Supplementary Methods

Collection of Blood Samples and Evaluation of Clinical and Biological Parameters

Whole blood from the subjects was collected by direct venipuncture into a test tube containing sodium citrate as an anticoagulant, or into a specific test tube for serum collection. A total volume of 40 mL was drawn from each RA patient. To minimize variations in blood composition due to diet or circadian rhythm, all samples were collected in the early morning following an 8-hour fasting period.

Blood samples were processed to isolate serum by centrifugation at 3,500 rpm for 10 minutes at room temperature, within 4 hours of collection. Serum was transferred to new RNase-free tubes and either used immediately for *ex vivo* studies or stored at −80°C until subsequent analysis. For the longitudinal paired samples, baseline and 6-month serum samples from each patient were processed, stored, and handled under identical conditions. Samples used for proteomic and 3D synovial experiments underwent equivalent freeze–thaw procedures. Serum storage duration ranged from approximately 24 to 48 months.

PBMCs were isolated using Ficoll-Paque gradient (Lymphoprep, Axis-Shield, Dundee, UK). Initially, 10 mL of blood was mixed with an equal volume of phosphate-buffered saline (PBS, pH 7.4) and carefully overlaid on an equal volume of Ficoll medium. That was centrifuged at 2,500 rpm for 20 minutes at 20°C. The PBMC layer was aspirated and washed with PBS by centrifugation at 1,800 rpm for 5 minutes at 4°C.

In addition, neutrophils were also obtained from the Ficoll-Paque gradient using specific cell separation commercial kits (Miltenyi Biotech) and the autoMACS Pro Separator (Miltenyi Biotech, Bergisch Gladbach, Germany), following a previously established protocol[1].

Proliferation assay

PBMCs and neutrophils proliferation was assessed using the Cell Proliferation Reagent WST-1 (4-[3-(4-iodophenyl)-2-(4-nitrophenyl)-2H-5-tetrazolio]-1, 3-benzene disulfonate) (F. Hoffmann-La Roche Ltd, Basel, Switzerland). WST-1 is a colorimetric, non-radioactive assay that quantifies cell proliferation based on metabolic activity.

The assay was performed according to the manufacturer's instructions. Briefly, 1.5×10^5 cells were seeded in each well of a 96-well plate with 100 µL of phenol red-free RPMI 1640 medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) to minimize interference during spectrophotometric measurements. A well containing only the medium, without cells, was used as a blank control. Subsequently, 10 µL of WST-1 reagent was added to each well (1:10 dilution), and the plate was incubated at 37°C in a humidified atmosphere with 5% CO_2 for 1 hour.

Absorbance was measured at 450 nm using a BioTek-PowerWave XS plate reader (BioTek Instruments, Inc., Winooski, VT, USA), with a reference wavelength of 690 nm to correct for background noise. Changes in cell proliferation mediated by JAKi and TNFi were determined by comparing WST-1 activity to RA autologous serum-treated controls, which were defined as 100%. Each experiment was performed in triplicate to ensure reproducibility and reliability.

Adhesion assay

PBMCs and neutrophils adhesion assays were performed using the Vibrant™ Cell Adhesion Assay Kit (Molecular Probes, Eugene, OR, USA) to assess changes in cell adhesion after JAKi and TNFi *ex vivo* treatment. Briefly, 3.0×10^5 cells were harvested and labeled with calcein AM (5 µM per cell) for 30 minutes. After washing with phenol red-free RPMI 1640 medium (Gibco,

Thermo Fisher Scientific, Waltham, MA, USA), the cell suspension was plated immediately into a 96-well black/clear bottom plate. A well containing only the medium (without cells) was used as a blank. After 90 minutes, non-adherent calcein-labeled cells were removed by gently washing twice with phenol red-free RPMI 1640 medium.

Adherent cells were quantified using a fluorescence microplate reader (TECAN GENious plus, Tecan Group Ltd., Männedorf, Switzerland) at 485 nm excitation and 535 nm emission. JAKi/TNFi-mediated changes in cell adhesion were determined by comparing the calcein AM signal to RA autologous serum-treated controls, defined as 100%. All experiments were performed in triplicate.

Quantitative real-time PCR

Selected genes involved in the proliferation and adhesion of PBMCs and neutrophils treated *ex vivo* with JAKi and TNFi were analyzed using quantitative real-time RT-PCR (Supplementary Table S10), with GAPDH as the housekeeping gene and the QuantStudio 7 Flex Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA).

mRNA expression levels were normalized to GAPDH and quantified using the 2-ΔΔCt method, as described elsewhere.

NETosis derived products

NETosis-derived products were quantified in the supernatants from *ex vivo* experiments. Nucleosomes were measured using the Cell Death Detection ELISAplus kit for human cell death detection (Roche, Basel, Switzerland), and citrullinated histone H3 levels were quantified using the Citrullinated Histone H3 ELISA (Cayman Chemical, Ann Arbor, MI, USA) following the manufacturer's recommendations.

Signal detection from both assays were detected spectrophotometrically using a VANTAstar plate reader (BMG lab-tech, Ortenberg, Germany).

Proteome assay in supernatants and serum samples

PEA technology provides high sensitivity, specificity, and multiplexing capacity while requiring minimal sample volume. PEA technology combines an antibody-based immunoassay with the powerful properties of qPCR. This method relies on a dual-recognition immunoassay, where two matched antibodies labelled with unique DNA oligonucleotides simultaneously bind to a target protein in solution. This brings the two antibodies into proximity, allowing their DNA oligonucleotides to hybridize, serving as template for a DNA polymerase-dependent extension step. This creates a double-stranded DNA “barcode” which is unique for the specific antigen and quantitatively proportional to the initial concentration of target protein. The hybridization and extension are immediately followed by PCR amplification.

The final concentration of the PCR product is directly proportional to the initial concentration of the target protein. Relative protein levels were expressed on a Log2-based normalized protein expression (NPX) scale. To ensure data quality, samples were randomly distributed across two plates, maintaining the proportional representation of study groups. Each 96-plex panel contained 96 assays, including four internal controls for systematic quality control. Additionally, interplate controls were incorporated to compute NPX values, along with negative controls and a duplicate external control. Intensity normalization was performed to align the median NPX of each protein across plates with the overall median, ensuring consistency and reliability in the analysis. During this process, the data were adjusted so that the median NPX for each protein on each plate matched the global median. This adjustment further enhanced the reliability and consistency of the analysis across all assays.

Machine-learning modelling and model comparison

To evaluate the ability of the *ex vivo* proteomic profile to discriminate treatment-specific response patterns, three supervised learning approaches with different levels of model complexity were compared: logistic regression (LR), random forest (RF), and a feed-forward neural network (NN). The treatment-specific outcome comprised two response profiles: remission under JAKi or non-remission under TNFi (JAKi-response/TNFi-non-response profile) and non-remission under JAKi or remission under TNFi (JAKi-non-response/TNFi-response profile). All models were developed using the same set of 50 candidate proteins. Treatment group was not included as a predictor.

Nested cross-validation and feature selection

Model development and evaluation were performed using repeated nested stratified cross-validation. The outer validation procedure consisted of 10-fold cross-validation repeated 20 times. At each repetition, patients were partitioned into 10 stratified folds. Each fold was used once as the test set, while the remaining nine folds constituted the corresponding training set.

Feature selection and model optimization were performed exclusively within the training data of each outer fold. An inner 3-fold stratified cross-validation was conducted within each outer training set. Candidate proteins were ranked according to their association with the treatment-specific response profile using Welch's t-tests calculated exclusively within the corresponding inner-training partition. Models including the 3, 5, 7, 9, 12, 15, or 20 top-ranked proteins were evaluated, and the number yielding the highest mean validation AUC across the inner folds was selected. In case of ties, the more parsimonious configuration was preferred.

Once the optimal number of proteins had been determined, protein ranking was recalculated using the complete outer-training dataset, and the corresponding top-ranked proteins were retained. The final model was fitted using the outer-training set and subsequently evaluated on the held-out outer-test set. Thus, patients in the outer-test set were not involved in feature selection, preprocessing, or model optimization.

For logistic regression and neural network models, protein values were standardized using the mean and standard deviation estimated exclusively from the corresponding training partition, and the same parameters were subsequently applied to the validation or test data.

Comparison of predictive algorithms

Logistic regression was fitted using maximum-likelihood estimation, with the number of included proteins optimized within the inner cross-validation. Model convergence and the absence of non-finite coefficients were verified.

A feed-forward neural network was evaluated as a nonlinear alternative. The number of selected proteins, hidden-layer size (1, 2, 3, or 5 neurons), and weight-decay parameter (0.01, 0.1, or 0.5) were jointly optimized within the inner cross-validation. Neural networks were trained using a maximum of 200 iterations.

Random forest was evaluated as a nonlinear ensemble approach. The number of selected proteins, the number of variables randomly considered at each split (mtry: 1, 2, 3, 4, 5, 7, 9, 12, 15, or 20, restricted to the number of selected proteins), and minimum node size (1, 5, or 10) were optimized within the inner cross-validation. Random forests were fitted using the Gini splitting rule, with 500 trees during inner optimization and 1,000 trees for the final outer-fold models.

The same outer and inner cross-validation partitions and nested feature-selection strategy were used for all three algorithms, allowing direct comparison of their predictive performance.

Model performance

For each outer fold, predicted probabilities were obtained for patients in the held-out test set. Predictions from the 10 outer folds were combined to generate one out-of-fold prediction for each patient within each repetition. Model discrimination was primarily assessed using the area under the receiver operating characteristic curve (AUC).

Across the 20 repetitions, AUC values were summarized using the mean, median, standard deviation, 2.5th and 97.5th percentiles, and minimum and maximum values. Classification performance was additionally assessed using balanced accuracy, sensitivity, specificity, and overall accuracy at a probability threshold of 0.50. The JAKi-response/TNFi-non-response profile (remission under JAKi or non-remission under TNFi) was considered the positive class, whereas the JAKi-non-response/TNFi-response profile (non-remission under JAKi or remission under TNFi) was considered the reference class. Classification metrics were calculated separately for each repetition and summarized as mean (SD) across the 20 repetitions.

Mean ROC curves were generated from the complete out-of-fold predictions of each repetition. ROC curves were linearly interpolated onto a common false-positive-rate grid, and the mean true-positive rate was calculated at each point. Variability across the 20 repetitions was represented by the 2.5th–97.5th percentile range.

Finally, protein-selection stability was assessed separately for each algorithm by calculating the frequency with which each protein was selected across the 200 outer training partitions (10 folds × 20 repetitions).

All analyses were performed in R version 4.5.1 using the *caret* (version 7.0-1), *pROC* (version 1.19.0.1), *nnet* (version 7.3-20), and *ranger* (version 0.18.0) packages. Random seeds were fixed throughout the analysis to ensure reproducibility, with a seed of 982 used to generate the outer cross-validation partitions.

Supplementary references

1. De La Rosa IA, Perez-Sanchez C, Ruiz-Limon P, Patiño-Trives A, Torres-Granados C, Jimenez-Gomez Y, et al. Impaired microRNA processing in neutrophils from rheumatoid arthritis patients confers their pathogenic profile. Modulation by biological therapies. haematol.; 2020;105:2250–61. https://doi.org/10.3324/haematol.2018.205047