Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrated multi‐omics defines immune programmes underlying progression to rheumatoid arthritis in anti-citrullinated protein antibody positive individuals with arthralgia and palindromic rheumatism.

Leon Chang
University of Leeds https://orcid.org/0009-0003-5223-7616
Fareeha Tariq
Department of Rheumatology, King's College Hospital, London, UK https://orcid.org/0009-0001-0194-2927
Amreen Azeem
University of Leeds
Daniel Clayton
University of Leeds
Kierran Flack
University of Leeds
Thomas Young
University of Leeds https://orcid.org/0009-0003-9081-4920
Alice Ibbtoson
University of Leeds
Ana Martinez-Rodriguez
University of Leeds
Samuel Lara-Reyna
University of Leeds https://orcid.org/0000-0002-9986-5279
Aisha Aman
University of Leeds
Emilia Cross
University of Leeds
Laurence Duquenne
University of Leeds
Andrea Di Matteo
University of Leeds
Paul Emery

Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds
https://orcid.org/0000-0002-7429-8482

Kulveer Mankia

National Institute for Health Research Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust

Sinisa Savic

S.Savic@leeds.ac.uk

University of Leeds https://orcid.org/0000-0001-7910-0554

Article

Keywords:

Posted Date: March 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9108192/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Integrated multi-omics defines immune programmes underlying progression to rheumatoid arthritis in anti-citrullinated protein antibody positive individuals with arthralgia and palindromic rheumatism.

Authors:
Leon Chang[1], Fareeha Tariq[1,2], Amreen Azeem[1], Daniel Clayton[1], Kierran Flack[1], Thomas Young[1], Alice Ibbotson[1], Ana Martinez-Rodriguez[1], Samuel Lara-Reyna[1], Aisha Aman[1]; Emilia Cross[1], Laurence Duquenne[1], Andrea Di Matteo[1], Paul Emery[1,3], Kulveer Mankia[1,3*], Sinisa Savic[1,3*].

[1]University of Leeds, Leeds Institute of Rheumatic and Musculoskeletal Medicine, Leeds, United Kingdom
[2]King's College Hospital NHS Foundation Trust, London, United Kingdom
[3]National Institute for Health and Care Research (NIHR), Leeds Biomedical Centre, Leeds, United Kingdom

*Equal contributions

Correspondence: Sinisa Savic, Leeds Institute for Rheumatic and Musculoskeletal Medicine. Email: s.savic@leeds.ac.uk

Abstract

Rheumatoid arthritis (RA) develops through a prolonged preclinical phase in which systemic immune dysregulation precedes joint inflammation. To clarify the molecular programmes driving divergent outcomes in anti-citrullinated protein antibody (ACPA) positive individuals, we performed integrated multi-omics profiling of 192 ACPA+ at-risk individuals, including those with arthralgia and palindromic rheumatism (PR). Analyses across genomic, transcriptomic, proteomic, and functional modalities revealed broad innate-immune activation relative to healthy controls. Genetic risk factors, including shared epitope alleles and rare germline variants in inflammation-related pathways, distinguished future progressors from non-progressors, while clonal haematopoiesis contributed to a latent factor associated with progression. Multi-omics factor analysis identified an innate-dominant immune programme associated with delayed progression, with adaptive-immune interactions accelerating the transition to RA. Distinct interferon-enriched and cytokine-driven latent factors characterised PR, indicating unique immune programmes associated with this episodic phenotype, even when measured between flares. These findings define coordinated immune programmes in ACPA+ individuals with arthralgia and PR that precede the development of clinical RA, providing a framework for biomarker development and targeted RA prevention strategies.

Introduction

Rheumatoid arthritis (RA) is a chronic, immune-mediated inflammatory disease characterised by persistent synovitis, progressive joint destruction, and extra-articular manifestations. Regarded as a prototypical autoimmune disorder, it is increasingly recognised as a biologically heterogenous state that evolves through distinct phases, spanning preclinical autoimmunity and including later autoinflammatory amplification[1]. RA pathogenesis is a multistage process that unfolds over years, long before clinical symptoms emerge. A synergistic interplay between genetic risk factors conferred by HLA-DRB1 shared epitope (SE) alleles and non-HLA SNPs in genes such as *PTPN22, CTLA4*, and *STAT4* and environmental exposures such smoking, mucosal dysbiosis, and diet are believed to induce post-translational citrullination and carbamylation of self-proteins generating autoantigens that breach immune tolerance[2-4]. The ensuing loss of tolerance leads to generation of autoantibodies such as anti-citrullinated protein antibodies (ACPA) and rheumatoid factors (RF) which mark the earliest serological evidence of systemic autoimmunity that precedes clinical symptoms and the development of arthritis[5].

During this pre-arthritis phase, ACPA and RF are believed to form immune complexes that activate complement and Fc-receptor-bearing myeloid cells, amplifying the release of inflammatory cytokines such as tumour necrosis factor (TNF), interleukin-6 (IL-6), and interleukin-1β (IL-1β). This cytokine milieu often correlates with symptoms of arthralgia, stiffness and fatigue, which characterise the so-called *at-risk* phase of RA, and reflects systemic autoimmunity in the absence of clinically evident synovitis[6-10]. Within this continuum, some individuals manifest palindromic rheumatism (PR), characterised by intermittent but self-resolving flares of joint pain, swelling, and erythema, with no residual joint damage[11]. PR shares genetic and serological risk factors with RA, including ACPA and HLA-DRB1 SE alleles, yet differs in its inflammatory pathotype, with evidence implicating a role of inflammasome activation and MEFV gene variants, particularly in ACPA-negative individuals[12,13]. Comparative genotyping of polymorphisms further supports this divergence with distinct allele prevalence in immune regulatory loci in *TNFAIP3*, *IL10*, and *IRF5* between RA and PR[14]. Despite these distinctions, the autoimmune and autoinflammatory pathways that drive the transition from systemic autoimmunity to clinical arthritis in both ACPA+ arthralgia and PR remain incompletely understood.

Somatic mutations are increasingly recognised as drivers of acquired autoinflammatory phenotypes that intersect with the rheumatic disease spectrum. VEXAS (vacuoles, E1 enzyme, X-linked, autoinflammatory, somatic) syndrome caused by *UBA1* mutations in haematopoietic progenitors and acquired NLRP3-associated autoinflammatory disease exemplify the role of somatic mosaicism in dysregulation of innate immune homeostasis. These mutations induce sustained IL-1β, NF-kB, and type I interferon (IFN) activation within myeloid lineages, leading to cytopenia and arthritis[15-18]. Beyond these rare syndromic presentations, clonal haematopoiesis-associated (CH) somatic mutations that arise in haematopoietic regulator genes such as *DNMT3A*, *TET2*, and *ASXL1,* accumulate with increasing age and are also overrepresented in established RA compared with age-matched controls[19-21]. Mechanistically, CH promote sustained

proinflammatory milieu through epigenetic reprogramming of myeloid progenitors, leading to exaggerated NF-kB signalling, NLRP3 inflammasome activation and enhanced production of IL-1β and IL-6[22,23]. Of note, these mutations also overlap with those seen in myelodysplastic syndromes, suggesting a shared clonal architecture in linking aberrant haematopoiesis with chronic inflammation[24].

Based on these observations, we hypothesised that molecular perturbations characteristic of established RA arises in at-risk individuals before the onset of clinical arthritis, and that somatic mutations within the haematopoietic and auto-inflammatory pathways may contribute towards disease progression. To address this, we employed an integrated multi-omics approach to characterise the mutational landscape and its downstream impact on global transcriptomic, proteomic and immunological profiles in ACPA+ individuals with arthralgia and PR who are at-risk of progressing to RA.

Results

Patient cohort description

Our study consists of samples collected from 192 at-risk individuals who have tested positive for anti-cyclic citrullinated peptide (anti-CCP+) antibodies – the clinical assay used to detect ACPA. These individuals were prospectively recruited prior to onset of inflammatory arthritis, of whom 98 cases had PR. Progression to clinical inflammatory arthritis occurred in 51% (49/94) of ACPA+ individuals with arthralgia and 49% (48/98) of PR cases, monitored over a 110-month period. The male-female ratio was approximately 1:2 (65 Male to 127 Female) and the mean cohort age was 50.1 years at the time of recruitment. A summary of the cohort can be found in Supplementary Table S1.

Unimodal signatures in at-risk individuals

Data was collected for seven molecular modalities—HLA genotypes, rare germline variant burden, somatic CHIP prevalence, peripheral blood transcriptomics, circulating inflammatory cytokines, IFN protein scores, and inflammasome activity. Each modality was analysed independently to characterise modality-specific patterns of variation within the cohort prior to integration.

Inherited genetic risk factors

HLA-DRB1 genotypes were inferred from DNA samples available for 183 of the 192 at-risk individuals. The distribution of SE dosage differed significantly by progression status (χ^2=9.72, p=0.008, Figure 1A), with SE homozygosity and HLA-DRB1*04:01 alleles enriched among progressors compared with non-progressors (47% vs 20%, FDR<0.002). To assess broader genetic predisposition beyond HLA risk, we investigated rare germline variation across the exome and identified rare, potentially pathogenic variants within the cohort (variant inclusion criteria defined in methods).

Rare germline variant burden was quantified at the gene level in areas of biological relevance using (i) a virtual panel of RA-associated genes and (ii) an expanded discovery gene panel associated with inflammation and interferonopathy (Supplementary Table S2), defined as the number of genes per individual harbouring at least one qualifying variant. Within the ACPA+ arthralgia group, rare variant burden differed significantly with

progression status in the discovery gene panel, with more than half of ACPA+ arthralgia future progressors harbouring at least one qualifying variant (Fisher's exact test, p=0.0096; Figure 1B). This suggests a genetically primed inflammatory background within this subset of individuals. In contrast, no enrichment was observed in the RA-associated gene panel and in patients with PR.

Two variants, *ADAM17* c.2017G>A (p.(Val673Ile)) and *ALPK1* c.200A>G (p.(Gln67Arg)) were significantly enriched in ACPA+ progressors with arthralgia, observed in four individuals each and absent in all non-progressors (p=0.0088). In-silico predictions suggest deleterious/damaging effects, but both variants are currently classified as likely benign/benign in ClinVar and are therefore unlikely to be disease-causing. However, *ADAM17* and *ALPK1* have been implicated in RA and osteoarthritis pathogenesis respectively[25,26], and their enrichment in future progressors suggests they may still contribute as potential risk factors or disease modifiers. All qualifying discovery panel variants identified in the ACPA+ progressors with arthralgia are provided in Supplementary table S3. Rare germline variant burden within these gene-panels were used as input for downstream multi-omics integration.

Exploratory analysis of all ACPA+ individuals in an exome-wide, panel-agnostic manner identified seven recurrent missense germline variants located in *COG1*, *COL4A3*, *DCDC2B*, *DNAH12*, *HAVCR1*, *HEATR5A*, and *SRRM2*, each present in eight or more individuals (Supplementary Table S4). Among these genes, *HAVCR1*, *HEATR5A* and *SRRM2* have reported associations with autoimmune or inflammatory phenotypes[27-29], whereas the remainder currently lack clear biological links to autoimmunity.

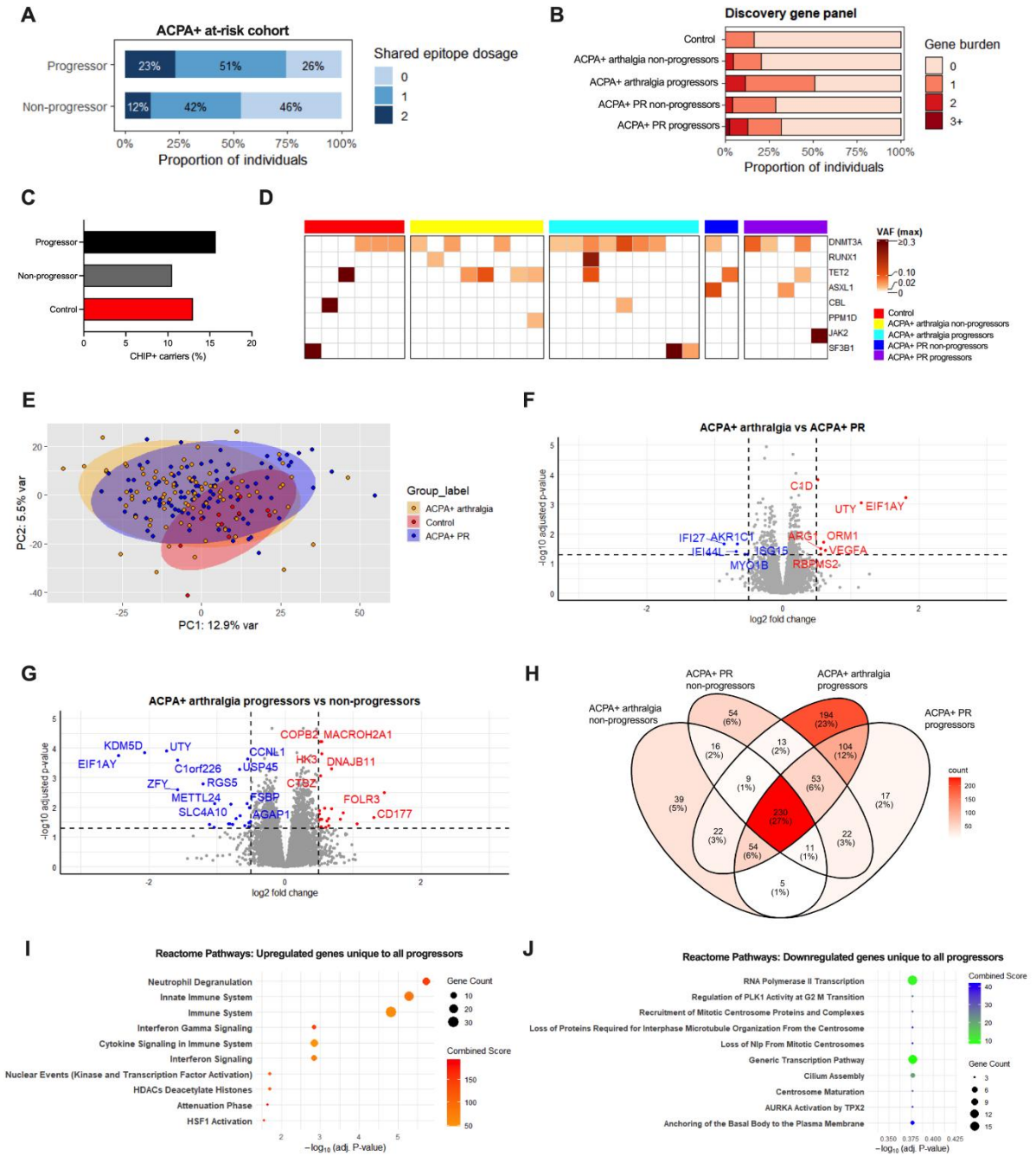


Figure 1. Summary of immunogenetic, mutational, transcriptomic, and functional differences across clinical phenotypes. *Genomic profiling was carried out on 183 at-risk individuals (ACPA+ arthralgia non-progressors n=44; ACPA+ arthralgia progressors n=43; ACPA+ PR non-progressors n=49 and ACPA+ PR progressors n=47) and 49 controls. (A) Shared epitope dosage distribution showing enrichment of SE homozygosity in progressors. (B) Rare germline variant burden within the discovery gene panel across study groups. (C) CH prevalence in age-matched progressors, non-progressors, and controls. (D) Somatic variant patterns across CH driver genes, with each column representing an individual. Bulk RNA-seq was carried out on 191 at-risk individuals (ACPA+ arthralgia non-progressors n=45; ACPA+ arthralgia progressors n=48; ACPA+ PR non-progressors n=50 and ACPA+ PR progressors n=48) and 11 controls. (E) PCA of ACPA+ arthralgia, ACPA+ PR and control transcriptomes corrected for sex. (F) Volcano*

plot providing an overview of differentially expressed genes identified between the ACPA+ arthralgia and PR groups. (G) Volcano plot providing an overview of differentially expressed genes identified between the ACPA+ progressors and non-progressors with arthralgia. (H) Venn diagram illustrating the quantity of differentially expressed genes intersecting from each ACPA+ group versus healthy controls comparison. (I) Reactome pathway enrichment analysis for the 104 genes differentially expressed in both ACPA+ progressor groups relative to HCs, highlighting innate-immune and IFN-mediated signatures among upregulated genes.

Clonal haematopoiesis-associated somatic variants

CH prevalence was assessed using a conservative definition restricted to 12 established canonical driver genes (*DNMT3A*, *TET2*, *ASXL1*, *PPM1D*, *SF3B1*, *TP53*, *SRSF2*, *JAK2*, *IDH1*, *IDH2*, *CBL*, *RUNX1*) and a ≥2% variant allele fraction (VAF) threshold[30,31]. Compared to non-progressors, at-risk progressors (ACPA+ arthralgia and PR combined) exhibited a modest 5.2% higher CH prevalence, although this difference did not reach statistical significance (Figure 1C). Notably, progressors from the ACPA+ arthralgia group had the highest CH prevalence at 20.9%, exceeding estimates reported in age-matched populations (mean 55.8 years of age)[21]. This was followed by ACPA+ arthralgia non-progressors (16.7%), ACPA+ PR progressors (10.9%) and PR non-progressors (4.3%).

Independent of progression status, CH prevalence was significantly higher among ACPA+ individuals with arthralgia than to those with PR, despite comparable mean ages (18.7% vs 7.5%, p = 0.025; 52.3 vs 47 years, Figure 1D). This pattern suggests that clonal myeloid priming may play a more prominent role in driving RA-like innate inflammatory states than to episodic PR phenotypes. The longer duration of subclinical inflammation observed in the ACPA+ arthralgia group may additionally contribute to their increased CH prevalence. Outside of CH driver genes, no enrichment of somatic variants was observed in genes associated with autoinflammatory conditions.

Peripheral blood transcriptomics

Principal component analysis showed substantial overlap in transcriptomic profiles between ACPA+ individuals with arthralgia and those with PR (Figure 1E), with only 13 genes differentially expressed between the groups. Among these, several IFN-stimulated genes—*IFI27*, *IFI44L*, and *ISG15*—displayed reduced expression in the ACPA+ arthralgia group compared to individuals with PR (Figure 1F). When examining gene expression by progression status within each clinical stratum, we observed modest transcriptional alterations in the ACPA+ arthralgia cohort, identifying 46 differentially expressed genes between progressors and non-progressors (Figure 1G), and only three such genes in the PR group.

In contrast, more pronounced transcriptional differences were observed when each ACPA+ group was compared to healthy controls (HCs). The intersection of each comparison identified 819 genes (230 upregulated; 589 downregulated) that were consistently differentially expressed across all ACPA+ groups relative to HCs. Notably, 104 genes were upregulated in both ACPA+ arthralgia and PR progressors and absent in

non-progressors (Figure 1H). Subsequent enrichment analysis of this gene set demonstrated significant overrepresentation in innate immune pathways, including neutrophil-associated, interferon-responsive, and cytokine-mediated processes (Figure 1I). This indicates that despite clinical differences, ACPA+ future progressors with arthralgia and PR share an innate-immune transcriptional programme, prior to the onset of inflammatory arthritis. This analysis was repeated for the genes downregulated in both progressor groups, but no significant pathway enrichment was observed (Figure 1J).

Inflammatory cytokine expression

Serum protein expression of 45 cytokines relevant to inflammatory processes, immune responses and related diseases was assessed in all study groups. Heatmap visualisation shows distinct elevation of cytokines in both progressor phenotypes, with modest heterogeneity across PR status (Figure 2A). Compared to controls, IL6, HGF and OSM concentrations were significantly elevated in all progressor groups. Within clinical stratum, CXCL9, CXCL11, IL6, IL10 and MMP12 concentrations were significantly elevated with progression status in the ACPA+ arthralgia group. Whilst CXCL9, CXCL10 and IL10 were significantly elevated in ACPA+ PR progressors when compared to PR non-progressors.

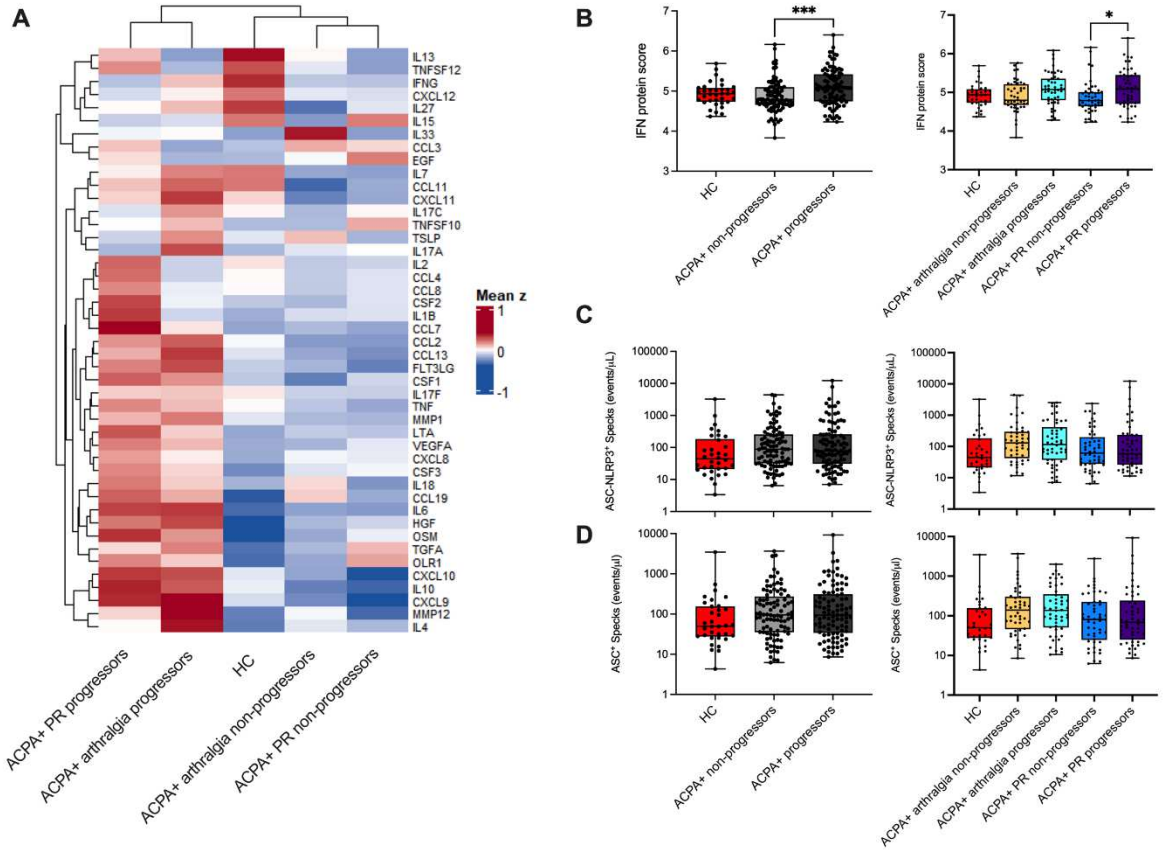


Figure 2. Cytokine profiles, IFN activity, and ASC responses across clinical groups. *Inflammatory cytokine profiling, chemokine-derived IFN scoring and measurements of ASC specks inflammasome activity was carried out on 191 at-risk individuals (ACPA+ arthralgia non-progressors n=46; ACPA+ arthralgia progressors n=48; ACPA+ PR non-progressors n=49 and ACPA+ PR progressors n=48) and controls (n=35). (A) Heatmap showing standardized (z-score) cytokine and chemokine expression across clinical groups, with hierarchical clustering of both*

*samples and analytes. (B) Boxplots comparing IFN protein scores between healthy controls, non-progressors, and progressors (left), and across extended phenotypes (right). (C) Boxplots illustrating ASC-NLRP3+ speck events/µL and (D) non-specific ASC+ speck events/µL between healthy controls, non-progressors, and progressors (left), and across extended phenotypes (right). Significant differences are indicated using asterisks; *** denotes p<0.001; * denotes p<0.05.*

Chemokine-derived IFN scores

Serum IFN scores derived from chemokine concentrations (CCL8, CXCL10, CCL19, CCL2, CXCL11) differed significantly between progressors, non-progressors, and HCs (ANOVA p=0.0011). Post-hoc comparisons indicated higher mean IFN scores in progressors when compared to non-progressors (p=0.0009, Figure 2B). When stratified by PR status, significantly elevated IFN scores were seen only in PR-positive individuals (p = 0.025).

Inflammasome activity

Inflammasome activity was assessed by quantifying circulating apoptosis-associated speck-like protein containing a CARD (ASC) speck, including both NLRP3-bound and non-specific ASC assemblies. The effect sizes for both ASC-NLRP3 and non-specific ASC specks were modest across the cohort and did not reach statistical significance, but trends were directionally consistent with enhanced inflammasome activity in progressors (Figure 2C-D).

Stratification by PR status revealed a significant difference in ASC speck distributions across study groups (Kruskal-Wallis, p=0.042); however, post hoc pairwise comparisons were not statistically significant after correction for multiple testing. Collectively, these results are consistent with low grade inflammasome-mediated activity in at-risk individuals, but the modest effect sizes limit the utility of ASC specks as a standalone preclinical biomarker.

Germline burden-Phenotype association analysis

As an exploratory analysis, rare germline variant burden within the discovery panel was evaluated in relation to downstream immune phenotypes to assess whether cumulative germline variation was associated with measurable immune perturbation. Burden was correlated to inflammasome activity (ASC-NLRP3 specks, non-specific ASC specks), IFN protein scores, and cytokines previously identified as altered with progression status (CXCL9, IL6, CXCL10, IL10, MMP12, CXCL11). Spearman correlation coefficients were uniformly low for all variables $|\rho| < 0.12$ and none reached statistical significance (data not shown).

These findings suggest that rare germline variation within the discovery panel does not independently associate with steady-state downstream functional immune measures using univariate correlation approaches. We therefore applied a multi-omic integration framework to assess whether germline variation may contribute to immune states in a higher-dimensional context.

Multi-omics integration identifies latent immune programmes

To investigate coordinated immune variation across the seven unimodal datasets, we performed an integrative analysis using multi omics factor analysis (MOFA). This approach allowed us to identify shared axes of variation spanning transcriptional, proteomic, cytokine, inflammasome, and genetic features across the 192 anti CCP+ at risk individuals (Figure 3A). The contribution of each data modality to the integrated dataset is shown in Figure 3B, highlighting balanced sample availability and enabling cross-modality comparisons.

After model training, the MOFA framework identified ten latent factors, each representing a distinct, data driven biological programme capturing covariation across multiple immune layers (Figure 3C). Inspection of variance decomposition revealed that the total variance captured by the MOFA model was predominantly explained by peripheral blood gene expression (33.9%) and interferon (IFN) protein scores (21.2%, Figure 3D). Inflammasome activity (15.2%) and inflammatory cytokine profiles (13.1%) formed secondary but substantial contributors to the integrated structure. Genetic modalities, including somatic CH prevalence, HLA-DRB1 genotypes, and rare germline variant burden, collectively explained a smaller proportion of total variance (<5% each) (Figure 3D).

Together, these ten MOFA-derived factors provide a coordinated representation of multi-layered immune biology in individuals at risk of RA progression, forming the basis for downstream associations with clinical outcomes and progression risk stratification.

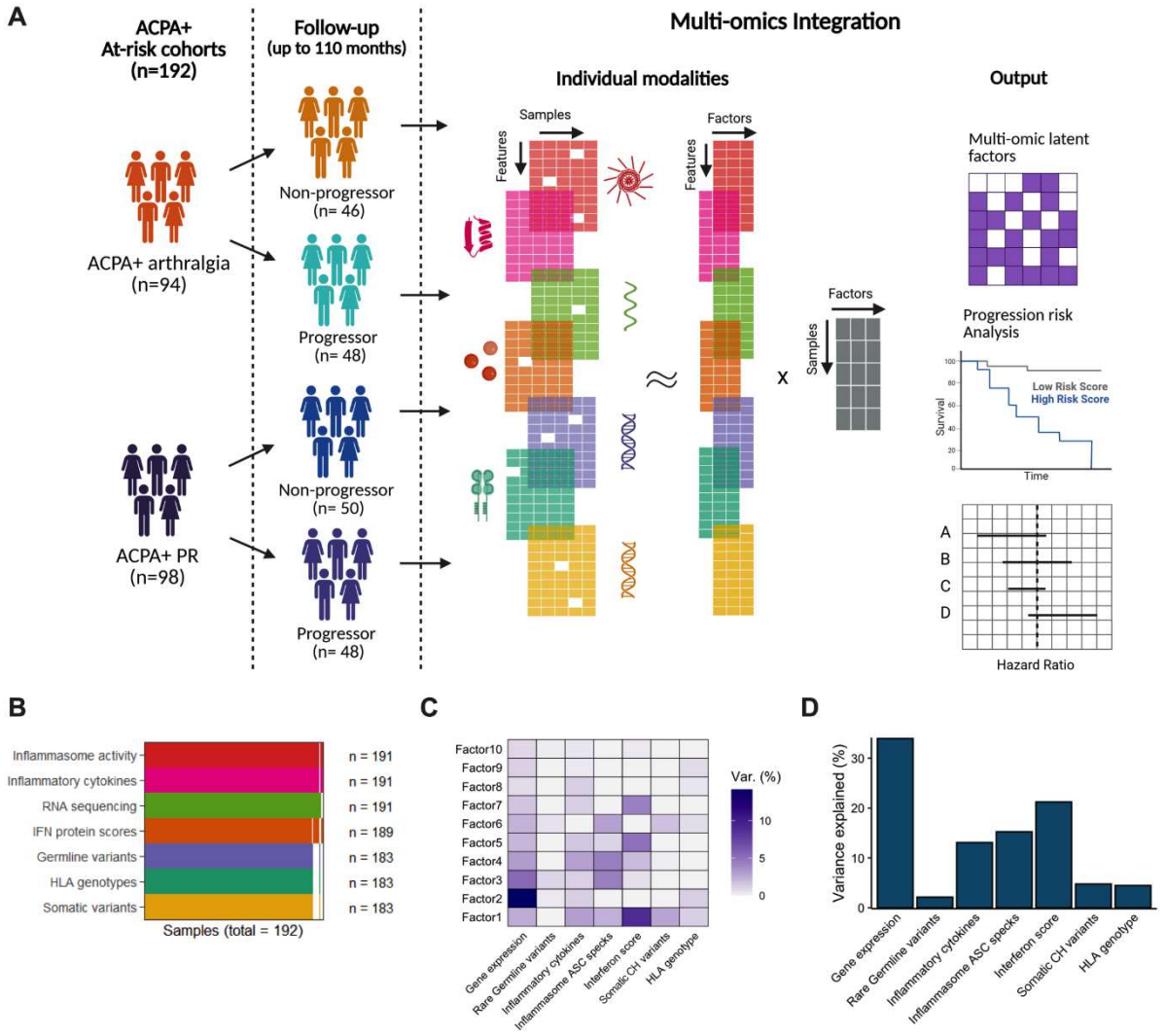


Figure 3. Study design, multi-omics integration and variance structure of latent immune factors. *(A) Overview of the study cohort, follow-up, and multi-omics workflow. ACPA+ at-risk individuals (with arthralgia or PR) were followed longitudinally and profiled across seven molecular modalities. These datasets were integrated using multi-omics factor analysis (MOFA) to derive latent immune factors and assess their association with progression risk. (B) Data availability across modalities, showing high completeness for inflammasome activity, cytokines, transcriptomics, IFN scores, germline variants, HLA genotypes and somatic variants. (C) Heatmap of modality contributions to each latent factor, illustrating coordinated multi-layered variance structure across immune pathways. (D) Proportion of total variance captured by each molecular modality in the integrated model, highlighting dominant contributions from transcriptomics, IFN scores, inflammasome activity and cytokines.*

Association of latent factors to progression to RA

Each latent factor was examined against time to progression using a multivariable Cox regression adjusting for age, sex, smoking and PR (Figure 4A). Factor 6 was found significantly associated with a reduced hazard of progression (hazard ratio [HR] per standard deviation [SD] 0.56, 95% CI 0.40–0.79; $p < 0.001$) and no evidence of proportional hazards violation was observed. Factor 6 values were subset at the median into 'high' and 'low' groups for Kaplan–Meier analysis, demonstrating significant separation in time to progression (log-rank $p = 0.014$, Figure 4B). Median progression times were approximately 1.79 years (21.5 months) in the low Factor 6 group, and 6.7 years (80.3 months) in the high Factor 6 group. Thus, higher Factor 6 values were associated both with a lower overall risk of progressing and, among those who did progress, with a substantially longer time to progression. The association between Factor 6 and progression was not modified by palindromic status (interaction $p = 0.72$).

Variance decomposition indicated that Factor 6 was driven predominantly by inflammasome activity and gene expression modalities, with secondary contributions from somatic CH variants, HLA genotype, and rare germline variants (Figure 4C). Factor 6 had positive loadings with both ASC-NLRP3 and non-specific ASC speck events, as well as CH carrier status, indicating that both events are core components of this latent axis (Figure 4D-E). Conversely, gene-expression loadings showed that the negative direction of Factor 6 was enriched for oxidative phosphorylation, fatty-acid metabolism, mTORC1 signalling, IL-6–JAK–STAT3 signalling, and related transcriptional programmes that broadly support metabolically sustained immune activation, including pathways commonly engaged during adaptive immune responses (Figure 4F). Negative loadings were also observed for SE carriage, driven particularly by HLA-DRB1*04:01 alleles, and for rare germline variant burden (Figure 4G–I). Together, these results suggest that the genetic background associated with increased RA susceptibility exerts an opposing influence on the biology represented by Factor 6.

Collectively, these results indicate that Factor 6 captures an innate-dominant, inflammasome-associated immune state linked to clinical progression (Figure 5). Low Factor 6 scores, characterised by gene expression programmes that favour adaptive immune activity together with inherited genetic risk features, were associated with earlier time to progression, highlighting the interplay between innate and adaptive immune pathways.

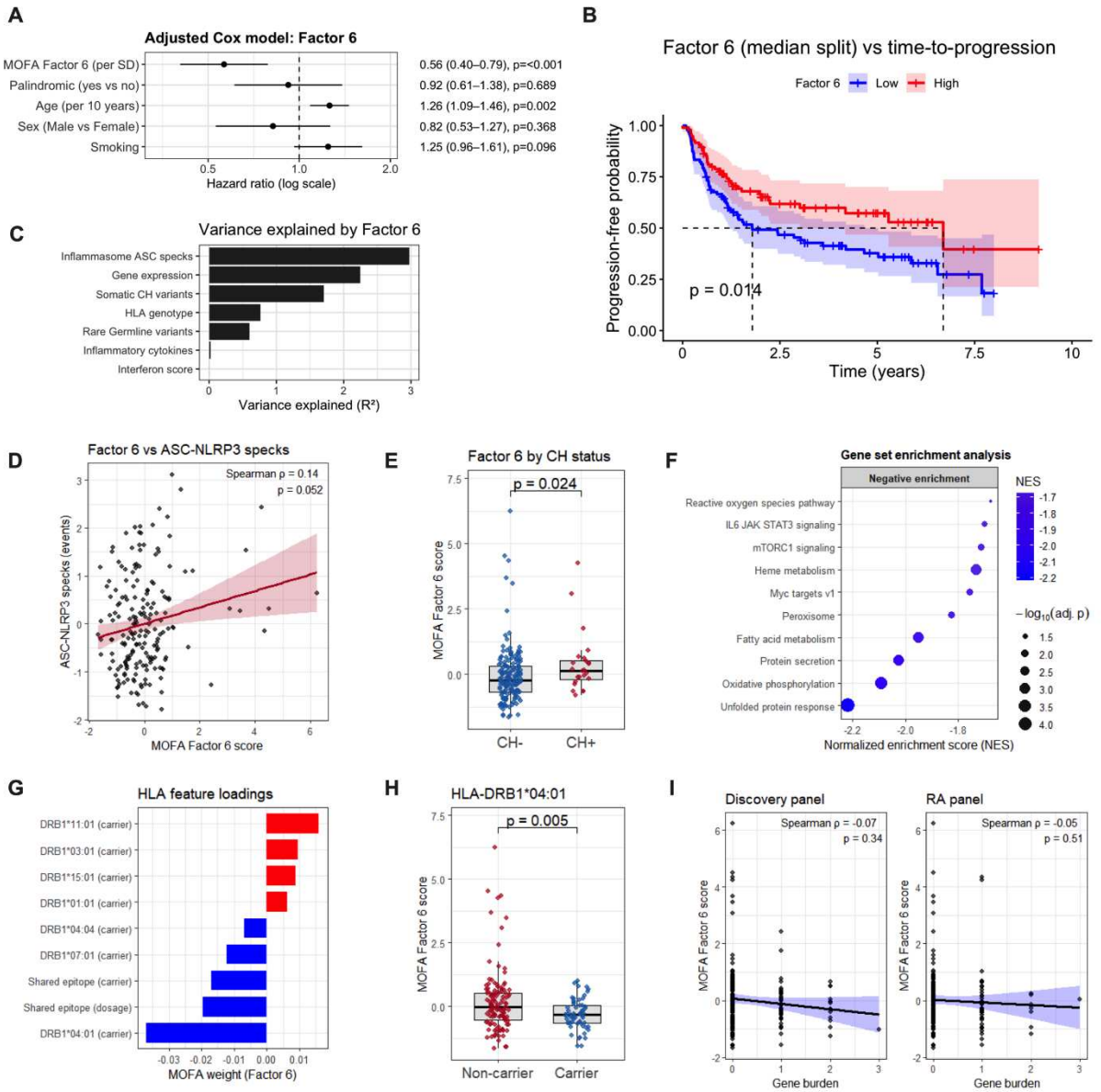


Figure 4. Factor 6 identifies an innate-dominant immune programme linked to progression risk and time-to-progression in ACPA+ individuals. *(A) Cox model showing Factor 6 is associated with reduced progression risk after adjustment for clinical covariates. (B) Kaplan–Meier plot (median split) demonstrating delayed progression in individuals with high Factor 6 scores and earlier progression in those with low Factor 6 scores. Tick marks indicate censored observations. (C) Variance decomposition highlighting major contributions from inflammasome activity, gene expression, somatic CH variants and modest contributions from HLA genotype and rare germline variants. (D–I) Unimodal correlates of Factor 6: positive associations (highlighted in red) with ASC–NLRP3 specks and CH status, and negative associations (highlighted in blue) with adaptive-immune pathways, shared epitope alleles and rare germline variant burden.*

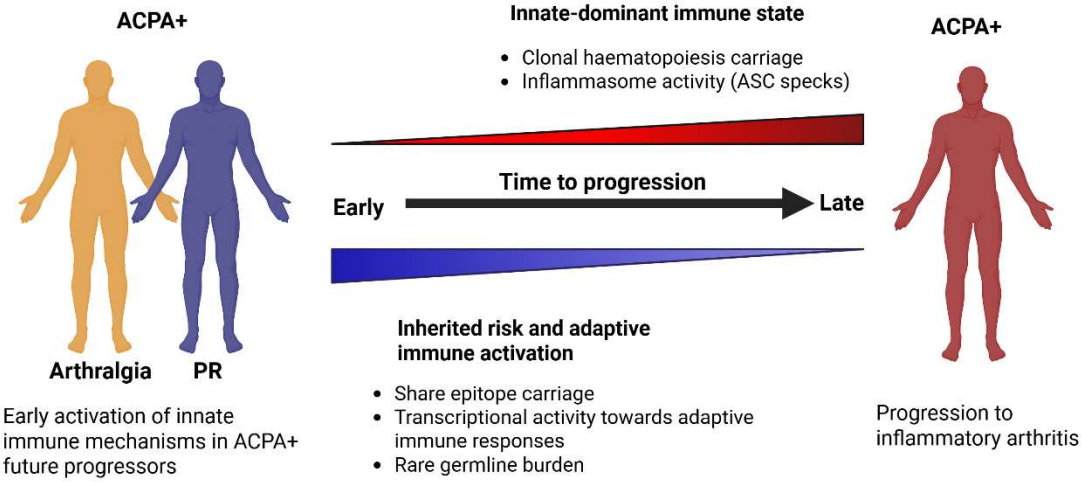


Figure 5. Schematic representation of the Factor 6 immune programme and its relationship to progression in ACPA+ individuals. *Conceptual model illustrating how the Factor 6 immune programme is associated with delayed progression to inflammatory arthritis in ACPA+ individuals. Unimodal analysis showed early innate-immune signatures in at-risk and PR participants at baseline. Prolonged innate-dominant immune states characterised by CH carriage and inflammasome activity are associated with high Factor 6 scores and delayed clinical progression. Lower Factor 6 scores align with participants harbouring inherited genetic risk features (shared epitope carriage and rare germline variant burden) and adaptive-immune transcriptional activity, which are associated with earlier transition to clinical arthritis. Together, the schematic highlights the opposing contributions of innate-dominant versus adaptive-immune mechanisms to progression timing in preclinical RA.*

Latent factors associated with palindromic rheumatism

Three distinct latent factors were independently associated with PR status (Figure 6A). In multivariable logistic regression models adjusted for age, sex, and smoking status, Factor 8 (odds ratio [OR] 1.89, 95% CI 1.38–2.65; FDR-adjusted p = 0.001) and Factor 7 (OR 1.60, 95% CI 1.14–2.33; FDR-adjusted p = 0.033) were positively associated with PR, whereas Factor 10 was negatively associated (OR 0.57, 95% CI 0.41–0.78; FDR-adjusted p = 0.004). There was no correlation between Factor 7 and Factor 8, indicating that they are two orthogonal immune programmes independently associated with PR, supporting biological heterogeneity between within the phenotype (Figure 6B).

Variance decomposition demonstrated that the PR-associated latent factors captured partially overlapping but biologically distinct multi-omics programmes. These factors were primarily driven by different combinations of IFN activity, inflammatory cytokines, and gene expression features, with substantial heterogeneity in their underlying structure. Factor 7 was dominated by IFN activity (Figure 6C-D) and displayed strong positive loadings for IFN-associated cytokines (Figure 6E) and a corresponding positive enrichment of IFN-related gene expression signatures (Figure 6F-G). This indicates that Factor 7 reflects a classical type I IFN–driven immune programme, which remained elevated in individuals with PR features.

In contrast, Factor 8 was driven primarily by inflammatory cytokines with attenuated IFN-related gene expression (Figure 6H-I). Despite Factor 7 and Factor 8 sharing positive loadings with a subset of inflammatory cytokine (Figure 6E and 6J), the underlying biology

of Factor 8 appears to be less associated with IFN and instead had a stronger genetic component, with positive contributions from HLA genotype, particularly the DRB1*03:01 allele (Figure 6K). This HLA-specific enrichment was corroborated by stratified analyses showing higher Factor 8 scores in PR individuals (Figure 6L). These results position Factor 8 as an HLA-modulated inflammatory cytokine programme, distinct from the IFN-dominant axis captured by Factor 7.

Factor 10, which was inversely associated with PR status, captured a coordinated state driven primarily by gene expression features, followed by inflammatory cytokines, IFN activity and rare germline variant burden (Supplementary Figure S1). Gene expression features were positively enriched for pathways related to cellular stress responses, tissue remodelling, apoptosis, and broad inflammatory signalling, while showing negatively enrichment in IFN-related pathways. This pattern indicates that Factor 10 reflects a non-PR inflammatory state characterised by stress-adaptive and tissue-associated transcriptional programmes rather than IFN-driven immunity.

Together, PR-associated latent factors, although all inflammatory, represent divergent immune programmes: an IFN-centric axis (Factor 7), an HLA-influenced cytokine-dominated axis (Factor 8), and a stress-adaptive, tissue-remodelling inflammatory axis (Factor 10). This multi-factor architecture reflects the immunological heterogeneity within anti-CCP+ at-risk individuals and highlights distinct molecular routes toward PR versus non-PR immune states.

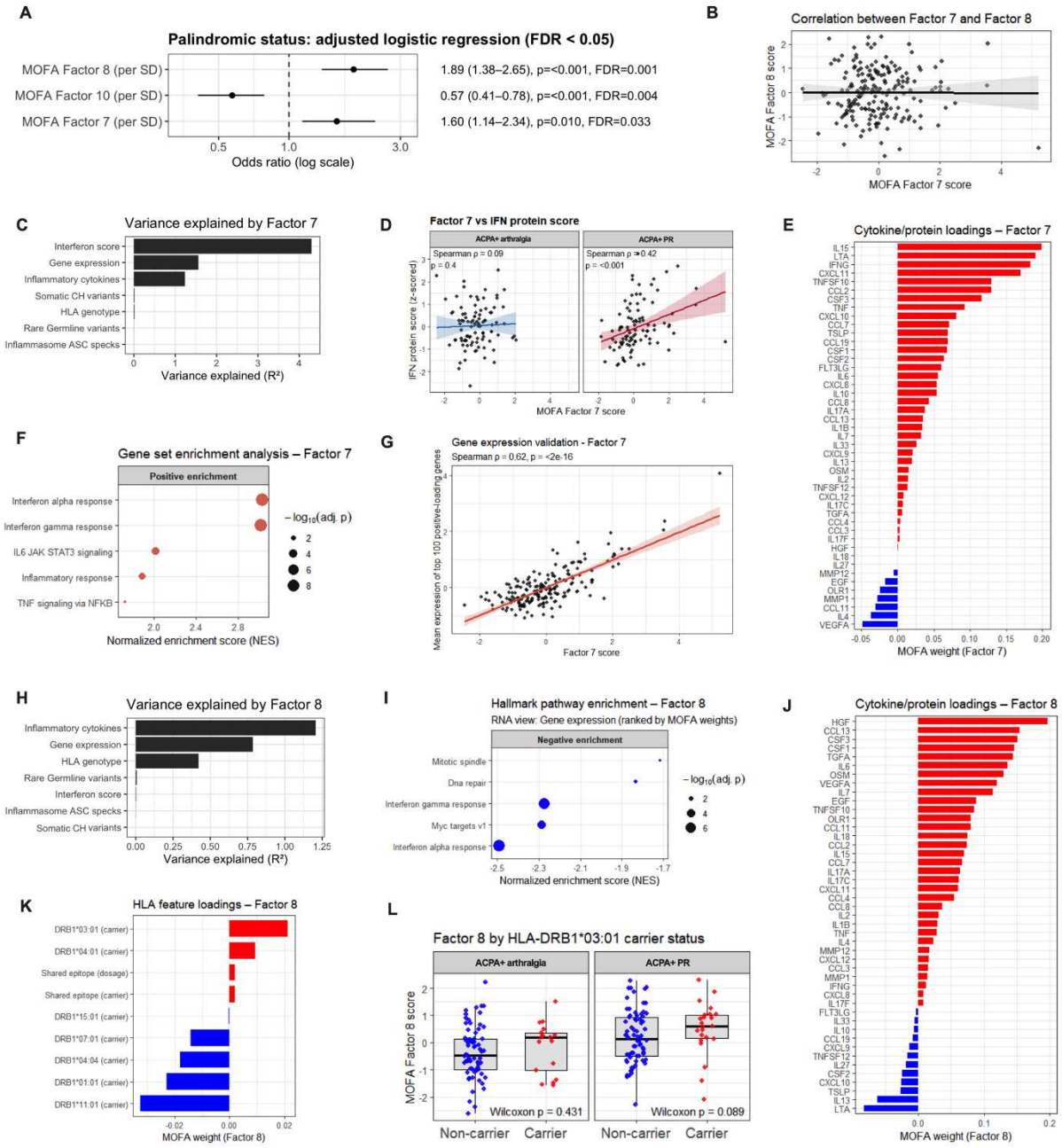


Figure 6. Latent factors associated with palindromic rheumatism (PR) and their molecular correlates. *(A) Logistic regression showing latent Factors 7, 8 and 10 significantly associated with PR status after adjustment for age, sex and smoking. (B) No correlation was observed between Factors 7 and 8, indicating distinct inflammatory programmes. (C–D) Variance contributions and correlation of Factor 7 with IFN protein scores, consistent with an IFN-enriched state in individuals with PR. (E–F) Cytokine loadings and transcriptional pathway enrichment for Factor 7 highlighting IFN-responsive and cytokine-mediated processes. (G) Positive association between Factor 7 and IFN-related gene expression. (H–I) Variance contributions and transcriptional pathway enrichment for Factor 8 showing broader cytokine-driven activation and negative associations with IFN pathways. (J) Cytokine/protein loadings for Factor 8. (K) HLA feature loadings showing positive contributions from HLA-DRB1*03:01 to Factor 8. (L) Factor 8 scores stratified by HLA-DRB1*03:01 carrier status in ACPA+ individuals with arthralgia and PR.*

Discussion

In this multi-omics study of ACPA+ at-risk individuals with arthralgia and PR, we demonstrate that molecular immune perturbations characteristic of established RA, are detectable years before the onset of clinically evident inflammatory arthritis. The integration of genomic, transcriptomic, proteomic and functional immunological layers, uncovered distinct latent immune programmes that define susceptibility, progression kinetics, and divergent inflammatory endotypes within the pre-RA continuum.

Across independent unimodal analyses, we observed consistent evidence of heightened innate immune activation preceding clinical disease. Peripheral blood transcriptomes revealed broad upregulation of IFN-responsive, neutrophil-associated and cytokine-mediated pathways in all study groups relative to HCs, indicating an early and coordinated activation of innate immunity prior to disease onset. This is consistent with previous serological studies showing elevation of cytokines, including IL-6, TNF-related factors, and IFN-induced chemokines years prior to diagnosis[7]. Our findings extend this by demonstrating that these molecular perturbations are encoded across multiple layers of immune regulation, from gene expression to serum protein activity.

Contributions from inherited genetic variation and somatic CH prevalence were also apparent in preclinical immune states. The SE alleles, particularly HLA-DRB1*04:01, were enriched in progressors, consistent with their known role in RA susceptibility and citrulline-specific autoreactivity[32]. While most non-HLA genetic associations in RA have been identified through genome-wide association studies and involve common variants, we examined the potential impact of low-frequency germline coding variants[33,34]. We did not observe enrichment within the RA-associated gene panel, suggesting that previously reported non-HLA susceptibility loci in RA may be largely driven by common germline variation rather than rare coding burden.

In contrast, rare germline variant burden in our immune-focused discovery gene panel differentiated at-risk progressors from non-progressors, suggesting that cumulative inherited burden within inflammation and IFN-related pathways may modulate the threshold for clinical transition. Comparable patterns have been reported in systemic lupus erythematosus (SLE), where aggregates of rare or low-frequency variants in complement, nucleic-acid–sensing, and interferon-regulatory genes contribute to heightened basal immune activation and autoantibody-driven pathology[35,36]. Conditions such as inflammatory bowel disease have also shown that rare damaging variants in innate-immune pathways increase susceptibility and shape inflammatory tone[37]. Together, these disease parallels support a broader model in which cumulative low-frequency germline variation functions as a modifier of underlying immune state across chronic inflammatory and autoantibody-mediated diseases, including preclinical RA.

Although somatic variant burden did not differ significantly with progression status, in either gene panel, at-risk progressors displayed notably higher burden in CH driver genes when compared with age-matched populations[21,38]. The prevalence of CH alone did not predict progression; however, its presence contributed substantially to a multi-omics latent factor associated with disease progression. This supports emerging evidence that

age-related haematopoietic mosaicism may amplify inflammatory states relevant to RA pathogenesis[20].

The primary mechanistic insight from this study is the latent factor associated with delayed progression to RA. Our multi-omics model captured an innate-dominant immune state characterised by inflammasome activity and CH carriage, suggesting that systemic inflammatory priming underpins disease transition. Individuals with low scores in this latent factor progressed to inflammatory arthritis substantially earlier and demonstrated increased transcriptional activity favouring adaptive immune pathways alongside higher rates of shared epitope carriage. This contrast suggests that individuals with at-risk immune states dominated by innate processes remain in a preclinical phase for longer, whereas those with stronger adaptive immune activation experience a more rapid transition. These results highlight the dynamic interplay between autoinflammatory and autoimmune mechanisms, whereby CH-inflammasome activity contributes to the underlying inflammatory state, and adaptive immune activation modulates the rate of disease onset.

Our study further delineates the unique molecular landscape of PR, featuring several latent factors independently associated with PR status. The latent factors enriched in PR reflected distinct inflammatory states: one dominated by IFN-responsive cytokines and gene expression, and another exhibiting a broader cytokine-driven inflammatory signature with additional contributions from HLA-DRB1*03:01. Although PR samples were collected during asymptomatic intervals rather than during flares, the presence of IFN-enriched programmes is notable, as type I IFN activity frequently accompanies episodic inflammatory flares in autoimmune diseases such as SLE, even when measured outside periods of clinical activity[39]. Peri-articular, non-erosive symptoms, intermittent flares, and responsiveness to hydroxychloroquine are also features shared between PR and SLE which suggests that subsets of PR may experience flare-associated immune states like those seen in other immune-complex–mediated conditions. Given that PR commonly occurs in individuals with very high autoantibody titres, immune-complex formation may represent an additional mechanistic link between IFN-related activity and PR pathophysiology. The contribution observed from HLA-DRB1*03:01, a major genetic risk factor implicated in SLE, further supports immunological overlap between PR and SLE pathology and warrants further investigation[40].

This study has demonstrated that integrated multi-omics profiling in populations at risk of RA provides a deeper resolution of disease-relevant immune states than any single modality analysed in isolation and offers a mechanistic foundation for biomarker development. Our identification of a latent factor predictive of progression kinetics also suggests that integrated immune programmes govern transition to clinical arthritis rather than specific pathways. Our characterisation of PR-associated immune factors also supports the clinical notion that PR should be considered a distinct clinical entity - not merely an early stage of RA but encompasses heterogenous endotypes that only partially overlap with ACPA+ individuals with arthralgia[41].

Lastly, several limitations must be acknowledged within the study. Our cohort is among the largest molecularly profiled ACPA+ at risk cohorts to date; however, replication in

independent longitudinal cohorts will be essential to validate the robustness of our findings. Functional validation of factor-defined immune programmes, in relation to progression kinetics, would also further clarify causal mechanisms. Furthermore, our study also focused on peripheral blood, and integration with data from synovial or mucosal tissue would produce a more comprehensive overview of systemic–tissue immune interactions in the early stages of disease progression.

In summary, this study reveals that distinct, multi-layered immune programmes shape the earliest phases of RA development and differentiate ACPA+ individuals with arthralgia destined for progression to clinical arthritis from those with stable autoimmunity or palindromic disease. These findings underscore the importance of integrative immunology in resolving heterogeneity within the preclinical RA spectrum and lay the groundwork for targeted prevention strategies tailored to specific immune trajectories.

Methods

Patient sample cohorts

This study includes 192 at-risk individuals recruited following a positive test for anti-CCP+ antibodies. This study was approved under the following ethics IDs: 06/Q1205/169 (CCP); 17/YH/0177 (CCP-NG); 09/H1307/98 (RADAR). A summary of the cohort can be found in Supplementary Table S1.

Genomic DNA and total RNA extraction

Genomic DNA was extracted from EDTA blood using the Wizard® Genomic DNA Purification Kit (Promega). Peripheral blood collected in PAXGENE RNA tubes were sent to Azenta Life Sciences GENEWIZ for sample processing and total RNA was extracted using QIAGEN RNeasy Kits.

Whole Exome Sequencing (WES)

DNA samples were quantified using Qubit® and libraries (2×150 bp) were prepared using the "Twist Comprehensive Exome 2.0 with spike-in" panel (37.64 Mb) and unique molecular indices (UMIs). Library quality was assessed by Agilent TapeStation and sequenced on an Illumina® NovaSeq™ at the Leeds Genomics Facility, targeting 100× mean coverage. Reads were aligned to GRCh38/hg38 using bwa-mem (v0.7.17); BAM files were coordinate-sorted and duplicates marked using samtools/picard (v3.1.0)[42,43].

Germline Variant Discovery

SNVs and indels were called using GATK HaplotypeCaller (v4.3.0.0), followed by joint genotyping and variant quality score recalibration (VQSR)[44]. Variants were annotated with Ensembl VEP v114 using allele frequencies from gnomAD v4.1. Germline variants were filtered using the following criteria: population allele frequency <1% in gnomAD exomes, sequencing depth (DP) ≥30, genotype quality (GQ) ≥20 and VQSR filter status of PASS. Missense variants were further filtered with CADD scores ≥20, excluding variants classified as "likely benign" by AlphaMissense, and those annotated as by tolerant/benign in both SIFT and PolyPhen-2. Predicted loss of function (pLoF) variants were defined by consequence (stop gained, frameshift, splice acceptor, splice donor, transcript ablation, or exon loss) and further restricted to genes with high loss-of-function intolerance (pLI score ≥0.9). To mitigate potential technical artifacts, rare germline variants with less than 40% VAF and those observed in more than 20 individuals within the cohort with were excluded.

Error-Corrected Target Panel Sequencing

Targeted deep error-corrected sequencing was performed using a custom Twist capture panel covering genes implicated in CH, autoinflammatory disease, and interferonopathies (Supplementary Table 4). UMI-based libraries for the at-risk cohort were sequenced on the Illumina platform to a median raw depth of 7,506×. Duplex consensus reads were generated to suppress sequencing artifacts, yielding mean post-consensus depths of 481× in the at-risk study cohort and 237× in healthy controls.

Somatic Variant Discovery

Somatic variants were identified using GATK Mutect2 in tumour-only mode (GRCh38)[45]. Putative germline variants were excluded using a population allele frequency resource and an in-house panel of normals created using healthy controls (n=45). Detection power for low-frequency somatic variants was estimated using a binomial sampling model assuming a ≥2 alternate duplex consensus read criterion (k=2). Using the lower group's mean depth (237×) as a conservative reference, this corresponded to ~95% power to detect variants present at 2% variant allele frequency (VAF) at typical coverage.

Bulk RNA-Sequencing (RNA-seq)

Total RNA samples were processed for strand-specific RNA-seq with rRNA/globin depletion (~20M paired-end reads/sample). Libraries (2×150 bp) were sequenced on an Illumina® NovaSeq™. FASTQs were quality controlled (FastQC v0.12.1), trimmed (Cutadapt v5.0), and quantified by transcript-level pseudo-alignment (Salmon v.1.10.3)[46-48]. Mapped gene counts were exported to RStudio (v4.4.2) where differential expression analysis and pathway enrichment was achieved using DESeq2 (v1.46.0), and enrichR (v3.4) with the Reactome 2024 gene-set libraries[49,50]. Genes with adjusted p <0.05 and log2FC ≥0.5 were considered differentially expressed. Variance stabilising transformation was performed on the read counts and z-scored for multi-omic analysis. Selected RNA-seq findings were validated in a sample subset (n=44) using SYBR Green qPCR (QuantStudio 7; PowerTrack Master Mix) with *TBP* and *HPRT1* as reference genes (Supplementary Figure S2).

Human Leukocyte Antigen (HLA) Genotyping

HLA genotypes were inferred at two-field resolution using HLA-HD (v1.7.1) from whole exome sequencing data[51]. Reads mapping to the HLA region were extracted and realigned to known HLA allele sequences from the IMGT/HLA v3.58 reference database. Genotypes were validated using paired RNA-seq data analysed with arcasHLA (v0.6.0), achieving 98.2% concordance for HLA-DRB1 alleles (377/384 alleles)[52]. In cases of discordance, HLA-HD calls were retained. HLA-DRB1 alleles observed more than 10 times were tested for enrichment between progressors and non-progressors using Fisher's exact test with Benjamini–Hochberg correction.

Olink Proteomics Analysis

Serum proteins were quantified using the Olink Target 48 Cytokine and Immune Surveillance panel, covering 45 inflammation- and immunity-related targets. Group comparisons of cytokine levels were performed using Wilcoxon rank-sum tests on log2-transformed, z-scored protein values at the individual level. Comparisons included controls versus each clinical subgroup, and progression status comparisons within at-risk and PR individuals. P-values were corrected for multiple testing using the Benjamini–Hochberg procedure.

Inflammasome Activity (ASC Specks) Assay

ASC specks and NLRP3 inflammasome components were measured by flow cytometry using an in-house assay[53]. Serum samples were incubated with PE-conjugated anti-ASC and APC-conjugated anti-NLRP3 antibodies, then analysed on a CytoFLEX S cytometer.

Size calibration beads and ASC-stimulated/deficient THP-1 controls were used to define gating. Results are reported as non-specific ASC or ASC-NLRP3-positive events per µL.

Multi-omic factor analysis (MOFA)

MOFA is an unsupervised variational Bayesian framework that decomposes multiple data matrices into a shared set of latent factors representing sources of joint and modality-specific variation[54]. Seven molecular modalities were included: 1) Peripheral blood transcriptomics, 2) IFN protein scores, 3) Circulating inflammatory cytokines, 4) Inflammasome activity quantified by ASC speck formation, 5) CH variant prevalence, 6) HLA genotype features, and 7) Rare germline variant burden. All modalities were harmonised at the sample level, and features were scaled within each modality prior to model fitting. Count-based RNA-seq data were variance-stabilised and z-scored before inclusion. Binary or categorical genetic features (e.g. CH carrier status, HLA alleles) were encoded numerically to allow joint modelling.

MOFA models were trained using the MOFA2 (v1.16.0) framework and a Bernoulli likelihood for binary views and the Gaussian likelihood for all other views. The number of latent factors was set to $k = 10$ and model stability was assessed using the evidence lower bound (ELBO) data after remodelling at multiple seeds. No outcome or clinical variables were included during training to preserve the unsupervised nature of the analysis.

Variance decomposition and factor characterisation

For each latent factor, the proportion of variance explained was quantified globally and per modality. Factors were interpreted by examining feature loadings within each view, with positive and negative loadings analysed separately to capture opposing biological programmes. High value loadings were used to define factor-specific molecular signatures. Individual factor scores were extracted and used in downstream association analyses with clinical and immunological phenotypes, including time-to-progression outcomes.

Statistics

Baseline characteristics are summarized as counts/percentages for categorical variables and as means (s.d.) or medians (IQR) for continuous variables, as appropriate. The sample sizes for each group and investigation are defined in the figure legends. All data was generated from singular distinct aliquots of genomic DNA, total RNA or serum collected for all participants. Group differences in categorical variables were assessed using χ^2 tests (e.g. SE dosage by progression status) or Fisher's exact tests when expected counts were <5 (e.g. rare variant burden in gene panels). Variants reoccurring in >2 individuals were investigated for enrichment using Fisher's exact tests and population variant frequencies within gnomAD v4.1 (European Non-Finnish ancestry) for comparison. All p-values were adjusted for multiple comparisons using Bonferroni's correction ($p<0.05$/variants tested). In qPCR experiments, relative expression was calculated using the ΔΔCt method and compared to RNA-seq log2FC values using linear regression[55].

For serum cytokines (45-plex), pairwise group comparisons used the Wilcoxon rank–sum test with p-values adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure with FDR < 0.05 considered statistically significant. Chemokine-derived IFN scores were compared by one-way ANOVA with post-hoc pairwise comparisons and multiplicity control; stratified analyses by PR status repeated the same framework. Inflammasome activity (ASC-NLRP3 and non-specific ASC specks) was evaluated across groups using the Kruskal–Wallis test with post-hoc Dunn/Wilcoxon tests and FDR correction. Associations between germline variant burden and immune phenotypes was carried out using Spearman's correlation coefficients adjusted using Benjamini–Hochberg FDR.

Associations between latent factors and time to progression were tested using multivariable Cox proportional hazards models adjusted for age (years), sex (male or female), smoking status (current, previous, never), and PR. Proportional hazards assumptions were evaluated using Schoenfeld residuals. Effect sizes are reported as appropriate for each analysis, including hazard ratios (HR) and odds ratios (OR) per s.d. increase for regression models, Spearman's ρ for correlation analyses, and median or mean differences for group comparisons. Survival curves were visualized with Kaplan–Meier estimates using median dichotomization of factor scores and compared by log-rank tests. Associations between latent factors and PR status used multivariable logistic regression (covariate-adjusted) with OR per s.d. increase and FDR control across factors. All tests were two-sided with $\alpha = 0.05$ (after multiplicity adjustment where applicable).

Acknowledgements
Funding Information: SS is funded by Kennedy trust senior fellowship. LC, SLR are also supported by Kennedy Trust via Senior fellowship awarded to SS. The study is in part supported by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC) (NIHR203331). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Data Availability
Raw and processed bulk RNA-seq data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE319945. All source data underlying the figures and analyses in this study are available from the corresponding author upon reasonable request. Due to the identifiable nature of germline exome sequencing data, these datasets cannot be made publicly available but can be shared in controlled-access form upon request and subject to appropriate data-sharing agreements.

Code Availability
Source code for all analyses written in R has been uploaded to Github and is available at https://github.com/LIRMM-Savic-Group/Preclinical_RA_multiomic_integration_study.

Authorship Contributions: LC, FT, AA, DC, KF, TY, AI, AMR, SLR, AA, EC, conducted experiments. LC, FT, TY, SLR and TY analysed the data and produced the figures; LD, ADM and KM conducted clinical evaluation of the patients and curated the patient cohort. SS, KM and PE designed the study and provided senior oversight; AI, TY, SC and SS developed the figures and tables; LC and SS wrote the first draft; all authors edited and approved the final manuscript

Disclosure of Conflicts of Interest: Authors have no COI to declare in relation to this manuscript.

References

1 McGonagle, D., Watad, A. & Savic, S. Mechanistic immunological based classification of rheumatoid arthritis. *Autoimmunity Reviews* **17**, 1115-1123 (2018). https://doi.org/10.1016/j.autrev.2018.06.001

2 Smolen, J. S., Aletaha, D. & McInnes, I. B. Rheumatoid arthritis. *Lancet* **388**, 2023-2038 (2016). https://doi.org/10.1016/S0140-6736(16)30173-8

3 Guo, Q. *et al*. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. *Bone Res* **6**, 15 (2018). https://doi.org/10.1038/s41413-018-0016-9

4 Gregersen, P. K., Silver, J. & Winchester, R. J. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* **30**, 1205-1213 (1987). https://doi.org/10.1002/art.1780301102

5 Nielen, M. M. *et al*. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis Rheum* **50**, 380-386 (2004). https://doi.org/10.1002/art.20018

6 Sokolove, J., Zhao, X., Chandra, P. E. & Robinson, W. H. Immune complexes containing citrullinated fibrinogen costimulate macrophages via Toll-like receptor 4 and Fcgamma receptor. *Arthritis Rheum* **63**, 53-62 (2011). https://doi.org/10.1002/art.30081

7 Deane, K. D. *et al*. The number of elevated cytokines and chemokines in preclinical seropositive rheumatoid arthritis predicts time to diagnosis in an age-dependent manner. *Arthritis Rheum* **62**, 3161-3172 (2010). https://doi.org/10.1002/art.27638

8 Chalan, P. *et al*. Analysis of serum immune markers in seropositive and seronegative rheumatoid arthritis and in high-risk seropositive arthralgia patients. *Sci Rep* **6**, 26021 (2016). https://doi.org/10.1038/srep26021

9 Rakieh, C. *et al*. Predicting the development of clinical arthritis in anti-CCP positive individuals with non-specific musculoskeletal symptoms: a prospective observational cohort study. *Ann Rheum Dis* **74**, 1659-1666 (2015). https://doi.org/10.1136/annrheumdis-2014-205227

10 Di Matteo, A. *et al*. Ultrasound erosions in the feet best predict progression to inflammatory arthritis in anti-CCP positive at-risk individuals without clinical synovitis. *Ann Rheum Dis* **79**, 901-907 (2020). https://doi.org/10.1136/annrheumdis-2020-217215

11 Mankia, K. & Emery, P. Palindromic rheumatism as part of the rheumatoid arthritis continuum. *Nat Rev Rheumatol* **15**, 687-695 (2019). https://doi.org/10.1038/s41584-019-0308-5

12 Canete, J. D. *et al*. An unexpectedly high frequency of MEFV mutations in patients with anti-citrullinated protein antibody-negative palindromic rheumatism. *Arthritis Rheum* **56**, 2784-2788 (2007). https://doi.org/10.1002/art.22755

13 Canete, J. D. *et al*. Association of intermittent hydrarthrosis with MEFV gene mutations. *Arthritis Rheum* **54**, 2334-2335 (2006). https://doi.org/10.1002/art.21923

14 Asgharvand-Hajeb, A. *et al*. Molecular analysis of TNFAIP3, PTPN22, IRF5, and IL-10 gene polymorphisms in palindromic rheumatism. *Clin Rheumatol* **44**, 4759-4766 (2025). https://doi.org/10.1007/s10067-025-07652-0

15 Savic, S. *et al*. Autoimmune-autoinflammatory rheumatoid arthritis overlaps: a rare but potentially important subgroup of diseases. *RMD Open* **3**, e000550 (2017). https://doi.org/10.1136/rmdopen-2017-000550

16 Beck, D. B. *et al*. Somatic Mutations in UBA1 and Severe Adult-Onset Autoinflammatory Disease. *N Engl J Med* **383**, 2628-2638 (2020). https://doi.org/10.1056/NEJMoa2026834

17 Kosmider, O. *et al*. VEXAS syndrome is characterized by inflammasome activation and monocyte dysregulation. *Nat Commun* **15**, 910 (2024). https://doi.org/10.1038/s41467-024-44811-4

18 Jin, C. *et al*. NLRP3 inflammasome plays a critical role in the pathogenesis of hydroxyapatite-associated arthropathy. *Proc Natl Acad Sci U S A* **108**, 14867-14872 (2011). https://doi.org/10.1073/pnas.1111101108
19 Savola, P. *et al*. Clonal hematopoiesis in patients with rheumatoid arthritis. *Blood Cancer J* **8**, 69 (2018). https://doi.org/10.1038/s41408-018-0107-2
20 Hiitola, E. *et al*. Clonal hematopoiesis is associated with distinct rheumatoid arthritis phenotypes. *Sci Adv* **11**, eadt9846 (2025). https://doi.org/10.1126/sciadv.adt9846
21 Jaiswal, S. *et al*. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371**, 2488-2498 (2014). https://doi.org/10.1056/NEJMoa1408617
22 Fuster, J. J. *et al*. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**, 842-847 (2017). https://doi.org/10.1126/science.aag1381
23 Sano, S. *et al*. CRISPR-Mediated Gene Editing to Assess the Roles of Tet2 and Dnmt3a in Clonal Hematopoiesis and Cardiovascular Disease. *Circ Res* **123**, 335-341 (2018). https://doi.org/10.1161/CIRCRESAHA.118.313225
24 Steensma, D. P. *et al*. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9-16 (2015). https://doi.org/10.1182/blood-2015-03-631747
25 Ishii, S. *et al*. ADAM-17 is expressed on rheumatoid arthritis fibroblast-like synoviocytes and regulates proinflammatory mediator expression and monocyte adhesion. *Arthritis Res Ther* **20**, 159 (2018). https://doi.org/10.1186/s13075-018-1657-1
26 Liu, X. *et al*. ALPK1 Accelerates the Pathogenesis of Osteoarthritis by Activating NLRP3 Signaling. *J Bone Miner Res* **37**, 1973-1985 (2022). https://doi.org/10.1002/jbmr.4669
27 Sun, W. *et al*. An Integrated Analysis of HAVCR1 with a Focus on Immunological and Prognostic Roles in Breast Cancer. *Cancer Inform* **24**, 11769351251393148 (2025). https://doi.org/10.1177/11769351251393148
28 Zeng, R. *et al*. Investigating Causality and Shared Genetic Architecture between Neurodegenerative Disorders and Inflammatory Bowel Disease. *Aging Dis* **14**, 1349-1359 (2023). https://doi.org/10.14336/AD.2022.12209
29 Xu, S. *et al*. SRRM2 organizes splicing condensates to regulate alternative splicing. *Nucleic Acids Res* **50**, 8599-8614 (2022). https://doi.org/10.1093/nar/gkac669
30 Kar, S. P. *et al*. Genome-wide analyses of 200,453 individuals yield new insights into the causes and consequences of clonal hematopoiesis. *Nat Genet* **54**, 1155-1166 (2022). https://doi.org/10.1038/s41588-022-01121-z
31 Stengel, A. *et al*. Mutational patterns and their correlation to CHIP-related mutations and age in hematological malignancies. *Blood Adv* **5**, 4426-4434 (2021). https://doi.org/10.1182/bloodadvances.2021004668
32 Holoshitz, J. The rheumatoid arthritis HLA-DRB1 shared epitope. *Curr Opin Rheumatol* **22**, 293-298 (2010). https://doi.org/10.1097/BOR.0b013e328336ba63
33 Kronzer, V. L., Sparks, J. A., Raychaudhuri, S. & Cerhan, J. R. Low-frequency and rare genetic variants associated with rheumatoid arthritis risk. *Nat Rev Rheumatol* **20**, 290-300 (2024). https://doi.org/10.1038/s41584-024-01096-7
34 Ishigaki, K. *et al*. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat Genet* **54**, 1640-1651 (2022). https://doi.org/10.1038/s41588-022-01213-w
35 Belot, A. *et al*. Contribution of rare and predicted pathogenic gene variants to childhood-onset lupus: a large, genetic panel analysis of British and French cohorts. *Lancet Rheumatol* **2**, e99-e109 (2020). https://doi.org/10.1016/S2665-9913(19)30142-0
36 Jiang, S. H. *et al*. Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus. *Nat Commun* **10**, 2201 (2019). https://doi.org/10.1038/s41467-019-10242-9

37 Rivas, M. A. *et al*. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* **43**, 1066-1073 (2011). https://doi.org/10.1038/ng.952
38 Buscarlet, M. *et al*. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood* **130**, 753-762 (2017). https://doi.org/10.1182/blood-2017-04-777029
39 Banchereau, J. & Pascual, V. Type I interferon in systemic lupus erythematosus and other autoimmune diseases. *Immunity* **25**, 383-392 (2006). https://doi.org/10.1016/j.immuni.2006.08.010
40 Miglioranza Scavuzzi, B. *et al*. The lupus susceptibility allele DRB1*03:01 encodes a disease-driving epitope. *Commun Biol* **5**, 751 (2022). https://doi.org/10.1038/s42003-022-03717-x
41 Sanmarti, R. *et al*. Palindromic Rheumatism: Just a Pre-rheumatoid Stage or Something Else? *Front Med (Lausanne)* **8**, 657983 (2021). https://doi.org/10.3389/fmed.2021.657983
42 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009). https://doi.org/10.1093/bioinformatics/btp324
43 Danecek, P. *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021). https://doi.org/10.1093/gigascience/giab008
44 Poplin, R. *et al*. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178 (2018). https://doi.org/10.1101/201178
45 Benjamin, D. *et al*. Calling Somatic SNVs and Indels with Mutect2. *bioRxiv*, 861054 (2019). https://doi.org/10.1101/861054
46 Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. (2010).
47 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **17** (2011).
48 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017). https://doi.org/10.1038/nmeth.4197
49 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8
50 Kuleshov, M. V. *et al*. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016). https://doi.org/10.1093/nar/gkw377
51 Kawaguchi, S., Higasa, K., Shimizu, M., Yamada, R. & Matsuda, F. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat* **38**, 788-797 (2017). https://doi.org/10.1002/humu.23230
52 Orenbuch, R. *et al*. arcasHLA: high-resolution HLA typing from RNAseq. *Bioinformatics* **36**, 33-40 (2020). https://doi.org/10.1093/bioinformatics/btz474
53 Topping, J. *et al*. FACS-based detection of extracellular ASC specks from NLRP3 inflammasomes in inflammatory diseases. *Clin Exp Immunol* **219** (2025). https://doi.org/10.1093/cei/uxae117
54 Argelaguet, R. *et al*. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* **14**, e8124 (2018). https://doi.org/10.15252/msb.20178124
55 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408 (2001). https://doi.org/10.1006/meth.2001.1262

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SuppFigS1.png
- SuppFigS2.png
- SuppTableS1BaselineDemographicsSummary.docx
- SuppTableS2Panels.docx
- SuppTableS3germlinevariantsarthralgiaprogressors.docx
- SuppTableS4germlineexomewide.docx
- SuppTableS5somaticCH.docx