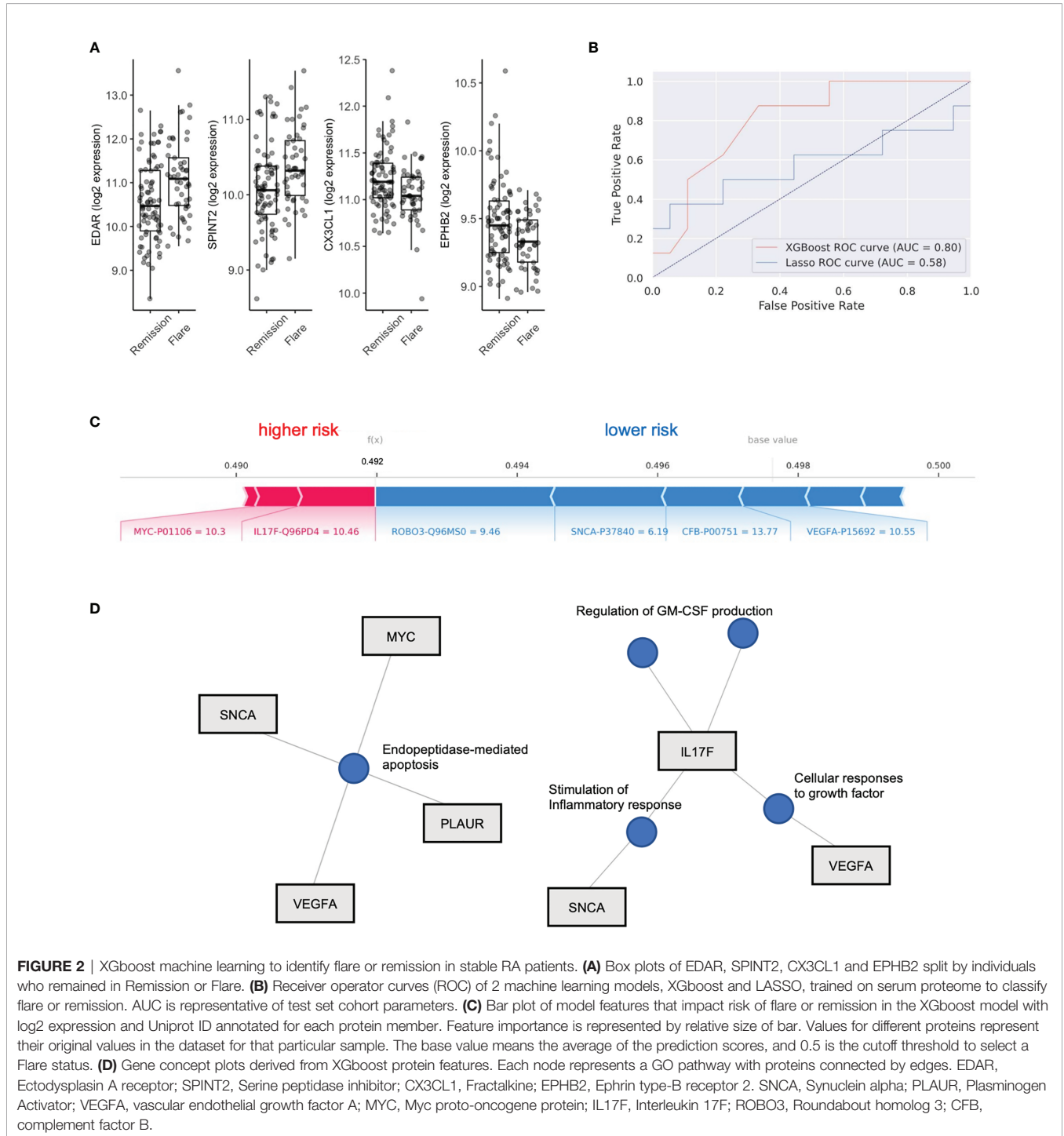
We next aimed to use the serum proteome to identify biomarkers associated with future disease flare in stable RA, given that clustering did not clearly associate with risk of flare. We identified DEP's between these groups (**Table S2**) and observed



upregulation in Ectodysplasin A receptor (EDAR, FC = 1.20) and Serine peptidase inhibitor (SPINT2, FC 1.1), and downregulation of Fractalkine (CX3CL1, FC 0.95) and Ephrin type-B receptor 2 (EPHB2, FC 0.95) in individuals who eventually went on to flare (**Figure 2A**). However, after adjustment for multiple comparisons, none of the differentially expressed proteins reached statistical significance. This suggests that although subtle differences exist in

the serum proteome between individuals who experience future flare, it's unlikely that singular biomarkers accurately predict this outcome in this population.

To test this hypothesis, we explored the use of two machine learning algorithms, LASSO and XGBoost, to build predictive models that classify future flare using baseline serum proteomics. We generated two models, both of which were validated on a test



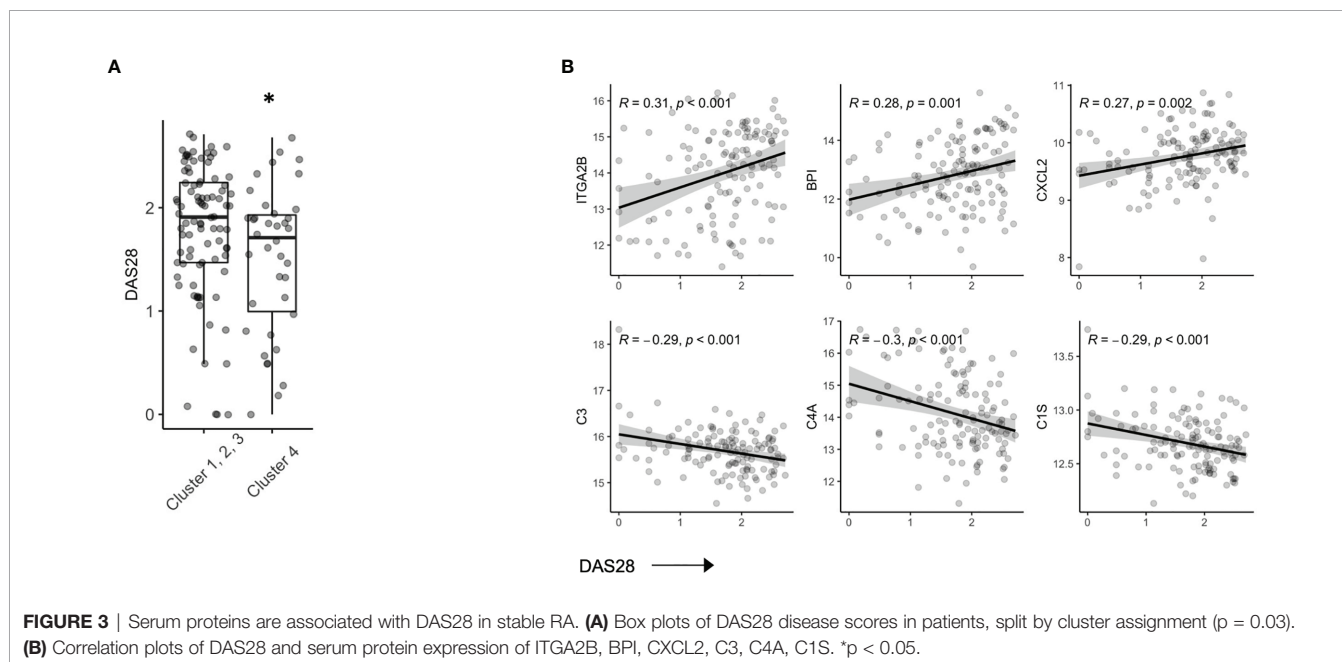
cohort (20% of total cohort). The LASSO model achieved 69.2% accuracy, with an AUC of 0.58, along with sensitivity of 0.5 and specificity of 0.78 based on the test cohort (Features in **Table S3**). XGboost delivered a model with higher specificity (0.78) than sensitivity (0.63) and an overall accuracy of 73.1% with an AUC 0.80. Therefore, we found that the XGboost model outperformed the LASSO model by the metric area under the curve (**Figure 2B**, AUC, 0.80 vs 0.58), accuracy (73.1% vs 69.2%) and sensitivity (0.63 vs 0.5).

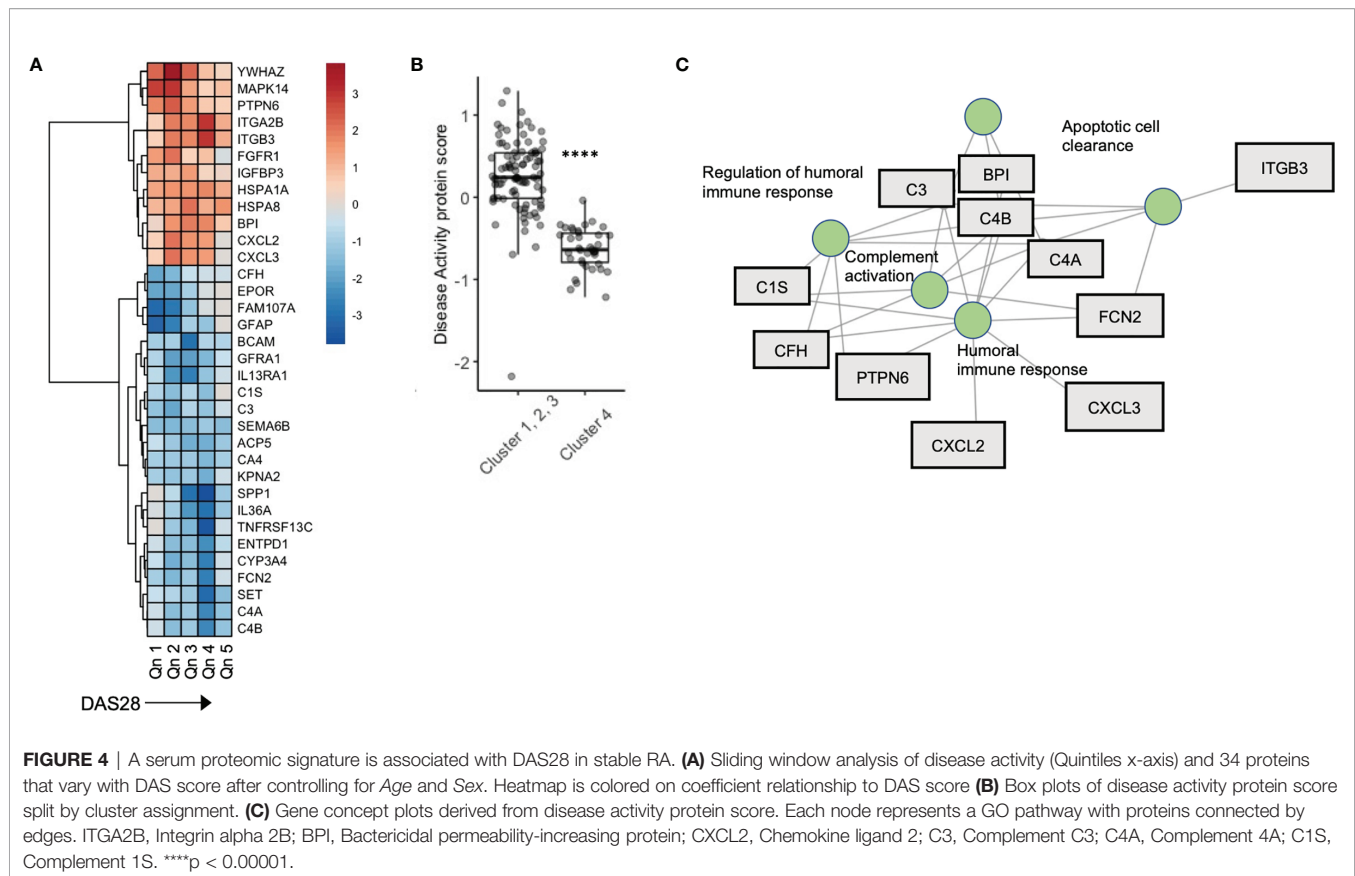
To interpret this XGBoost model, we explored the impact of essential features in terms of SHAP values on the classifier's output for a single prediction which are shown in **Figure 2C**. We identified Interleukin 17F (IL17F) and Myc proto-oncogene protein (MYC) expression as indicators of future flare, while Roundabout homolog 3 (ROBO3), Synuclein alpha (SNCA), complement factor B (CFB) and vascular endothelial growth factor A (VEGF-A) expression were indicators of sustained remission (**Figure 2C** and **Figure S3**). Given the small number of proteins that derived our boosted model, we next explored whether there were any functional links between these proteins. We developed gene concept plots to identify potential protein interactions and found that SNCA, MYC, VEGFA and PLAUR were connected by a single pathway, *endopeptidase mediated apoptosis*. We then analyzed IL17F restricted networks, given its conflicting role in RA (33–36), and that its expression was associated with future flare in our model. We found that IL17F interacted with VEGFA through *growth factor* function, and with SNCA through common effects on the *inflammatory response* (**Figure 2D**). IL17F independently regulated *GM-CSF production*, a key driver of RA disease activity through the recruitment of neutrophils (37). Overall, this network analysis suggests that cellular apoptosis and GM-CSF production may be associated with future disease flares in RA patients who are otherwise stable.

Disease Activity in Stable RA Is Reflected in the Serum Proteome

In our hierarchical clustering, we observed a lower mean DAS28 score in Cluster 4 compared to the remaining 3 clusters (**Figure 3A**). RA patients who achieve DAS28 defined remission often have residual disease activity, however, since this population is not typically the focus of translational studies, little is known regarding biomarkers that are reflective of ongoing disease activity. Indeed, we found that several protein members correlated with DAS28 score (**Figure S4**), including Integrin alpha 2B (ITGA2B), Bactericidal permeability-increasing protein (BPI) and chemokine ligand 2 (CXCL2, Pearson R, all p value < 0.01). Further, complement proteins (C3, C4A, C1S) were all negatively associated with DAS28 score, suggesting activation and consumption of complement proteins (**Figure 3B**) were indicators of disease. There was no indication that these parameters varied based on ACPA status (**Figure S5**).

To further explore the relationship between disease activity and serum biomarkers, we developed a sliding window model (SWAN) which examined protein variability across DAS28 quintiles, after controlling for Age and Sex. Across DAS28 a total of 34 proteins varied significantly with disease activity (**Figure 4A**). We used these 34 protein members to annotate a meta-protein expression score (the mean expression profiles of the 34 proteins), which correlated with DAS28 ($R = 0.45$, $p < 0.001$, **Figure S6**). To test the robustness of this finding, we sampled 5000 random sets with 34 proteins in each set and correlated their mean expression with DAS28 scores. We found a range of $-0.37 - 0.27$, associated with a low probability (0) that the correlation of 0.45 would occur by chance (**Figure S7**). This protein disease activity score was significantly lower in Cluster 4 compared to the remaining 3 clusters (**Figure 4B**), concordant with their lower DAS28 scores. Gene concept plots revealed that





these 34 proteins interacted through nodes that included *humoral immunity*, *apoptosis*, and *complement activation* (Figure 4C). Overall, these data suggest that stable RA disease activity is marked by complement consumption and humoral immune responses, which is reflected in a serum protein signature that is detectable in patients with stable RA.

DISCUSSION

With the advent of multiple targeted therapies, alone and in combination, most RA patients should reasonably expect to achieve low disease activity or clinical remission status. To date, few studies have sought to understand the heterogeneity of the biological pathways underpinning clinical remission in this rapidly expanding population of RA patients. When, and in whom to attempt withdrawal of therapy has become a compelling clinical question. On the one hand, there is justifiable concern regarding reactivation of systemic and articular inflammation. On the other hand, ongoing use of DMARDs and/or biologics is associated with increased risk of infectious complications (38), malignancy (39) and cost (40). Strategies for successful taper however remain ill-defined and, importantly, lack precision (6, 41, 42). The RETRO clinical trial has previously generated predictive models that were based on clinical parameters and serum studies (16, 17). ACPA seropositivity appears to be an important indicator for

increased risk of future relapse (17), while clinical parameters have modest predictive value even when combined with advanced machine learning techniques (43). It remains unclear if these indicators are clinically applicable and generalizable to a wide range of RA patient populations. The focus of this study was to use a broad-based serum proteomic approach to better understand the underlying heterogeneity amongst RA patients who are in sustained clinical remission, prior to their participation in a clinical trial of therapy withdrawal. Our results identify proteomic signatures reflecting biological mechanisms that are associated with ongoing disease stability off therapy, or alternatively, the risk of future disease relapse.

Our XGboost model suggested that individual circulating serum biomarkers are unlikely, on their own, to be predictive of future stability or relapse after therapy withdrawal. In spite of this, combinations of proteomic biomarkers identified by the machine learning achieved relatively high AUC and accuracy in predicting outcomes. Indeed, this is a testament to the power of rapidly evolving machine learning algorithms that are being developed for many clinical problems (44). Network analysis of proteins derived from machine learning suggested that inflammatory forms of cellular death was an indicator for risk future disease flare. Indeed, apoptosis is escaped by pathogenic fibroblast-like synoviocytes and likely contributes to their aggressive and hyperplastic phenotype in RA (45) and this may point to systemic FLS as a potential source of these proteins (46). Hierarchical clustering identified proteomic

clusters which were defined, in part, by clinical characteristics. Cluster 4 represented a patient group with lower DAS28 scores amongst the remaining patient cohort. Our disease activity signature, found to be lower in Cluster 4, suggested that elements of humoral immunity and complement activation might facilitate disease activity in otherwise stable RA patients. This suggests that activating pathways differentially regulate disease activity in this subset of RA patients, as many of the well-known disease activity markers in RA suggest that innate immune responses associate with DAS28 scores (47–49).

The analyses we undertook utilized the SOMAscan aptamer-based technology to interrogate 1307 distinct serum proteins. Although this provided us with a robust array of biomarkers, it is well recognized that these represent only a fraction of the human proteome, and that larger arrays that span a larger proportion of the circulating proteome may help generate even more accurate predictive algorithms. Moreover, there remains an incomplete understanding of how aptamer-based detection of each individual analyte correlates with other detection methodologies such as those that are antibody based (24). Due to our modest sample size, we observed strongly statistically significant association although R values related to DAS28 scores were all below 0.3. However, we expect the R values may increase using a larger sample size while the significant association will be still held. Finally, SOMA proteins may bias over-representation analysis based on the selected proteins which are included in the set. Notably, network analysis was used in this study to connect proteins of interest through biological nodes. These results would not be impacted by inherent bias in the SOMA protein set.

In conclusion, we applied an unsupervised, high-throughput proteomics assay to delineate biomarkers and pathways that reflect the biological heterogeneity present in RA patients who are collectively deemed to be in stable clinical remission. Based on this, we used supervised machine learning to develop robust models that predicted ongoing disease stability after therapy withdrawal as opposed to future disease flare. Although it is premature to try and define the potential clinical utility of these models, they do provide an important impetus for further studies that aim to further define a biological definition of remission in RA patients that can ultimately guide clinical decision making.

REFERENCES

- McInnes IB, Schett G. The Pathogenesis of Rheumatoid Arthritis. *N Engl J Med* (2011) 365:2205–19. doi: 10.1056/NEJMra1004965
- Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating Rheumatoid Arthritis to Target: 2014 Update of the Recommendations of an International Task Force. *Ann Rheum Dis* (2016) 75:3–15. doi: 10.1136/annrheumdis-2015-207524
- Dougados M, Aletaha D, van Riel PL. Disease Activity Measures for Rheumatoid Arthritis. *Clin Exp Rheumatol* (2007) 25:S22–9.
- Combe B, Rinciveau N, Benessiano J, Berenbaum F, Cantagrel A, Daures JP, et al. Five-Year Favorable Outcome of Patients With Early Rheumatoid Arthritis in the 2000s: Data From the ESPOIR Cohort. *J Rheumatol* (2013) 40:1650–7. doi: 10.3899/jrheum.121515
- Aga AB, Lie E, Uhlig T, Olsen IC, Wierod A, Kalstad S, et al. Time Trends in Disease Activity, Response and Remission Rates in Rheumatoid Arthritis During the Past Decade: Results From the NOR-DMARD Study 2000–2010. *Ann Rheum Dis* (2015) 74:381–8. doi: 10.1136/annrheumdis-2013-204020

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Friedrich-Alexander-University of Erlangen-Nuremberg. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LO'N: data analysis, manuscript construction. PH: data analysis. QL: data analysis. MI: data analysis. VS: data analysis. JR: data collection and clinical trial. AH: data collection and clinical trial. VA: manuscript, data analysis. IS: manuscript, data analysis. HE-G: manuscript, data analysis. GS: data collection and clinical trial. JW: oversight, manuscript, data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

The authors above have no relevant financial disclosures or benefits from commercial sources that could create a potential conflict of interest. The entirety of this work was funded by a grant obtained by JA Wilkins through the Canadian Foundation for Innovation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.729681/full#supplementary-material>

- Galvao TF, Zimmermann IR, da Mota LM, Silva MT, Pereira MG. Withdrawal of Biologic Agents in Rheumatoid Arthritis: A Systematic Review and Meta-Analysis. *Clin Rheumatol* (2016) 35:1659–68. doi: 10.1007/s10067-016-3285-y
- Ghiti Moghadam M, Vonkeman HE, Ten Klooster PM, Tekstra J, van Schaardenburg D, Starman-Kool M, et al. Stopping Tumor Necrosis Factor Inhibitor Treatment in Patients With Established Rheumatoid Arthritis in Remission or With Stable Low Disease Activity: A Pragmatic Multicenter, Open-Label Randomized Controlled Trial. *Arthritis Rheumatol* (2016) 68:1810–7. doi: 10.1002/art.39626
- Smolen JS, Nash P, Durez P, Hall S, Ilivanova E, Irazoque-Palazuelos F, et al. Maintenance, Reduction, or Withdrawal of Etanercept After Treatment With Etanercept and Methotrexate in Patients With Moderate Rheumatoid Arthritis (PRESERVE): A Randomised Controlled Trial. *Lancet* (2013) 381:918–29. doi: 10.1016/S0140-6736(12)61811-X
- Emery P, Hammoudeh M, FitzGerald O, Combe B, Martin-Mola E, Buch MH, et al. Sustained Remission With Etanercept Tapering in Early Rheumatoid Arthritis. *N Engl J Med* (2014) 371:1781–92. doi: 10.1056/NEJMoa1316133

10. Tweehuysen L, van den Ende C, Beeren F, Been E, van den Hoogen F, den Broeder A. Little Evidence for Usefulness of Biomarkers for Predicting Successful Dose Reduction or Discontinuation of a Biologic Agent in Rheumatoid Arthritis. *Arthritis Rheum* (2017) 69:301–8. doi: 10.1002/art.39946
11. Boja E, Hiltke T, Rivers R, Kinsinger C, Rahbar A, Mesri M, et al. Evolution of Clinical Proteomics and Its Role in Medicine. *J Proteome Res* (2011) 10:66–84. doi: 10.1021/pr100532g
12. Moaddel R, Ubaida-Mohien C, Tanaka T, Lyashkov A, Basisty N, Schilling B, et al. Proteomics in Aging Research: A Roadmap to Clinical, Translational Research. *Aging Cell* (2021) 20:e13325. doi: 10.1111/acel.13325
13. O'Neil LJ, Spicer V, Smolik I, Meng X, Goel RR, Anaparti V, et al. Association of a Serum Protein Signature With Rheumatoid Arthritis Development. *Arthritis Rheumatol* (2021) 73:78–88. doi: 10.1002/art.41483
14. Olivier M, Asmis R, Hawkins GA, Howard TD, Cox LA. The Need for Multi-Omics Biomarker Signatures in Precision Medicine. *Int J Mol Sci* (2019) 20:4781. doi: 10.3390/ijms20194781
15. Haschka J, Englbrecht M, Hueber AJ, Manger B, Kleyer A, Reiser M, et al. Relapse Rates in Patients With Rheumatoid Arthritis in Stable Remission Tapering or Stopping Antirheumatic Therapy: Interim Results From the Prospective Randomised Controlled RETRO Study. *Ann Rheum Dis* (2016) 75:45–51. doi: 10.1136/annrheumdis-2014-206439
16. Rech J, Hueber AJ, Finzel S, Englbrecht M, Haschka J, Manger B, et al. Prediction of Disease Relapses by Multi-biomarker Disease Activity and Autoantibody Status in Patients With Rheumatoid Arthritis on Tapering DMARD Treatment. *Ann Rheum Dis* (2016) 75:1637–44. doi: 10.1136/annrheumdis-2015-207900
17. Figueiredo CP, Bang H, Cobra JF, Englbrecht M, Hueber AJ, Haschka J, et al. Antimodified Protein Antibody Response Pattern Influences the Risk for Disease Relapse in Patients With Rheumatoid Arthritis Tapering Disease Modifying Antirheumatic Drugs. *Ann Rheum Dis* (2017) 76:399–407. doi: 10.1136/annrheumdis-2016-209297
18. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* (2010) 5:e15004. doi: 10.1371/journal.pone.0015004
19. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 Rheumatoid Arthritis Classification Criteria: An American College of Rheumatology/European League Against Rheumatism Collaborative Initiative. *Arthritis Rheum* (2010) 62:2569–81. doi: 10.1002/art.27584
20. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism Provisional Definition of Remission in Rheumatoid Arthritis for Clinical Trials. *Ann Rheum Dis* (2011) 70:404–13. doi: 10.1136/ard.2011.149765
21. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, et al. Plasma Proteomic Signature of Age in Healthy Humans. *Aging Cell* (2018) 17:e12799. doi: 10.1111/acel.12799
22. Sathyan S, Ayers E, Gao T, Weiss EF, Milman S, Verghese J, et al. Plasma Proteomic Profile of Age, Health Span, and All-Cause Mortality in Older Adults. *Aging Cell* (2020) 19:e13250. doi: 10.1111/acel.13250
23. Govaere O, Cockell S, Tiniakos D, Queen R, Younes R, Vacca M, et al. Transcriptomic Profiling Across the Nonalcoholic Fatty Liver Disease Spectrum Reveals Gene Signatures for Steatohepatitis and Fibrosis. *Sci Transl Med* (2020) 12:eaba4448. doi: 10.1126/scitranslmed.aba4448
24. Lollo B, Steele F, Gold L. Beyond Antibodies: New Affinity Reagents to Unlock the Proteome. *Proteomics* (2014) 14:638–44. doi: 10.1002/pmic.201300187
25. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple Imputation by Chained Equations: What Is it and How Does it Work? *Int J Methods Psychiatr Res* (2011) 20:40–9. doi: 10.1002/mpr.329
26. 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) 43:e47. doi: 10.1093/nar/gkv007
27. Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
28. Lehallier B, Gate D, Schaum N, Nanasi T, Lee SE, Yousef H, et al. Undulating Changes in Human Plasma Proteome Profiles Across the Lifespan. *Nat Med* (2019) 25:1843–50. doi: 10.1038/s41591-019-0673-2
29. Chen C, Zhang Q, Yu B, Yu Z, Lawrence PJ, Ma Q, et al. Improving Protein-Protein Interactions Prediction Accuracy Using XGBoost Feature Selection and Stacked Ensemble Classifier. *Comput Biol Med* (2020) 123:103899. doi: 10.1016/j.combiomed.2020.103899
30. Lee TF, Chao PJ, Ting HM, Chang L, Huang YJ, Wu JM, et al. Using Multivariate Regression Model With Least Absolute Shrinkage and Selection Operator (LASSO) to Predict the Incidence of Xerostomia After Intensity-Modulated Radiotherapy for Head and Neck Cancer. *PLoS One* (2014) 9:e89700. doi: 10.1371/journal.pone.0089700
31. Lundberg SM EG, Lee S. Consistent Individualized Feature Attribution for Tree Ensembles. (2018) arXiv2019. <https://arxiv.org/abs/1802.03888>
32. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism Provisional Definition of Remission in Rheumatoid Arthritis for Clinical Trials. *Arthritis Rheum* (2011) 63:573–86. doi: 10.1002/art.30129
33. Chen S, Blijdorp IC, van Mens LJJ, Bowcutt R, Latuhihin TE, van de Sande MGH, et al. Interleukin 17A and IL-17F Expression and Functional Responses in Rheumatoid Arthritis and Peripheral Spondyloarthritis. *J Rheumatol* (2020) 47:1606–13. doi: 10.3899/jrheum.190571
34. Robert M, Miossec P. IL-17 in Rheumatoid Arthritis and Precision Medicine: From Synovitis Expression to Circulating Bioactive Levels. *Front Med (Lausanne)* (2018) 5:364. doi: 10.3389/fmed.2018.00364
35. Marwa OS, Kalthoum T, Wajih K, Kamel H. Association of IL17A and IL17F Genes With Rheumatoid Arthritis Disease and the Impact of Genetic Polymorphisms on Response to Treatment. *Immunol Lett* (2017) 183:24–36. doi: 10.1016/j.imlet.2017.01.013
36. Taams LS. Interleukin-17 in Rheumatoid Arthritis: Trials and Tribulations. *J Exp Med* (2020) 217:e20192048. doi: 10.1084/jem.20192048
37. Liu R, Lauridsen HM, Amezcua RA, Pierce RW, Jane-Wit D, Fang C, et al. IL-17 Promotes Neutrophil-Mediated Immunity by Activating Microvascular Pericytes and Not Endothelium. *J Immunol* (2016) 197:2400–8. doi: 10.4049/jimmunol.1600138
38. Wallis D. Infection Risk and Biologics: Current Update. *Curr Opin Rheumatol* (2014) 26:404–9. doi: 10.1097/BOR.0000000000000072
39. Maneiro JR, Souto A, Gomez-Reino JJ. Risks of Malignancies Related to Tofacitinib and Biological Drugs in Rheumatoid Arthritis: Systematic Review, Meta-Analysis, and Network Meta-Analysis. *Semin Arthritis Rheum* (2017) 47:149–56. doi: 10.1016/j.semarthrit.2017.02.007
40. Barnabe C, Zheng Y, Ohinmaa A, Crane L, White T, Hemmelgarn B, et al. Effectiveness, Complications, and Costs of Rheumatoid Arthritis Treatment With Biologics in Alberta: Experience of Indigenous and Non-Indigenous Patients. *J Rheumatol* (2018) 10:1344–52. doi: 10.3899/jrheum.170779
41. Subesinghe S, Scott IC. Key Findings From Studies of Methotrexate Tapering and Withdrawal in Rheumatoid Arthritis. *Expert Rev Clin Pharmacol* (2015) 8:751–60. doi: 10.1586/17512433.2015.1077698
42. Smolen JS, Aletaha D. Rheumatoid Arthritis Therapy Reappraisal: Strategies, Opportunities and Challenges. *Nat Rev Rheumatol* (2015) 11:276–89. doi: 10.1038/nrrheum.2015.8
43. Vodencarevic A, Tascilar K, Hartmann F, Reiser M, Hueber AJ, Haschka J, et al. Advanced Machine Learning for Predicting Individual Risk of Flares in Rheumatoid Arthritis Patients Tapering Biologic Drugs. *Arthritis Res Ther* (2021) 23:67. doi: 10.1186/s13075-021-02439-5
44. Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, Oskotsky B, et al. Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. *JAMA Netw Open* (2019) 2:e190606. doi: 10.1001/jamanetworkopen.2019.0606
45. Korb A, Pavenstadt H, Pap T. Cell Death in Rheumatoid Arthritis. *Apoptosis* (2009) 14:447–54. doi: 10.1007/s10495-009-0317-y
46. Orange DE, Yao V, Sawicka K, Fak J, Frank MO, Parveen S, et al. RNA Identification of PRIME Cells Predicting Rheumatoid Arthritis Flares. *N Engl J Med* (2020) 383:218–28. doi: 10.1056/NEJMoa2004114
47. Hammer HB, Fagerhol MK, Wien TN, Kvien TK. The Soluble Biomarker Calprotectin (an S100 Protein) Is Associated to Ultrasonographic Synovitis Scores and Is Sensitive to Change in Patients With Rheumatoid Arthritis Treated With Adalimumab. *Arthritis Res Ther* (2011) 13:R178. doi: 10.1186/ar3503

48. Chen YS, Yan W, Geczy CL, Brown MA, Thomas R. Serum Levels of Soluble Receptor for Advanced Glycation End Products and of S100 Proteins Are Associated With Inflammatory, Autoantibody, and Classical Risk Markers of Joint and Vascular Damage in Rheumatoid Arthritis. *Arthritis Res Ther* (2009) 11:R39. doi: 10.1186/ar2645
49. Bach M, Moon J, Moore R, Pan T, Nelson JL, Lood C. A Neutrophil Activation Biomarker Panel in Prognosis and Monitoring of Patients With Rheumatoid Arthritis. *Arthritis Rheumatol* (2020) 72:47–56. doi: 10.1002/art.41062

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 O'Neil, Hu, Liu, Islam, Spicer, Rech, Hueber, Anaparti, Smolik, El-Gabalawy, Schett and Wilkins. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.