

A longitudinal cohort study uncovers plasma protein biomarkers predating clinical onset and treatment response of rheumatoid arthritis

Received: 25 March 2024

Accepted: 9 July 2025

Published online: 21 July 2025



Siyu He^{1,8}, Chenxi Zhu^{1,2,8}, Yi Liu^{1,8}, Zhiqiang Xu^{1,3,8}, Rui Sun^{1,8}, Bin Yang⁴, Xin Guo¹, Martin Herrmann^{1,5}, Luis E. Muñoz⁵, Inger Gjerdtsson⁶, Rikard Holmdahl⁷, Lunzhi Dai¹✉ & Yi Zhao^{1,2}✉

Rheumatoid arthritis (RA) is a systemic inflammatory condition posing challenges in identifying biomarkers for onset, severity and treatment responses. Here we investigate the plasma proteome in a longitudinal cohort of 278 RA patients, alongside 60 at-risk individuals and 99 healthy controls. We observe distinct proteome signatures in at-risk individuals and RA patients, with protein levels alterations correlating with disease activity, notably at DAS28-CRP thresholds of 3.1, 3.8 and 5.0. The combination of methotrexate (MTX) and leflunomide (LEF) modulates proinflammatory pathways, whereas MTX plus hydroxychloroquine (HCQ) impact energy metabolism. A machine-learning model is trained for predicting responses, and achieves average receiver operating characteristic (ROC) scores of 0.88 (MTX + LEF) and 0.82 (MTX + HCQ) in the testing sets. The efficiency of these models is further validated in independent cohorts using enzyme-linked immunosorbent assay data. Overall, our study unveils distinct plasma proteome signatures across various stages and subtypes of RA, providing valuable biomarkers for predicting disease onset and treatment responses.

Rheumatoid arthritis (RA) is a persistent and progressive bundle of systemic inflammatory diseases mainly affecting joints. There are considerable challenges in understanding its etiology, subtype heterogeneity, diagnostic biomarkers, and optimal treatment targets^{1–3} with different pathogenic characteristics at different disease stages⁴. A crucial prelude to RA onset is the “at-risk” phase, characterized by

elevated autoantibody levels^{5,6}. Studying this at-risk phase is essential for unraveling the disease’s developmental nuances and identifying interventions to prevent or mitigate its impact.

The management of RA pivots on a protocolized treat-to-target strategy, where conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) play a central role¹. A substantial proportion

¹Department of Rheumatology and Immunology and National Clinical Research Center for Geriatrics, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, China. ²Department of Rheumatology and Immunology, Clinical Institute of Inflammation and Immunology, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, Sichuan, China. ³Tianfu Jincheng Laboratory, Chengdu, China. ⁴Department of Laboratory Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan, China. ⁵Department for Internal Medicine 3, University Hospital Erlangen, and Deutsches Zentrum für Immuntherapie; Friedrich-Alexander Universität Erlangen-Nürnberg (FAU), Erlangen, Germany. ⁶Department of Rheumatology and Inflammation Research, Institute for Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. ⁷Section of Medical Inflammation Research, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden. ⁸These authors contributed equally: Siyu He, Chenxi Zhu, Yi Liu, Zhiqiang Xu, Rui Sun. ✉e-mail: lunzhi.dai@scu.edu.cn; zhaoyi-rheuma@wchscu.cn

of patients (30–60%) exhibit suboptimal responses to csDMARDs combinations⁷. Previous studies have investigated the influence of clinical parameters such as sex, disease duration, disease activity, and rheumatoid factor levels on the prediction of patient response to csDMARDs^{8,9}. Additionally, a range of clinical measures, including ultrasound, T-cell subset, and patient-reported outcome measures, have been utilized to predict sustained remission rates for patients treated with csDMARDs¹⁰.

Plasma proteomics has emerged as a powerful and promising tool for assessing human health and disease conditions^{11–16}. However, the majority of current proteomic studies in RA predominantly employ cross-sectional designs to identify disease risk factors or biomarkers^{15,16}. There is a growing need for longitudinal studies to investigate the clinical onset and treatment response in RA patients via omics strategies. Unfortunately, progress in this area has been impeded by restricted cohort sizes, underscoring the critical requirement for large-scale cohort studies^{17–20}.

In this study, we compare the plasma proteomic profiles of healthy persons, at-risk individuals, and RA patients, identifying key protein patterns associated with disease progression and anti-citrullinated peptide autoantibodies (ACPAs) status. We further monitor RA patients longitudinally under csDMARDs treatment and uncover distinct protein markers predictive of therapeutic response to methotrexate (MTX) combined with leflunomide (LEF) or hydroxychloroquine (HCQ). These findings support the development of protein-based tools for early disease monitoring and treatment optimization in RA.

Results

Proteomic characterization in at-risk, ACPA-positive and ACPA-negative RA individuals

We recruited 278 RA patients from the western region of China; among them, 231 were females (83%) (Fig. 1a, b and Supplementary Data 1). The average age of RA patients was 51 years, ranging from 16 to 77 years. The disease activity score in 28 joints with C-reactive protein (DAS28-CRP) varied from 1.24 to 8.39, with an average value of 3.53. The ACPA-negative individuals were slightly older, with an average age of 52 (*vs* 51 for ACPA-positive RA patients) and lower DAS28-CRP scores of 3.07 (*vs* 3.71 for ACPA-positive RA patients) (Supplementary Table 1). Patients included in the study had not received csDMARDs treatment for at least 6 months prior to the collection of plasma samples. Among the 206 RA patients with follow-up data, 140 had one follow-up sample at 3–6 months, and 59 had two follow-up samples at 6–9 months after receiving MTX monotherapy or csDMARDs combination treatments. In addition, we recruited 60 at-risk individuals, 38 of whom were followed up for 5–7 years, and 99 healthy controls for comparative analysis. The average age of healthy controls was 51 years, with a range of 38–76 years (79 females and 20 males). The average age of 60 at-risk individuals was 48 years (32 females and 28 males), ranging from 29 to 74 years (Supplementary Table 1).

Next, we performed tandem mass tag (TMT)-based proteomics analysis of these plasma samples (Fig. 1a and Supplementary Data 2). Correlation analysis of quality control samples (Supplementary Fig. 1a), common reference samples (Supplementary Fig. 1b), and replicate samples revealed the high quality of our mass spectrometry (MS) data (Supplementary Fig. 1c). The observed stability in the distribution of normalized protein abundance also indicated minimal batch effects (Fig. 1c). A total of 2504, 2022, and 1924 proteins were identified from RA, at-risk individuals and healthy individuals, respectively (Fig. 1d). Proteins quantified in more than 50% of the samples in each group of individuals, totaling 996 plasma proteins, were used for the subsequent data analysis.

Plasma proteome fluctuations from health to RA onset

We initially performed hierarchical clustering on plasma proteome data from 182 ACPA-positive RA, 67 ACPA-negative RA, 60 at-risk

individuals, and 99 healthy controls (Fig. 2a), revealing clear distinctions between these groups (Fig. 2b). Comparative analyses identified a number of differentially expressed proteins (DEPs) and pathways between ACPA-positive RA patients, ACPA-negative RA patients, at-risk individuals and healthy controls (two-sided Student's *t* test, $p < 0.05$) (Fig. 2c and Supplementary Fig. 1d,e). Then, we combined proteins that differed between healthy and other groups and performed pathway enrichment analysis (Fig. 2c). This analysis revealed the upregulation of proteins associated with neutrophil degranulation, cellular stress responses, and cross-presentation of soluble exogenous antigens in both ACPA-positive RA patients and at-risk individuals^{21–23}. However, ACPA-positive RA patients presented more intense immune and acute-phase responses (Fig. 2c). In contrast, the downregulated proteins were primarily involved in metabolic dysregulation, redox processes such as hydrogen peroxide catabolism, and protein processing, suggesting increased endoplasmic reticulum stress^{24–26}. Notably, proteins specifically elevated in at-risk individuals were linked to RNA metabolism, which is recognized for its connection to inflammation²⁷. Additionally, we observed the upregulation of ROBO receptor signaling, which inhibits osteogenic differentiation, and axon guidance pathways, both of which are known to be upregulated in RA^{28,29} (Fig. 2c).

The differences in proteome profiles between ACPA-positive and ACPA-negative RA patients remain poorly understood, despite variations in clinical characteristics, disease progression, and treatment response. We observed a stronger inflammatory response in ACPA-positive RA patients, which remained significant even after adjusting for the DAS28-CRP between the two subsets. These findings suggest that increased inflammation is an intrinsic effect of the ACPA-positive phenotype, independent of disease activity (Supplementary Fig. 1f).

Autoimmune disorders may share common pathogenic mechanisms. Therefore, we studied whether the top enriched proteins in RA patients also showed abnormal expression in patients with other autoimmune diseases, including primary Sjögren's syndrome, systemic sclerosis, idiopathic inflammatory myopathy, and systemic lupus erythematosus, compared with healthy controls. The results confirmed the RA specificity of these DEPs, as most did not significantly differ between the patients with other autoimmune diseases and healthy controls (Supplementary Fig. 1g).

Age and sex can significantly impact proteome analysis³⁰. However, we did not observe significant age differences between healthy controls, at-risk individuals, and RA patients (Supplementary Fig. 1h). Consequently, we analyzed proteomic differences stratified by sex (Supplementary Fig. 1d, e). Compared with those in the other groups, most DEPs in the ACPA-positive RA group were consistent regardless of sex, showing a common trend of increased neutrophil degranulation, complement cascade regulation, and acute-phase response. Compared with RA patients, at-risk individuals presented higher levels of ROBO receptor signaling and RNA metabolism, with RNA metabolism being more elevated in males. In contrast, axon guidance was more pronounced in male RA patients, indicating increased bone remodeling pressure³¹. ACPA-negative RA patients exhibited a distinct increase in lipid metabolism, with elevated fatty acid β -oxidation specifically in males (Supplementary Fig. 1e).

The preclinical phase of RA is a crucial period for identifying pathogenic mechanisms and potential prevention targets. We followed up 38 at-risk individuals, of whom 8 developed RA (converters). These converters exhibited significantly lower complement component levels, suggesting depletion due to immune complex formation during the transition to RA³². Additionally, metabolism-related proteins such as PSMB7 were upregulated, indicating immunoproteasome activation³³ (Fig. 2d, e). For 3 of these converters, we collected plasma samples to compare proteomic differences before and after disease onset (Fig. 2f). Commonly identified proteins between converters and non-converters, as well as before

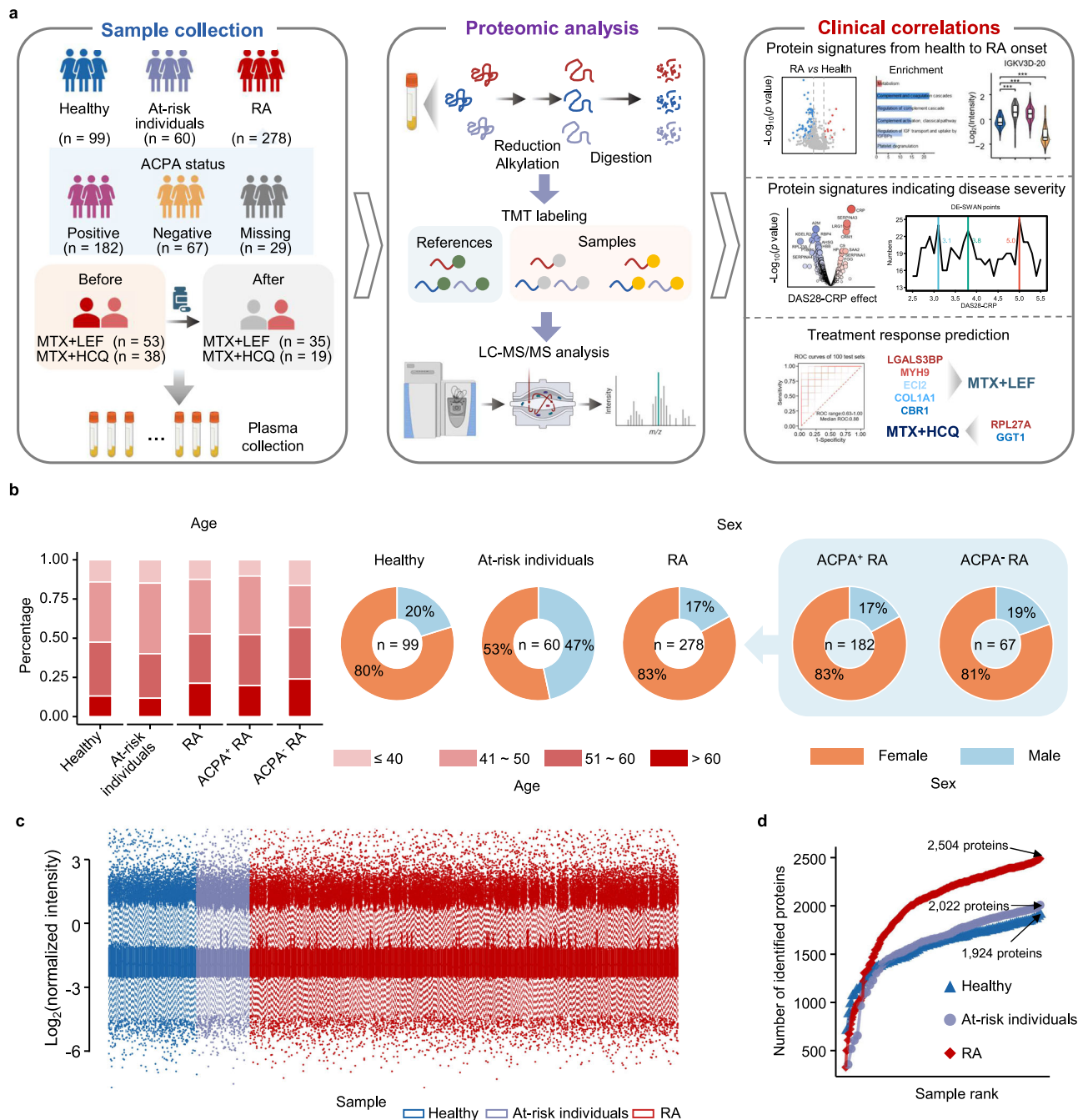


Fig. 1 | Proteomic analysis workflow and quality control. **a** Schematic design of the study (created with BioRender.com. Sun, R. (2025) <https://BioRender.com/6iee3nb>). **b** Bar plot (left) and pie charts (right) depicting the age and sex group distribution across different clinical subgroups. **c** Distribution of log₂-transformed protein ($n = 996$) intensities normalized to those of common reference samples.

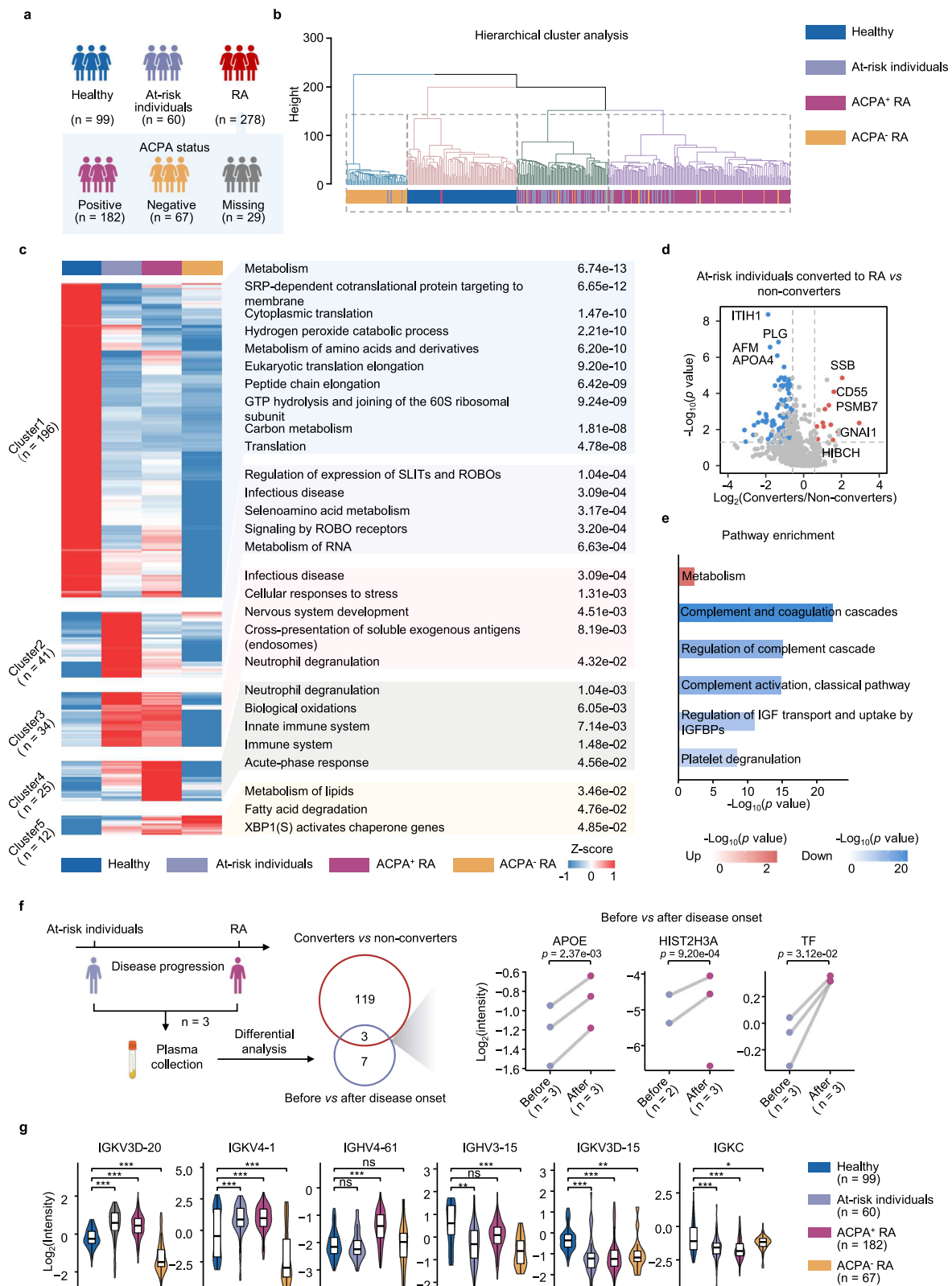
Box plots showing the median (center line), the 25th and 75th percentiles (bounds of box), and the minimum and maximum values (whiskers). **d** Cumulative number of identified proteins for healthy controls (blue, $n = 99$), at-risk individuals (violet, $n = 60$), and RA patients (red, $n = 278$). ACPA⁺ indicates ACPA-positive, and ACPA⁻ indicates ACPA-negative. Source data are provided as a Source Data file.

and after RA onset, included APOE, HIST2H3A, and TF. These findings highlight the roles of lipid metabolism dysregulation, neutrophil extracellular trap formation, and iron homeostasis in RA development^{34–36}.

IgG has dual roles in the pathogenesis of RA^{37,38}. We identified specific IgG segments with varying levels in the disease groups compared with those in the healthy controls. Specifically, IGKV3D-20, IGKV4-1, and IGHV4-61 increased in ACPA-positive RA or at-risk individuals, whereas IGHV3-15, IGKV3D-15, and IGKC decreased. Additionally, all the differential IgG segments were the lowest in ACPA-negative RA patients (Fig. 2g).

Identification of proteins associated with disease activity

Next, we investigated the proteins associated with disease activity. We observed significant sex differences in the DAS28-CRP scores among ACPA-positive RA patients, with higher disease activity in males than in females. In contrast, ACPA-negative RA did not show such sex-related differences (Fig. 3a, b). Owing to disease activity increasing with age, specifically in ACPA-positive females (Fig. 3c, d and Supplementary Fig. 2a), differentially expressed sliding window analysis (DE-SWAN) was conducted exclusively on female ACPA-positive RA patients. This analysis revealed a rapid decrease in the number of age-associated proteins after the age of 45 in females (Fig. 3e). In our study, this age



categorization further revealed disparities in DAS28-CRP, where females younger than 45 years presented reduced disease activity relative to their counterparts older than 45 years (Fig. 3f). In terms of clinical indicators, both the tender joint count (TJC) and CRP level exhibited similar trends, with both increasing in females over 45 years of age (Fig. 3f).

To reduce the influence of sex and age on protein calculations associated with disease activity, we adjusted for age and sex in subsequent analyses. Through multiple linear modeling, we initially identified the proteins associated with DAS28-CRP. Among these proteins, more were negatively correlated with DAS28-CRP (Fig. 3g). Proteins positively correlated with DAS28-CRP, such as CRP, LRG1,

Fig. 2 | Plasma proteomic heterogeneity during RA development. **a** Number of individuals in different clinical subgroups. **b** Dendrogram illustrating hierarchical clustering of proteomic data across samples. **c** Heatmap displaying unsupervised k-means clustering of proteins across healthy individuals, at-risk individuals, ACPA-positive RA patients and ACPA-negative RA patients (two-sided Student's *t* test, $p < 0.05$ and 1.5-fold change). The top enriched pathways for each cluster are shown (two-sided Fisher's exact test, $p < 0.05$). **d** Volcano plot of differentially expressed proteins (DEPs) between at-risk individuals who converted to RAs (converters) and non-converters (two-sided Student's *t* test, $p < 0.05$ and 1.5-fold change). The red and blue dots represent upregulated and downregulated proteins, respectively. **e** Bar plot displaying the top enriched pathways of DEPs between converters and non-converters (two-sided Fisher's exact test). **f** Schematic (left) of proteomic

analysis design for samples collected from three at-risk individuals before and after RA onset (created with BioRender.com. Sun, R. (2025) <https://BioRender.com/6iee3nb>). Venn diagram (middle) showing overlapped DEPs in two comparisons: red circle includes DEPs between converter and non-converter; blue circle includes DEPs before and after RA onset in three converters. Scatter plot (right) displaying the intensity of the overlapped DEPs before and after RA onset (two-sided Student's *t* test). **g** Violin plot displaying the intensity of antibody segments across four clinical groups (two-sided Student's *t* test). Box plots inside showing the median (center line), the 25th and 75th percentiles (bounds of box), and the minimum and maximum values (whiskers). ACPA* indicates ACPA-positive, and ACPA- indicates ACPA-negative. Significance is indicated as follows: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$, ns means $p \geq 0.05$. Source data are provided as a Source Data file.

ORM1, SERPINA4, and C9, were primarily associated with the acute-phase response and immune system. Conversely, the proteins that were negatively correlated with DAS28-CRP were involved mainly in biosynthesis and metabolism (Fig. 3h and Supplementary Fig. 2b). We further performed correlation analysis based on specific DAS28-CRP parameters (Fig. 3i). Among the proteins most significantly correlated with the clinical parameters, SERPINA3, LRG1 and HP were positively correlated with CRP, whereas ACOX1 and LRG1 were positively correlated with the swollen joint count (SJC). Moreover, HSD17B10 and RPL23A were negatively correlated with visual analogue scale (VAS) and TJC, respectively (Fig. 3i). ACPA-negative RA exhibited a consistent positive correlation between an intensified immune response and DAS28-CRP. Unexpectedly, almost no proteins were negatively correlated with DAS28-CRP or its four parameters (Supplementary Fig. 2c–f).

Due to differences in DAS28-CRP scores between ACPA-positive females under and over 45 years old, we investigated the impact of age on disease activity in this group. Overlap analysis of proteins associated with both age and DAS28-CRP was conducted (Fig. 3j). We found that CRP, SERPINA3, SAA2, and HP levels increased with age and were positively correlated with disease activity. Conversely, A2M, AHSG, and TF decreased with age and were negatively related to disease activity, highlighting the specific impact of aging on disease progression. Additionally, APOC3, RBP4, FN1 and NCL increased with age but were negatively correlated with disease activity. The age-related increase in these protective proteins warrants further investigation to understand the underlying mechanisms involved.

Decipher nonlinear proteomic fluctuations across DAS28-CRP

The relationship between plasma proteins and DAS28-CRP is intricate, extending beyond linear associations. To decode the complexity of proteomic dynamics fluctuating with DAS28-CRP, the most important parameter for assessing disease activity, two strategies have been applied.

First, to investigate proteomic differences based on clinical classification, we divided ACPA-positive RA patients into four groups based on DAS28-CRP: (I) remission (<2.6), (II) low ($2.6–3.2$), (III) moderate ($3.2–5.1$), and (IV) high (>5.1)³⁹. To reduce the complexity inherent in the proteome, we used unsupervised hierarchical clustering to group proteins with similar trajectories, resulting in six distinct clusters (Fig. 4a). Proteins associated with acute-phase responses, innate immunity, and neutrophil activity displayed increasing trends as disease activity increased in Clusters3. In Cluster6, proteins involved in carbon metabolism, IGF transport, and glycolysis consistently decreased with increasing DAS28-CRP. Proteins in Cluster5 and Cluster2, which are involved in pyruvate metabolism, ROBO signaling, and translation-related processes, initially increased from remission to low activity and then decreased. Notably, the fluctuations in complement in Cluster1 and Cluster4 suggest a dynamic balance between the activation and consumption of complement components as disease activity levels change. A similar analysis of ACPA-negative RA patients

revealed differences from ACPA-positive RA patients. ACPA-negative RA patients generally presented increased innate immune activity that decreased with increasing disease activity, weakened adaptive immune responses such as antigen presentation and T-cell receptor signaling, and a notable increase in amino acid metabolism and axon guidance (Supplementary Fig. 3). Overall, these results indicate that some plasma protein changes with increasing DAS28-CRP are nonlinear.

Second, given the nonlinear trends of most proteins across DAS28-CRP, as visualized by locally estimated scatterplot smoothing (LOESS)-estimated trajectories (Fig. 4b), we used DE-SWAN analysis to capture localized fluctuations at a smaller scale⁴⁰. We analyzed protein levels within a 40-sample window, comparing two groups within segments of 20 samples and incrementally sliding the window by 0.1 DAS28-CRP values from low to high disease activity. This analysis identified three key peaks at DAS28-CRP scores of 3.1, 3.8, and 5.0, revealing waves of protein level changes corresponding to these DAS28-CRP values (Fig. 4c, d). The peaks were related to distinct sets of proteins. At a DAS28-CRP score of 3.1, upregulated innate immune functions, such as complement activation and neutrophil degranulation, were observed, alongside inhibited anterograde transport. At a DAS28-CRP score of 3.8, inflammatory pathways were further upregulated, with impaired glucose metabolism. A DAS28-CRP score of 5.0 indicated elevated oxidative stress, with reduced ROBO signaling and protein metabolism (Fig. 4e). These dynamic and nonlinear changes in DAS28-CRP-associated proteins suggest that treatment strategies should be tailored to target specific proteins at different levels of disease activity.

Moreover, we assessed the correlations between the four components of DAS28-CRP and the proteins identified at the three peaks. Proteins correlated with the VAS significantly overlapped with DEPs at DAS28-CRP 3.8 and 5.0, while proteins related to other parameters showed greater overlap with DEPs at DAS28-CRP 5.0 (Fig. 4f). These findings suggest that the DAS28-CRP-related proteome exhibits distinct associations with different disease activity parameters.

Proteomic signatures for predicting treatment response via machine learning

MTX-based csDMARDs therapy is the first-line treatment, but the response rates to various combinations are not consistent. An in-depth analysis of the treatment response of longitudinal cohorts to csDMARDs is essential but remains unexplored. To address this issue, we used follow-up data from 206 patients treated with various csDMARDs. Subsequent assessments, following the European League Against Rheumatism (EULAR) criteria, were conducted after a period of more than three months⁴¹ (Supplementary Fig. 4a, b). We focused on the MTX + LEF ($n = 89$) and MTX + HCQ ($n = 64$) groups because of their adequate sample sizes for statistical analysis. RA patients with clinical remission and low disease activity were excluded because those with moderate to high disease activity were more likely to respond to treatment (Supplementary Fig. 4c, d). The age and sex

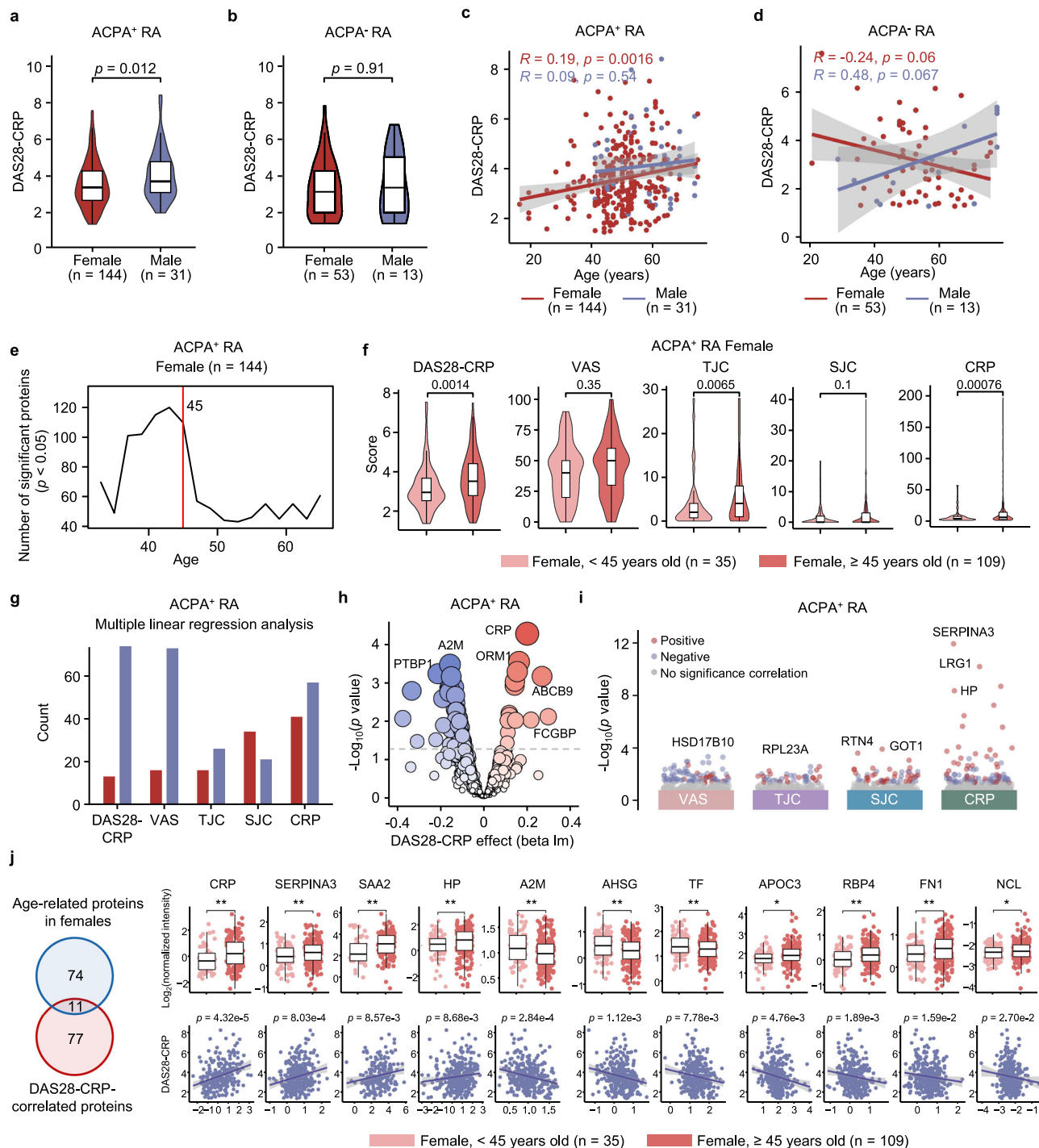


Fig. 3 | Impact of sex and age on disease activity and the proteome. Violin chart of the DAS28-CRP scores grouped by sex in ACPA-positive (a) or ACPA-negative (b) RA patients (two-sided Student's *t* test). Scatter plot with fitted regression lines illustrating Spearman's correlation (two-sided *p* value) between age and DAS28-CRP grouped by sex in ACPA-positive (c) or ACPA-negative RA (d). Gray band represents 95% confidence interval estimated using standard error of the mean (SEM). e DE-SWAN analysis of proteins across age in ACPA-positive females, with a peak at age 45 indicated by the red line. f Violin plot illustrating age-specific differences in DAS28-CRP and 4 clinical indicators (VAS, SJC, TJC and CRP) in ACPA-positive females aged above and below 45 years (two-sided Student's *t* test). g–i Multiple linear regression analysis (adjusted for age and sex, two-sided *p* value < 0.05) between DAS28-CRP indicators and proteins in ACPA-positive RA patients (*n* = 175). Bar plot showing the number of proteins significantly correlated with DAS28-CRP indicators (g), bubble plot displaying the regression analysis

between proteins and DAS28-CRP (h), and dot plot visualizing the regression analysis between proteins and VAS, TJC, SJC and CRP (i). j Venn diagram showing the overlap of proteins that exhibit significant changes between ACPA-positive females below and above 45 years or are significantly correlated with DAS28-CRP (left). Boxplots (right upper) displaying the normalized intensity of overlapped proteins across ACPA-positive females below and above 45 years (two-sided Student's *t* test). Scatter plots (right below) showing regression analysis between proteins and DAS28-CRP (adjusted for age and sex, two-sided *p* value). Gray band represents 95% confidence interval estimated using SEM. ACPA⁺ indicates ACPA-positive, and ACPA⁻ indicates ACPA-negative. Significance is indicated as follows: **p* < 0.05 and ***p* < 0.01. For box plots shown in (a, b, f, j), the center line represents the median; the bounds of the box indicate the 25th and 75th percentiles; and the whiskers extend to the minimum and maximum values. Source data are provided as a Source Data file.

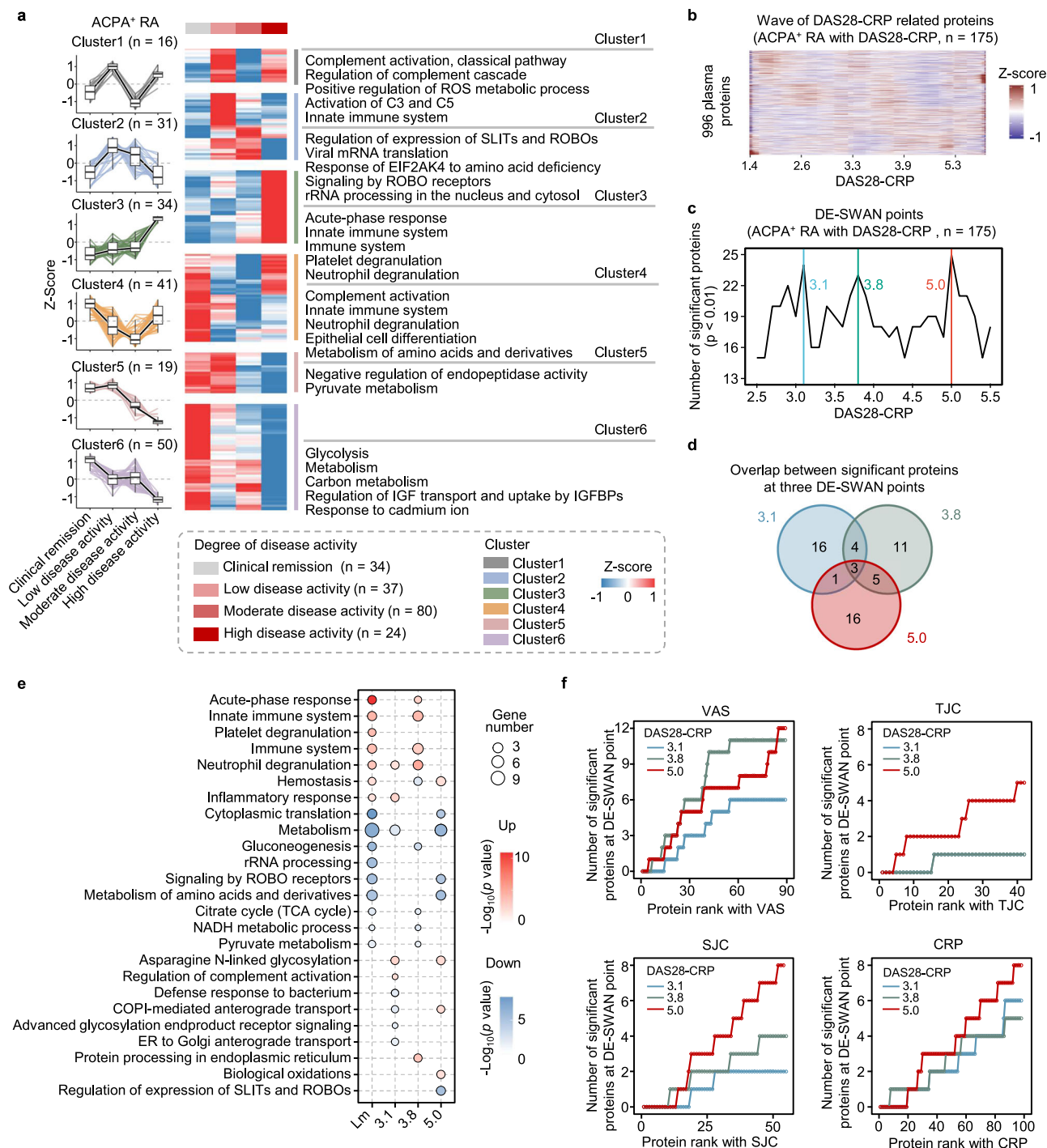


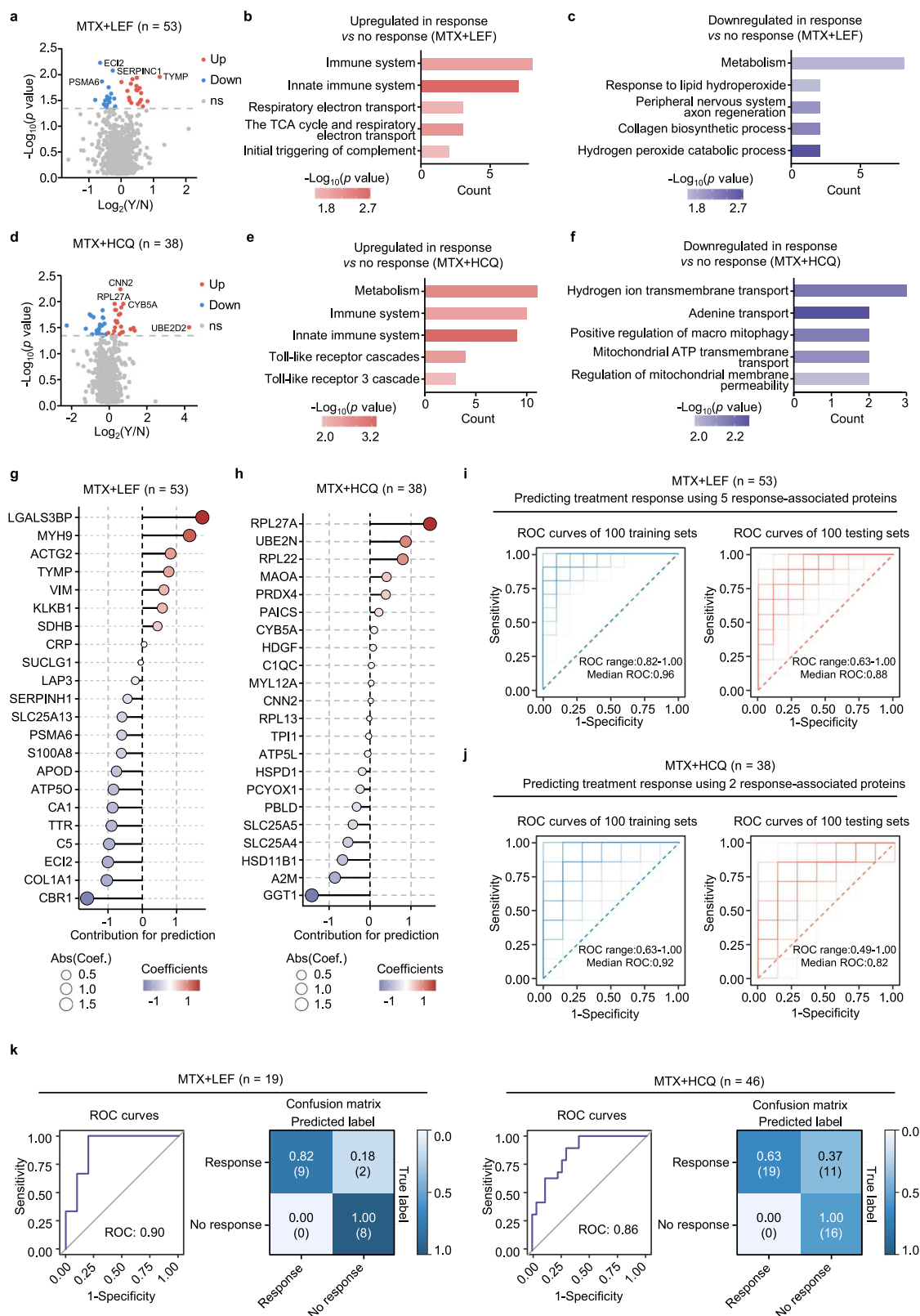
Fig. 4 | In-depth exploration of disease activity-related protein dynamics.

a Unsupervised *k*-means clustering analysis of DEPs across four disease activity groups (two-sided Student's *t* test, *p* < 0.05). The expression patterns of disease-related proteins in distinct clusters are shown on the left, with enriched pathways (more than 5 proteins) for each cluster on the right (two-sided Fisher's exact test). Box plots inside showing the median (center line), the 25th and 75th percentiles (bounds of box), and the minimum and maximum values (whiskers). **b** Heatmap visualizing protein trajectories across DAS28-CRP. The trajectories of 996 proteins are estimated using LOESS. **c** The number of DEPs across disease activity levels. DE-SWAN identified three local peaks at DAS28-CRP values of 3.1, 3.8, and 5.0.

d Overlap of proteins with significant differential expression at the three local peaks. **e** Bubble plot visualizing the enriched pathways of significant proteins identified through linear regression with DAS28-CRP and at three peaks in DE-SWAN (two-sided Fisher's exact test, *p* < 0.05). **f** Line plot visualizes the results of the linear regression analysis of proteins (significant at DAS28-CRP values of 3.1, 3.8, and 5.0 in DE-SWAN) with VAS, TJC, SJC, and CRP. The cumulative number of overlapped proteins that are significant either at DE-SWAN points or in relation to the four parameters is shown, with proteins ranked based on significance from the linear regression models (adjusted for age and sex, two-sided *p* value < 0.05). Source data are provided as a Source Data file.

differences between responders and non-responders were not significant in either group (Supplementary Fig. 4e, f). Initially, we conducted differential analyses between responders and non-responders without considering sex and age effects. In patients responsive to

MTX + LEF treatment, there were increased proteins related to immunity and energy metabolism, alongside decreased proteins related to lipid oxidation (Fig. 5a–c). In patients responsive to MTX + HCQ treatment, we detected elevated protein levels associated with



metabolism, immunity, and toll-like receptor cascades, and reduced protein levels associated with transport pathways (Fig. 5d–f). Furthermore, we analyzed these differences between responders and non-responders in the ACPA-positive RA group, which had a sufficient sample size for statistical analysis. MTX+LEF responders showed increased complement activation, fibrinolysis, and autophagy, with downregulated metabolic and glycolytic pathways (Supplementary

Fig. 5a, b), while MTX+HCQ responders exhibited upregulated immune activation and downregulated mitochondrial transport pathways (Supplementary Fig. 5c, d). Given that sex may affect treatment response⁴², we also examined its impact on response-related proteomics. In female responders to MTX+LEF, we observed elevated protein transport and inflammatory pathways, while male responders showed increased endocytosis (Supplementary Fig. 5a, b). Among the

Fig. 5 | Machine learning-driven discovery of key proteins for predicting the response to csDMARDs treatment. **a** Volcano plot of DEPs between response and no response to MTX + LEF treatment (two-sided Student's *t* test, $p < 0.05$). Y = response, N = no response. Enrichment analysis of upregulated (**b**) and down-regulated (**c**) proteins in response *vs* no response to MTX + LEF treatment (two-sided Fisher's exact test, $p < 0.05$). **d** Volcano plot of DEPs between response and no response to MTX + HCQ treatment (two-sided Student's *t* test, $p < 0.05$). Y = response, N = no response. Enrichment analysis of upregulated (**e**) and down-regulated (**f**) proteins in response *vs* no response to MTX + HCQ treatment (two-

sided Fisher's exact test, $p < 0.05$). LASSO regression analysis showing the contribution of DEPs to treatment response prediction in the MTX + LEF (**g**) and MTX + HCQ (**h**) groups. ROC curves illustrating the predictive performance of the LASSO model for MTX + LEF (**i**) and MTX + HCQ (**j**) responses, using the top 5 or 2 proteins, respectively, in both the training (left) and testing (right) sets, with 10-fold cross-validation repeated 100 times. **k** ROC curve showing model performance after integrating protein levels measured by ELISA. The confusion matrix displays sensitivity and specificity at the optimal cutoff for the MTX + LEF (left) and MTX + HCQ (right) groups. Source data are provided as a Source Data file.

MTX + HCQ responders, females presented increased nonsense-mediated decay and decreased amino acid metabolism (Supplementary Fig. 5c, d).

Furthermore, we developed models using plasma proteins to predict treatment response. By employing least absolute shrinkage and selection operator (LASSO) feature selection on characteristic proteins, we constructed linear regression models and calculated the contribution scores of these proteins to the models. Proteins with absolute contribution values greater than 1 were ultimately selected for model construction (Fig. 5g, h). We ensured equal numbers of responders and on-responders in both the training and testing sets. After 10-fold cross-validation to determine the optimal regularization parameter, we performed 100 iterations to generate an average receiver operating characteristic (ROC) curve, ensuring stable and reliable predictions (Supplementary Fig. 6a). In the model for predicting the MTX + LEF treatment response, five proteins were used, with LGALS3BP and MYH9 increased in responders, while ECI2, COL1A1, and CBR1 decreased in responders. For the MTX + HCQ treatment, RPL27A was a positive predictor and GGT1 was a negative predictor. The LASSO-selected proteins predictors all ranked within the top 10 across multiple other feature selection methods (random forest, recursive feature elimination combined with support vector machine, XGBoost, stability selection and elastic net), supporting their robustness (Supplementary Fig. 6b). Cross-validation yielded an average ROC of 0.96 for the training set and 0.88 for the testing set in MTX + LEF treatment groups (Fig. 5i). The predictive ROC values were 0.92 for training and 0.82 for testing in MTX + HCQ treatment groups (Fig. 5j). SHAP analysis was performed to interpret the contribution of individual proteins to the predictive models, confirming that their effect directions were consistent with those identified by feature selection (Supplementary Fig. 6c). In addition, we built prediction models with random forest and XGBoost using LASSO-identified features, but their median ROC values remained lower than those from LASSO (Supplementary Fig. 6d, e). These findings consistently highlight the superior predictive performance of LASSO. Furthermore, incorporating DAS28-CRP parameters (VAS, SJC, TJC, and CRP) into the protein features slightly improved the predictive performance, with median ROC values of 0.90 (*vs.* 0.88) for MTX + LEF and 0.84 (*vs.* 0.82) for MTX + HCQ in the testing sets (Supplementary Fig. 6f).

We validated our model performance in an independent cohort of 46 RA patients receiving MTX + HCQ and 19 patients receiving MTX + LEF. The enzyme-linked immunosorbent assay (ELISA) results revealed consistent biomarker changes with proteomic data between responders and non-responders (Supplementary Fig. 6g). Integrating these protein levels into our model maintained strong classification efficiency, with ROC values of 0.90 for MTX + LEF and 0.86 for MTX + HCQ. Using a confusion matrix to determine the optimal cutoff, the MTX + LEF model successfully identified 9 out of 11 responders with no false negatives. In contrast, the MTX + HCQ model exhibited a sensitivity of 0.63 and specificity of 1.0, which may be influenced by the smaller discovery cohort size (Fig. 5k). Overall, both LASSO models demonstrated robust predictive performance and can accurately predict treatment responses for the two most common MTX

combination therapies, offering valuable insights for personalized treatment strategies.

Proteomic changes after treatment in RA patients who respond

To investigate the proteomic changes after MTX + LEF or MTX + HCQ treatment in RA patients who responded, differential analyses were performed (Fig. 6a, b). We found that retinol metabolism and cytoplasmic translation increased, whereas actin cytoskeleton and acute-phase responses decreased in responders after MTX + LEF treatment (Fig. 6a, c). In contrast, mRNA metabolism, retinol metabolism and cell adhesion increased, while the complement pathway decreased in responders after MTX + HCQ treatment (Fig. 6b, d). Notably, these pathways did not show pronounced changes in non-responders following either treatment (Fig. 6e, f and Supplementary Fig. 7a, b).

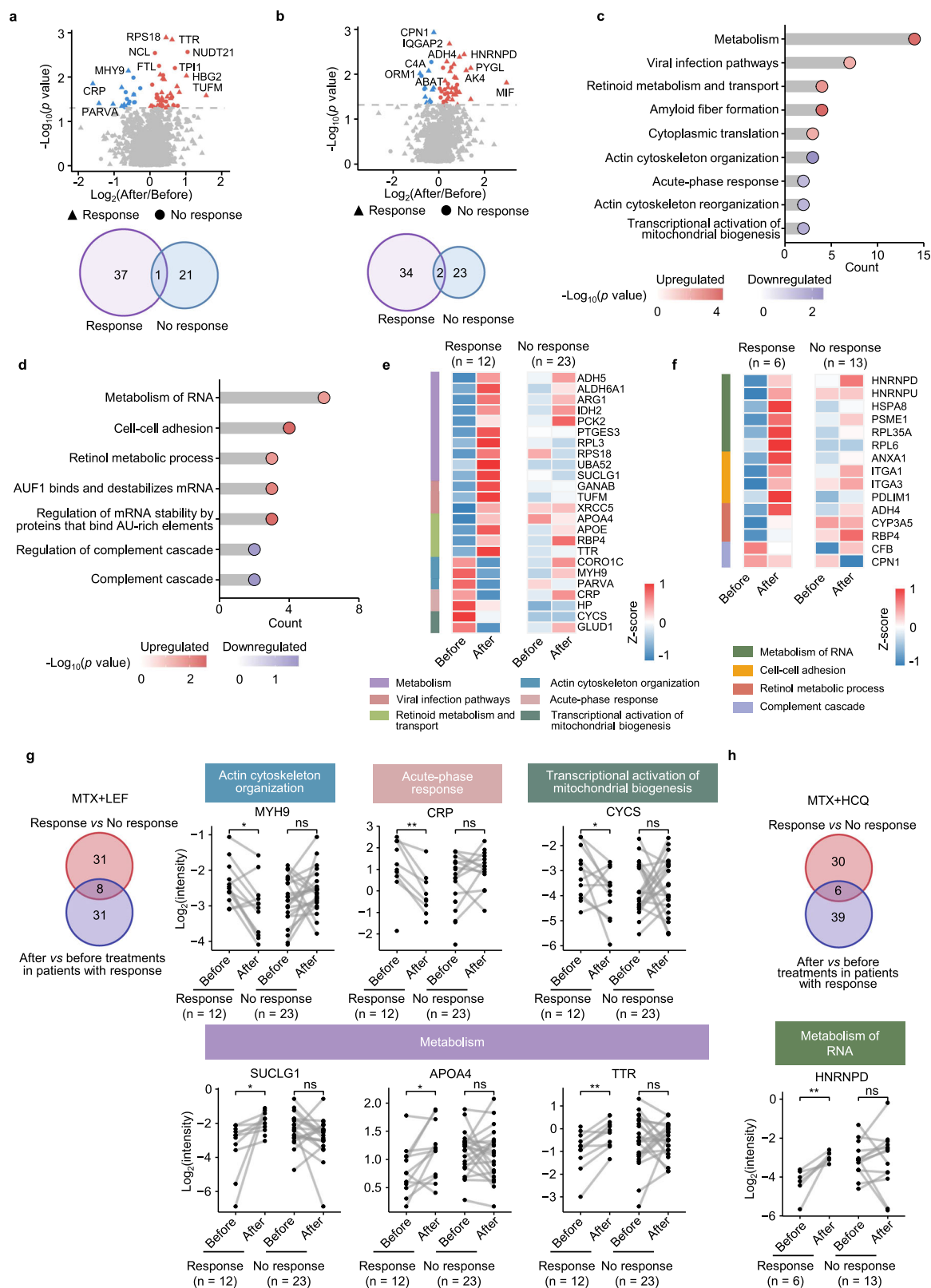
To identify pivotal factors contributing to pharmacological efficacy, we performed overlapping analysis between proteins associated with treatment response and those significantly changed after csDMARDs treatment. In the MTX + LEF group, eight common proteins involved in the acute response, actin cytoskeleton organization, mitochondrial biogenesis activation and metabolism were identified (Fig. 6g). In the MTX + HCQ group, six overlapping proteins were identified (Fig. 6h). These proteins may serve as potential targets for these two csDMARDs therapies.

Besides, we explored the effects of sex and ACPAs status on treatment-induced proteomic changes. Due to the limited number of ACPA-negative RA patients receiving both treatments and the limited number of males receiving MTX + HCQ treatment, these patients were not included in the analysis. In ACPA-positive RA patients receiving MTX + LEF, translation, amino acid metabolism, and axon guidance were increased. After MTX + LEF treatment, female responders presented elevated protein and RNA metabolism, whereas male responders showed increased actin cytoskeleton regulation (Supplementary Fig. 7c–e). In contrast, after MTX + HCQ treatment, RNA metabolism and axon guidance increased in ACPA-positive responders (Supplementary Fig. 7f, g). Our analysis indicates that sex has a certain impact on csDMARDs therapy-induced proteomic changes.

Discussion

Owing to the complexity and heterogeneity of the mechanisms underlying RA, as well as the inefficacy and various adverse reactions to medications, proteomics-driven precision medicine plays a crucial role in the personalized treatment of RA. This work yields several key findings. First, our study delineates the characteristic molecular profiles of each RA subtype, revealing potential therapeutic targets for interventions in the preclinical stages of RA, as well as in ACPA-negative RA. Second, we explore proteins that underwent linear and nonlinear changes with DAS28-CRP, identifying fluctuation peaks at scores of 3.1, 3.8, and 5.0. Third, treatment response-related proteins differ between the MTX + LEF and MTX + HCQ therapies, aiding in predictive model development and revealing potential molecular mechanisms to enhance treatment efficacy.

RA is characterized by aberrantly activated autoimmune responses. Recent studies have uncovered cellular dysfunctions in RA and



dysregulation of energy and nutrient metabolism^{43–45}, as well as protein processing⁴⁶. Our research reveals how these functions are affected at the protein level and their implications for RA progression and therapeutic interventions. The acute-phase response-related proteins not only showed significant associations with disease activity but also emerged as primary factors elucidating sex or age disparities in the DAS28-CRP.

Heterogeneity in RA is evident across different clinical phases and serological statuses^{47–49}. In our study, we find notable proteomic features related to these factors, which might help achieve better personalized precision medicine. First, we observe a notable increase in RNA metabolism in at-risk individuals, especially in males, along with the upregulation of the ROBO receptor signaling pathway, which inhibits osteogenic differentiation²⁹. Compared with those in both the RA and

Fig. 6 | Plasma protein signatures in csDMARDs-treated RA patients with different responses. **a** Volcano plots showing DEPs before and after MTX + LEF treatment, stratified by treatment response (response, $n = 12$; no response, $n = 23$) (paired two-sided Student's t test, $p < 0.05$). **b** Volcano plots showing DEPs before and after MTX + HCQ treatment, stratified by treatment response (response, $n = 6$; no response, $n = 13$) (paired two-sided Student's t test, $p < 0.05$). **c** Pathway enrichment analysis of DEPs before and after MTX + LEF treatment in response (two-sided Fisher's exact test). **d** Pathway enrichment analysis of DEPs before and after MTX + HCQ treatment in response (two-sided Fisher's exact test). **e** Heatmap

of the relative abundance of DEPs before and after MTX + LEF treatment, separated by response. **f** Heatmap of the relative abundance of DEPs before and after MTX + HCQ treatment, separated by response. Venn diagrams displaying overlap of treatment- and response-related proteins for MTX + LEF (response, $n = 12$; no response, $n = 23$) (**g**) and MTX + HCQ (response, $n = 6$; no response, $n = 13$) (**h**) therapies, grouped by response, with corresponding dot plots illustrating the differential expression of these proteins among groups (paired two-sided Student's t test). Significance is indicated as follows: $*p < 0.05$, $**p < 0.01$, ns means $p \geq 0.05$. Source data are provided as a Source Data file.

healthy groups, some proteins even reach their highest or lowest levels in the at-risk group. Although at-risk individuals are clinically considered to be in an intermediate stage, we believe that this represents a distinct biological stage with a unique protein expression profile rather than merely a transitional phase. These divergent proteins could serve as early biomarkers or therapeutic targets, potentially altering the disease course before clinical RA onset. Second, we reveal that lipid metabolism was elevated in ACPA-negative RA patients, suggesting increased metabolic demand or a modification in energy metabolism, which could present potential treatment targets. Moreover, IgG sequence diversity in autoimmune diseases has been demonstrated in studies of BCR sequences⁵⁰. We discover serum IgG segments with different levels among the clinical groups, indicating that autoantigen-driven antibody gene rearrangements underlie the transition from healthy to disease⁵¹.

Notably, our research demonstrates nonlinear changes in proteins associated with DAS28-CRP. We identified three protein dynamics peaks using DE-SWAN analysis, corresponding to DAS28-CRP scores of 3.1, 3.8, and 5.0. The 3.1 point closely approaches the widely used low disease activity point at 3.2. At this crest, we note an enhanced innate immune response. These changes are notably linked to the VAS score. Considering that the proteins at this stage may reflect the transition from moderate to mild disease activity, studying their molecular mechanisms may provide insights into the pathogenesis of patients with low disease activity, which will further help achieve remission, in line with the treat-to-target strategy⁵². A continued intensification of inflammation is observed at point 3.8, along with inhibited glucose metabolism. The limited correlation identified between the DAS28-CRP parameters and protein changes at 3.8 suggests a promiscuous mechanism in the moderate disease activity group. Notably, the 5.0 crest, which is close to the high disease activity cutoff, exhibits the strongest associations with the TJC and SJC. The protein changes include increased biological oxidation and decreased amino acid metabolism, translation, and ROBO signaling. These findings provide potential insights into the underlying mechanisms of severe disease status.

According to the recommendations, csDMARDs serve as the first line for treating RA⁵³, even though patients face challenges related to adverse reactions and suboptimal responsiveness. In this context, identifying distinct characteristics and predictive signatures for treatment response to these traditional drugs is crucial. Our analysis reveals the proteomic changes of commonly used therapies, including MTX + LEF, whose safety has been previously validated in Chinese cohorts^{54,55} and MTX + HCQ. These combinations effectively regulate immune functions, including complement activation, acute phase responses, and neutrophil degranulation, and they restore RNA metabolism. After identifying the characteristic proteins in the responsive population, we construct prediction models for MTX + LEF and MTX + HCQ treatment response. These models demonstrate promising efficacy and were subsequently validated in independent cohorts.

While this study provides valuable insights into both the pathogenic mechanisms and pharmacological strategies in RA, it is important to acknowledge several limitations, particularly the relatively

small sample sizes in certain subgroups, including those at risk before and after disease onset, as well as in the cohort used to validate the drug response prediction model. The limited sample sizes may be partially attributable to the small number of at-risk individuals who progress to clinical disease. Previous studies have shown that ACPA-positive individuals with arthralgia have an approximately 28% risk of developing RA⁵⁶. Although our at-risk individuals are asymptomatic, our follow-up data reveal that 8 out of 38 individuals (21.1%) progressed to RA, reflecting a consistent progression rate. Long-term follow-up (5–7 years) results in a limited number of samples available for comparison between converters and non-converters. Our focus on plasma proteomics within the circulatory system may have overlooked nuances present in the synovium⁵⁷, a critical site in the pathology of RA. These considerations provide avenues for future research to refine and expand our understanding of this complex bundle of autoimmune diseases.

Methods

Study design and ethics approval

Plasma samples were obtained from 99 healthy controls, 60 at-risk individuals, and 278 patients with RA. These samples were collected at West China Hospital of Sichuan University, following the approval of the Research Ethics Committee of West China Hospital at Sichuan University (Permission number: 2021(790)), and informed consent was obtained from all participants. Patients were diagnosed with RA by meeting the 2010 American College of Rheumatology / EULAR criteria. According to the EULAR, at-risk individuals can be defined by the presence of one or more of the following criteria: (a) genetic risk factors for RA, (b) environmental risk factors for RA, (c) systemic autoimmunity associated with RA, (d) symptoms without clinical arthritis, and (e) unclassified arthritis. In the context of our study, the at-risk individuals specifically corresponded to those in phase (c), characterized by systemic autoimmunity associated with RA⁵⁸. Healthy controls were age- and sex-matched individuals with no history or clinical evidence of autoimmune or rheumatic diseases⁵⁹. All participants were enrolled randomly without prior sex-based selection or stratification. Sex of participants was determined based on self-report. Blood collection adhered to standard venipuncture protocols, utilizing anticoagulant tubes. After centrifugation to obtain the supernatant, the samples were stored at -80°C until analysis. ACPAs levels were measured via the Elecsys anti-CCP assay (Roche Diagnostics, Mannheim, Germany) on the Cobas® e 801 modules, with results classified as either positive (≥ 17.0 U/mL) or negative (< 17.0 U/mL). The human tissues used for common reference samples were from distant normal tissues of cancer patients, with approval from the Research Ethics Committee of West China Hospital, Sichuan University (approval numbers: 2019(538) for liver, 2019(539) for lung, and 2020(374) for intestine). Normal kidney tissue was obtained from renal transplant donors with approval number 2019(748).

Protein extraction and digestion

The plasma samples were first thawed and then diluted 10-fold with precooled phosphate-buffered saline containing protease and phosphatase inhibitors. From each diluted plasma sample, a 16.7 μL aliquot

(~100 µg of protein) was further diluted to 100 µL with 100 mM triethylammonium bicarbonate (Sigma-Aldrich, Cat. No. T7408) buffer. The resulting samples were reduced at 56 °C for 1 h with 10 mM Tris (2-carboxyethyl) phosphine (Sigma-Aldrich, Cat. No. C4706), followed by alkylation with 17 mM iodoacetamide (Sigma-Aldrich, Cat. No. I6125) at room temperature in the dark for 35 minutes. Next, ~100 µg of protein from each sample was digested for 14 h at 37 °C with trypsin (Promega, Cat. No. V5117) at a ratio of 1:50 (w/w) (2 µg/µL). A C18 solid-phase extraction column (TECAN, CEREX 10 mg, Cat. No. 417-0101R) was used to desalt the tryptic peptides, and the samples were dried in a vacuum concentrator before isobaric labeling.

TMT labeling

TMT (Thermo Scientific, Product catalog number: 90066; Lot number: RJ236348) reagents were employed for isobaric labeling. To minimize cross-isotope contamination between the common internal reference and experimental samples, TMT-126 was used to label the common reference sample. The experimental samples were labeled with TMT-129 or TMT-131, and empty channels were strategically placed between them. Equal amounts of proteins derived from pooled plasma, liver, lung, kidney, and intestine tissues were combined to create reference samples. The utilization of this reference sample serves two main purposes. (I) It acts as a reference sample, reducing batch effects during the analysis of MS data. (II) It acts as a carrier protein to increase the composite intensity of low-abundance proteins in plasma and thus increases the likelihood of their detection by MS^{60–62}. This strategy allows for high-throughput identification and quantification of plasma proteins without the need to remove high-abundance plasma proteins. The excess TMT reagents were subsequently quenched, and the samples labeled with TMT-129 or TMT-131 as well as the reference sample were mixed, desalted and then dried via a speed–vacuum system.

LC–MS/MS analysis

Peptide samples were analyzed via a Q Exactive HF high-resolution MS coupled with an EASY-nLC 1200 nanoflow high-performance liquid chromatograph system (both Thermo Fisher Scientific). The samples were redissolved in loading buffer (2% ACN, 0.1% FA) and loaded onto a 75 µm × 2.5 cm homemade trap column (Spursil C18, 5 µm particle size, DIKMA, Cat. No. 85251) and coupled to a homemade capillary column (25 cm length × 75-µm inner diameter, Reprgsil-PurC18-AQ-1.9 µm: particle size, Dr. Maisch, Cat. No. r119.aq.0001). Separation was achieved via a gradient of 8–100% HPLC buffer B (0.1% formic acid, 2% DMSO in 80% acetonitrile) in buffer A (0.1% formic acid, 2% DMSO in 98% water). The gradient flow rate was set at 330 nL/min for 90 min, following this pattern: 0–3 min, 8–8% B; 3–20 min, 8–12% B; 20–80 min, 12–25% B; 80–85 min, 25–95% B; and 85–90 min, 100% B. Data-dependent acquisition (DDA) was configured in positive ion mode for a full mass spectrometry survey scan spanning from 350 to 1600 m/z, with a resolution of 60,000, a maximum injection time of 100 ms, and an automatic gain control (AGC) target value of 1e6. The top 20 MS precursors were chosen with a 0.4 m/z isolation window and fragmented with 30% normalized collision energy. The MS2 scans were carried out at a resolution of 30,000, an AGC target of 5e5, and a maximum injection time of 120 ms. Unassigned ions or those with a charge state of $z=1$ or 3–8 were excluded from MS/MS, and the intensity threshold was set to 2.8e5.

Database searching

For data analysis, the raw MS data were searched against the human UniProt sequence database via MaxQuant⁶³ (version 1.6.1.0). The first search mass tolerance, the main search peptide tolerance and the fragment ion mass tolerance were set at 10 ppm, 4.5 ppm and 0.02 Da, respectively. The database search included cysteine carbamidomethylation as a fixed modification, as well as methionine oxidation, TMT6-plex (Lys), and protein *N*-terminal acetylation as variable

modifications. Trypsin was selected as the protease, and two missing cleavages were allowed. A minimum peptide length of 6 amino acids was applied, and the peptide false discovery rate was set to 1%. Proteins with at least one unique peptide were preserved.

MS data processing

The protein levels within each TMT batch were normalized to their levels in the TMT-126-labeled internal reference. The datasets from all TMT batches were combined into an expression matrix, and a log2 transformation was applied to the merged data. To ensure reliable plasma protein identification in our study, we created a plasma protein database that incorporates proteins from Human Plasma Protein Project⁶⁴ and Human Protein Atlas^{65,66}, as well as those identified in previous plasma proteomes^{12,67–71}. Following an overlapping analysis between the identified proteins in this work and the proteins in the plasma protein databases, any uncertain plasma proteins identified by this strategy were excluded. Only proteins detected in more than 50% of the samples in each disease group were preserved, and the resulting matrix was imputed via the random forest function from the R-randomForest package version 4.6-14. This imputed matrix was used for subsequent data analyses.

Bioinformatics and statistical analysis

Differential expression analysis among various groups was tested by two-sided Student's *t* test. Spearman's correlation coefficients were employed to calculate the correlations between common internal references or between experimental samples. Gene Ontology term analysis⁷² and Reactome enrichment analysis were conducted via the Database for Annotation, Visualization, and Integrated Discovery (DAVID) Bioinformatics Resources. The *p* values for pathway enrichment analysis were calculated using the DAVID tool based on two-sided Fisher's exact test. The enrichment scores of various pathways in each sample were assessed via the ssGSEA algorithm⁷³ from the GSVA package (version 1.48.3).

To assess the impact of the DAS28-CRP score on protein expression, a linear regression model was applied, incorporating age and sex as covariates as follows:

$$\text{Protein level} \sim \alpha \cdot \text{DAS28} - \text{CRP} + \beta_1 \cdot \text{sex} + \beta_2 \cdot \text{age}$$

The proteins exhibiting significant positive or negative linear correlations ($p < 0.05$) were subsequently subjected to pathway enrichment analyses via DAVID.

DE-SWAN

To discern and quantify alterations in the plasma proteome concerning DAS28-CRP and age in females, the DE-SWAN method from the R package DE-SWAN (version 0.0.0.9001) was employed⁴⁰. The center of the analysis window was shifted in increments of 0.1 DAS28-CRP values, spanning from low to high, and the protein levels of the 20 samples closest to the window's center on each side were compared. The analysis was conducted via the following linear model:

$$\text{Protein level} \sim \alpha \cdot \text{DAS28} - \text{CRP}_{\text{Low/High}} + \beta_1 \cdot \text{sex} + \beta_2 \cdot \text{age}$$

Proteins exhibiting statistical significance ($p < 0.05$) within the peaks with the most substantial fluctuations (3.1, 3.8, 5.0) were selected for pathway enrichment analysis via DAVID.

Machine learning for treatment response

To prevent overfitting of the prediction model, we imposed feature penalties on the protein characteristics. We applied LASSO via the glmnet package⁷⁴ (version 4.1-4) in R to construct linear regression models for the MTX + LEF and MTX + HCQ treatment groups, which

were used to assess the contribution of the DEPs to the treatment response^{75–77}.

First, we standardized the proteomics data via scale normalization. We subsequently performed 10-fold cross-validation on the basis of the mean squared error (MSE) criterion to select the optimal lambda value (minimum MSE plus one standard deviation), with each observation assigned a weight of 1. (Parameters: alpha = 1, nfold = 10, family = “binomial”, type.measure = “mse”, s = “lambda.1se”, weights = 1 and alignment = “lambda”). Finally, to establish a reliable drug prediction model, we selected the most stable feature proteins through 50 random loops. The contribution value of each predictor (protein) in each prediction model was derived by averaging the coefficients across the 50 iterations, as expressed by the following formula:

Contribution = average (coefficient)

i: Number of random loops in the linear model

Proteins with absolute contribution values exceeding 1 were chosen as features for the formal prediction analysis. The protein features utilized included CBR1, LGALS3BP, MYH9, COL1A1 and ECI2 (MTX + LEF), along with GGT1 and RPL27A (MTX + HCQ).

To optimize the LASSO model, we used the cv.glmnet function to perform 10-fold cross-validation and identify the optimal regularization parameter (λ). λ_{\min} , which minimizes the cross-validation error, is selected as the optimal parameter. The function is run with the parameters nfold = 10, family = “binomial”, and alpha = 1 to apply LASSO regularization. This step ensures a balance between model complexity and predictive performance, preventing overfitting while maintaining accuracy. Once λ_{\min} is determined, it is used to build the final LASSO regression model, with lambda = λ_{\min} and alpha = 1, enforcing sparsity in the selected features. For model construction, samples are randomly divided into training and testing sets, ensuring equal numbers of responders and non-responders in each random sampling. The trained model is applied to predict response probabilities using the parameter type = “response”, producing robust and reliable probability estimates for drug response outcomes. Hyperparameter tuning ensures that the LASSO model is optimized for the dataset, improving its generalizability and predictive reliability. Finally, we used the multipleROC function from the pROC package⁷⁸ (version 1.18.5) to calculate the ROC curve. To estimate the confidence interval for each ROC, we performed 100 iterations and calculated the median ROC curve⁷⁵.

In addition to LASSO, feature selection was performed using random forest, recursive feature elimination combined with support vector machine (REF + SVM), XGBoost, stability selection, and elastic net. Random forest and XGBoost were also used for model prediction. For REF + SVM, the model iteratively removed the least important features and evaluated performance across feature subsets using 5-fold cross-validation, yielding a stable subset of informative features via the caret package⁷⁹ (version 7.0-1) in R. Stability selection was performed by repeatedly fitting LASSO models on subsampled datasets. We used the stabsel function combined with the lars.lasso fitting method, performing 100 subsampling iterations (sampling.type = “MB”) and setting the per-family error rate to 1. Features with selection frequencies exceeding 0.75 were considered stable and retained for downstream analysis via the stabs package⁸⁰ (version 0.6-4) in R. For elastic net, the optimal regularization strength (λ) was determined via 10-fold cross-validation, and features with non-zero coefficients at the lambda.1se value were retained as selected features via the glmnet package⁸¹ (version 4.1-4) in R. For Random Forest, we used the following settings: 500 trees, the square root of the number of features for splits (mtry), and a minimum node size of 1 via the randomForest⁸² package (version 4.7-1.1) in R. For XGBoost, we set the maximum tree depth to 4, the learning rate to 1, 10 boosting rounds, and 2 threads. The objective was binary logistic regression via the xgboost package⁸³ (version 1.7.8.1) in R.

To enhance the interpretability of the treatment response prediction models, we employed SHapley Additive exPlanations (SHAP) to

quantify the contribution of each feature to model outputs. Specifically, we used the fastshap package⁸⁴ (version 1.18.5) to compute SHAP values based on 50 simulations of a custom logistic regression prediction function (predict_median_logistic). The computed SHAP values, along with the original feature matrix, were used to construct a shapviz object for downstream visualization and interpretation.

Enzyme-linked immunosorbent assays

Serum concentrations of protein features, including COL1A1 (Solarbio, China, Cat. No. SEKH-0401), MHY9 (Signalway Antibody, Pearland, USA, Cat. No. EK15634), ECI2 (EIAab, Wuhan, China, Cat. No. EL6269h), LGALS3BP (Boster Biological Technology, Wuhan, China, Cat. No. EK1240), and CBR1 (COIBO BIO, China, Cat. No. CB16353-Hu) for MTX + LEF, GGT1 (Signalway Antibody, Pearland, USA, Cat. No. EK14228) and RPL27A (EIAab, Wuhan, China, Cat. No. E5486h) for MTX + HCQ, were quantified via a commercially available ELISA kit. The detailed protocols for each assay are accessible on the manufacturer’s website (Supplementary Table 2), and all procedures were conducted in strict accordance with the manufacturer’s instructions. The plasma samples were prepared at various concentrations to meet the required protein levels. Following the manufacturer’s protocol, 300 μ L of wash buffer was added to each well and incubated for 30 seconds. After the wash buffer was removed, the microplate was gently tapped dry on absorbent paper; this washing step was repeated twice. Then, 100 μ L of 2-fold serially diluted standards was added to the standard wells, and 100 μ L of sample was added to the sample wells. The plate was incubated at room temperature ($25 \pm 2^\circ\text{C}$). Subsequently, 100 μ L of biotinylated antibody solution was added to each well. The plate was sealed and incubated at room temperature for 90 min. Next, 100 μ L of the prepared avidin-biotin-peroxidase complex was added to each well, and the plate was covered with a plate sealer and incubated for 40 min at room temperature. Next, 90 μ L of tetramethyl benzidine dihydrochloride (TMB, NEOBIOSCIENCE, Cat. No. TMS.600) substrate solution was added to each well, and the plate was incubated in the dark at room temperature for 30 min. Finally, 100 μ L of stop solution was added to each well, ensuring that the stop solution was added in the same order as the TMB substrate. The optical density values were measured within 5 min via a microplate reader at a dual wavelength of 450 nm. Alternatively, the mean absorbance for each standard was plotted against the concentration. Four-parameter logistic regression was used on the standard curve generated with curve fitting software to interpolate the concentration of the sample.

Validation of the treatment response prediction model

On the basis of the results of the previous 100 training iterations using the proteomic data, the average coefficient for each protein feature was taken as the final coefficient for the drug prediction model. The protein concentrations detected by ELISA were standardized and then input into the model. ROC curve analysis was then performed to evaluate the sensitivity and specificity of the model’s classification. To further investigate the model’s sensitivity and specificity, a confusion matrix was constructed using the predicted probabilities from the test set. The probability threshold was estimated via the coords function in the pROC package⁷⁸ (version 1.18.5) and the Youden index, with the cutoff value determined via the Youden method^{85,86}. Differences in each biomarker between the responder and non-responder groups were assessed using a two-sided Mann–Whitney U test.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the iProX partner repository^{87,88}

with the dataset identifier [PXD048245](#). All data are included in the Supplementary Information or available from the authors, as are unique reagents used in this Article. The raw numbers for charts and graphs are available in the Source Data file whenever possible. Source data are provided with this paper.

Code availability

The source code, including differential computation, feature selection, prediction models and plotting, is publicly available on Zenodo at <https://doi.org/10.5281/zenodo.15717981>⁸⁹.

References

- Smolen, J. S. et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. *Ann. Rheum. Dis.* **82**, 3–18 (2023).
- Weyand, C. M. & Goronzy, J. J. The immunology of rheumatoid arthritis. *Nat. Immunol.* **22**, 10–18 (2021).
- Mun, S. et al. Serum biomarker panel for the diagnosis of rheumatoid arthritis. *Arthritis Res. Ther.* **23**, 31 (2021).
- Holmdahl, R., Malmström, V. & Burkhardt, H. Autoimmune priming, tissue attack and chronic inflammation - the three stages of rheumatoid arthritis. *Eur. J. Immunol.* **44**, 1593–1599 (2014).
- Rantapää-Dahlqvist, S., Boman, K., Tarkowski, A. & Hallmans, G. Up regulation of monocyte chemoattractant protein-1 expression in anti-citrulline antibody and immunoglobulin M rheumatoid factor positive subjects precedes onset of inflammatory response and development of overt rheumatoid arthritis. *Ann. Rheum. Dis.* **66**, 121–123 (2007).
- Kurki, P., Aho, K., Palosuo, T. & Heliövaara, M. Immunopathology of rheumatoid arthritis. *Antikeratin antibodies Preced. Clin. Dis. Arthritis Rheum.* **35**, 914–917 (1992).
- Kerschbaumer, A. et al. Efficacy of synthetic and biological DMARDs: a systematic literature review informing the 2022 update of the EULAR recommendations for the management of rheumatoid arthritis. *Ann. Rheum. Dis.* <https://doi.org/10.1136/ard-2022-223365> (2022).
- Sergeant, J. C. et al. Prediction of primary non-response to methotrexate therapy using demographic, clinical and psychosocial variables: results from the UK Rheumatoid Arthritis Medication Study (RAMS). *Arthritis Res. Ther.* **20**, 147 (2018).
- Anderson, J. J., Wells, G., Verhoeven, A. C. & Felson, D. T. Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration. *Arthritis Rheum.* **43**, 22–29 (2000).
- Gul, H. L. et al. Can biomarkers predict successful tapering of conventional disease-modifying therapy in rheumatoid arthritis patients in stable remission?. *Clin. Exp. Rheumatol.* **41**, 126–136 (2023).
- Sun, J. et al. Identification of novel protein biomarkers and drug targets for colorectal cancer by integrating human plasma proteome with genome. *Genome Med.* **15**, 75 (2023).
- Oh, H. S. et al. Organ aging signatures in the plasma proteome track health and disease. *Nature* **624**, 164–172 (2023).
- Davies, M. P. A. et al. Plasma protein biomarkers for early prediction of lung cancer. *EBioMedicine* **93**, 104686 (2023).
- Mazidi, M. et al. Plasma proteomics to identify drug targets for ischemic heart disease. *J. Am. Coll. Cardiol.* **82**, 1906–1920 (2023).
- O’Neil, L. J. et al. Association of a serum protein signature with rheumatoid arthritis development. *Arthritis Rheumatol.* **73**, 78–88 (2021).
- Hu, C. et al. Proteome profiling identifies serum biomarkers in rheumatoid arthritis. *Front. Immunol.* **13**, 865425 (2022).
- Maciejewski, M. et al. Prediction of response of methotrexate in patients with rheumatoid arthritis using serum lipidomics. *Sci. Rep.* **11**, 7266 (2021).
- Dervieux, T. et al. Pharmacogenetic and metabolite measurements are associated with clinical status in patients with rheumatoid arthritis treated with methotrexate: results of a multicentred cross sectional observational study. *Ann. Rheum. Dis.* **64**, 1180–1185 (2005).
- Xu, K. et al. Clinical markers combined with HMGB1 polymorphisms to predict efficacy of conventional DMARDs in rheumatoid arthritis patients. *Clin. Immunol.* **221**, 108592 (2020).
- Tasaki, S. et al. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. *Nat. Commun.* **9**, 2755 (2018).
- Paul, B. J., Kandy, H. I. & Krishnan, V. Pre-rheumatoid arthritis and its prevention. *Eur. J. Rheumatol.* **4**, 161–165 (2017).
- Okamoto, Y. et al. Association of sputum neutrophil extracellular trap subsets with IgA anti-citrullinated protein antibodies in subjects at risk for rheumatoid arthritis. *Arthritis Rheumatol.* **74**, 38–48 (2022).
- Wehr, P., Purvis, H., Law, S. C. & Thomas, R. Dendritic cells, T cells and their interaction in rheumatoid arthritis. *Clin. Exp. Immunol.* **196**, 12–27 (2019).
- Kovács, O. T. et al. Proteomic changes of osteoclast differentiation in rheumatoid and psoriatic arthritis reveal functional differences. *Front. Immunol.* **13**, 892970 (2022).
- James, J. et al. Redox regulation of PTPN22 affects the severity of T-cell-dependent autoimmune inflammation. *Elife* **11**, <https://doi.org/10.7554/eLife.74549> (2022).
- Miglioranza Scavuzzi, B. & Holoshitz, J. Endoplasmic reticulum stress, oxidative stress, and rheumatic diseases. *Antioxidants* **11**, <https://doi.org/10.3390/antiox11071306> (2022).
- Lai, H. C., Ho, U. Y., James, A., De Souza, P. & Roberts, T. L. RNA metabolism and links to inflammatory regulation and disease. *Cell Mol. Life Sci.* **79**, 21 (2021).
- Iwamoto, N. et al. Osteogenic differentiation of fibroblast-like synovial cells in rheumatoid arthritis is induced by microRNA-218 through a ROBO/Slit pathway. *Arthritis Res. Ther.* **20**, 189 (2018).
- Hao, R. et al. Identification of dysregulated genes in rheumatoid arthritis based on bioinformatics analysis. *PeerJ* **5**, <https://doi.org/10.7717/peerj.3078> (2017).
- Ferreira, M. B. et al. Sex differences in circulating proteins of patients with rheumatoid arthritis: a cohort study. *Int J. Rheum. Dis.* **25**, 669–677 (2022).
- Nakanishi, Y., Kang, S. & Kumanogoh, A. Crosstalk between axon guidance signaling and bone remodeling. *Bone* **157**, 116305 (2022).
- Arend, W. P. & Firestein, G. S. Pre-rheumatoid arthritis: predisposition and transition to clinical synovitis. *Nat. Rev. Rheumatol.* **8**, 573–586 (2012).
- Muchamuel, T. et al. A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. *Nat. Med.* **15**, 781–787 (2009).
- Dragoljevic, D. et al. Defective cholesterol metabolism in haematopoietic stem cells promotes monocyte-driven atherosclerosis in rheumatoid arthritis. *Eur. Heart J.* **39**, 2158–2167 (2018).
- O’Neil, L. J. et al. Neutrophil extracellular trap-associated carbamylation and histones trigger osteoclast formation in rheumatoid arthritis. *Ann. Rheum. Dis.* **82**, 630–638 (2023).
- Yang, L. et al. Auranofin mitigates systemic iron overload and induces ferroptosis via distinct mechanisms. *Sig. Transduct. Target Ther.* **5**, 138 (2020).
- He, Y. et al. A subset of antibodies targeting citrullinated proteins confers protection from rheumatoid arthritis. *Nat. Commun.* **14**, 691 (2023).
- Li, T. et al. Pathogenic antibody response to glucose-6-phosphate isomerase targets a modified epitope uniquely exposed on joint cartilage. *Ann. Rheum. Dis.* **82**, 799–808 (2023).

39. Arts, E. E., Fransen, J., Den Broeder, A. A., van Riel, P. & Popa, C. D. Low disease activity (DAS28 \leq 3.2) reduces the risk of first cardiovascular event in rheumatoid arthritis: a time-dependent Cox regression analysis in a large cohort study. *Ann. Rheum. Dis.* **76**, 1693–1699 (2017).
40. Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
41. Fransen, J. & van Riel, P. L. The disease activity score and the EULAR response criteria. *Rheum. Dis. Clin. North Am.* **35**, 745–757 (2009). vii–viii.
42. Saevarsdottir, S. et al. Predictors of response to methotrexate in early DMARD naive rheumatoid arthritis: results from the initial open-label phase of the SWEFOT trial. *Ann. Rheum. Dis.* **70**, 469–475 (2011).
43. Lei, Q. et al. Lipid metabolism and rheumatoid arthritis. *Front. Immunol.* **14**, 1190607 (2023).
44. Garcia-Carbonell, R. et al. Critical role of glucose metabolism in rheumatoid arthritis fibroblast-like synoviocytes. *Arthritis Rheumatol.* **68**, 1614–1626 (2016).
45. Xu, L. et al. Metabolomics in rheumatoid arthritis: advances and review. *Front. Immunol.* **13**, 961708 (2022).
46. Rahmati, M., Moosavi, M. A. & McDermott, M. F. ER stress: a therapeutic target in rheumatoid arthritis?. *Trends Pharm. Sci.* **39**, 610–623 (2018).
47. Ajeganova, S. & Huizinga, T. W. Rheumatoid arthritis: seronegative and seropositive RA: alike but different?. *Nat. Rev. Rheumatol.* **11**, 8–9 (2015).
48. van den Broek, M. et al. The association of treatment response and joint damage with ACPA-status in recent-onset RA: a subanalysis of the 8-year follow-up of the BeSt study. *Ann. Rheum. Dis.* **71**, 245–248 (2012).
49. Petrovska, N., Prajzlerova, K., Vencovsky, J., Senolt, L. & Filkova, M. The pre-clinical phase of rheumatoid arthritis: From risk factors to prevention of arthritis. *Autoimmun. Rev.* **20**, 102797 (2021).
50. Zhang, Y. & Lee, T. Y. Revealing the immune heterogeneity between systemic lupus erythematosus and rheumatoid arthritis based on multi-omics data analysis. *Int. J. Mol. Sci.* **23**, <https://doi.org/10.3390/ijms23095166> (2022).
51. Jankovic, M., Casellas, R., Yannoutsos, N., Wardemann, H. & Nussenzweig, M. C. RAGs and regulation of autoantibodies. *Annu. Rev. Immunol.* **22**, 485–501 (2004).
52. Felson, D. T. et al. American College of Rheumatology/European League against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Ann. Rheum. Dis.* **70**, 404–413 (2011).
53. Smolen, J. S. et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann. Rheum. Dis.* **79**, 685–699 (2020).
54. Deng, D. et al. Leflunomide monotherapy versus combination therapy with conventional synthetic disease-modifying antirheumatic drugs for rheumatoid arthritis: a retrospective study. *Sci. Rep.* **10**, 12339 (2020).
55. Kremer, J. M. et al. Concomitant leflunomide therapy in patients with active rheumatoid arthritis despite stable doses of methotrexate. A randomized, double-blind, placebo-controlled trial. *Ann. Intern. Med.* **137**, 726–733 (2002).
56. van de Stadt, L. A. et al. The extent of the anti-citrullinated protein antibody repertoire is associated with arthritis development in patients with seropositive arthralgia. *Ann. Rheum. Dis.* **70**, 128–133 (2011).
57. Xu, Z. et al. Integrative proteomics and N-glycoproteomics analyses of rheumatoid arthritis synovium reveal immune-associated glycopeptides. *Mol. Cell Proteom.* **22**, 100540 (2023).
58. Gerlag, D. M. et al. EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. *Ann. Rheum. Dis.* **71**, 638–641 (2012).
59. Radner, H., Neogi, T., Smolen, J. S. & Aletaha, D. Performance of the 2010 ACR/EULAR classification criteria for rheumatoid arthritis: a systematic literature review. *Ann. Rheum. Dis.* **73**, 114–123 (2014).
60. Yu, Q. et al. Sample multiplexing for targeted pathway proteomics in aging mice. *Proc. Natl Acad. Sci. USA* **117**, 9723–9732 (2020).
61. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).
62. Gong, Y. et al. Acetylation profiling by Iseq-Kac reveals insights into HSC aging and lineage decision. *Nat. Chem. Biol.*, <https://doi.org/10.1038/s41589-025-01916-1> (2025).
63. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
64. Deutsch, E. W. et al. Advances and Utility of the Human Plasma Proteome. *J. Proteome Res.* **20**, 5241–5263 (2021).
65. Uhlen, M. et al. The human secretome. *Sci. Signal* **12**, <https://doi.org/10.1126/scisignal.aaz0274> (2019).
66. Uhlen, M. et al. A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **366**, <https://doi.org/10.1126/science.aax9198> (2019).
67. Ashton, N. J. et al. A plasma protein classifier for predicting amyloid burden for preclinical Alzheimer's disease. *Sci. Adv.* **5**, eaau7220 (2019).
68. Eldjarn, G. H. et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).
69. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
70. Qu, Y. et al. Plasma proteomic profiling discovers molecular features associated with upper tract urothelial carcinoma. *Cell Rep. Med.*, 101166, <https://doi.org/10.1016/j.xcrm.2023.101166> (2023).
71. Niu, L. et al. Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat. Med.* **28**, 1277–1287 (2022).
72. Balakrishnan, R., Harris, M. A., Huntley, R., Van Auken, K. & Cherry, J. M. A guide to best practices for Gene Ontology (GO) manual annotation. *Database* **2013**, bat054–bat054 (2013).
73. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
74. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
75. Liang, L. et al. Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women. *Cell* **181**, 1680–1692.e1615 (2020).
76. Tibshirani, R. J. The lasso problem and uniqueness. *Electron J. Stat.* **7**, 1456–1490 (2012).
77. Yuan, M. & Lin, Y. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**, 49–67 (2005).
78. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma.* **12**, 77 (2011).
79. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
80. Hofner, B., Boccuto, L. & Göker, M. Controlling false discoveries in high-dimensional situations: boosting with stability selection. *BMC Bioinforma.* **16**, 144 (2015).

81. Tay, J. K., Narasimhan, B. & Hastie, T. Elastic net regularization paths for all generalized linear models. *J. Stat. Softw.* **106**, <https://doi.org/10.18637/jss.v106.i01> (2023).
82. Andy, L. Classification and regression by randomForest. *R. N.* **2**, 18 (2002).
83. Yan, Z. et al. XGBoost algorithm and logistic regression to predict the postoperative 5-year outcome in patients with glioma. *Ann. Transl. Med.* **10**, 860 (2022).
84. Covert, I., Kim, C. & Lee, S.-I. Learning to estimate Shapley values with vision transformers. In *Proc. International Conference on Learning Representations* (2023).
85. Alvez, M. B. et al. Next generation pan-cancer blood proteome profiling using proximity extension assay. *Nat. Commun.* **14**, 4308 (2023).
86. Schisterman, E. F., Perkins, N. J., Liu, A. & Bondell, H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology* **16**, 73–81 (2005).
87. Chen, T. et al. iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res.* **50**, D1522–D1527 (2022).
88. Ma, J. et al. iProX: an integrated proteome resource. *Nucleic Acids Res.* **47**, D1211–D1217 (2019).
89. He, S. A longitudinal cohort study uncovers plasma protein biomarkers predating clinical onset and treatment response of rheumatoid arthritis. *Zenodo*, <https://doi.org/10.5281/zenodo.15717988> (2025).

Acknowledgements

This work was supported by the National Key Research and Development Program of China (Nos. 2019YFE0108200 (Y.Z.)), Transformation Foundation of Tianfu Jincheng Laboratory (No. 2025ZH024 (L.D.)), Sichuan International Science and Technology Cooperation Project (Nos. 2022YFH0023 (Y.Z.), 2024YFFK0099 (L.D.), 2024YFHZ0231 (Y.Z.) and 2024JDHJ0044 (Y.Z.)), Science Popularization Base Project of Chengdu Science and Technology Bureau (No. 2022-GH03-00003-HZ (Y.Z.)), West China Hospital-Enterprise Cooperation Clinical Research Innovation Project (No. 21HXCX004 (Y.Z.)), the National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University (No. Z2024JC002 (L.D.)), and West China Hospital 135 project (Nos. ZYYC23013 (L.D.) and 21HXFH002 (Y.Z.)).

Author contributions

L.D. and Y.Z. designed the project and wrote the manuscript. S.H. and C.Z. analyzed the omics data and wrote the paper. Y.L. and R.S. col-

lected the samples and carried out the follow-ups. Z.X. contributed to the omics data acquisition. B.Y. and X.G. assisted in collecting the samples and verifying the clinical information. I.G., R.H., M.H. and L.E.M. revised the manuscript and provided helpful suggestions.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-62032-1>.

Correspondence and requests for materials should be addressed to Lunzhi Dai or Yi Zhao.

Peer review information *Nature Communications* thanks Liam O’Neil, Vincenzo Venerito and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025