Single-cell RNA-seq analysis reveals cell subsets and gene signatures associated with Rheumatoid Arthritis Disease Activity

Marie Binvignat, … , Mary C. Nakamura, Marina Sirota

*JCI Insight.* 2024. [https://doi.org/10.1172/jci.insight.178499](https://doi.org/10.1172/jci.insight.178499).

**Graphical abstract**

Find the latest version:

[https://jci.me/178499/pdf](https://jci.me/178499/pdf)
Single-cell RNA-seq analysis reveals cell subsets and gene signatures associated with Rheumatoid Arthritis Disease Activity

Marie Binvignat†,1,2,3, Brenda Y. Miao†,1, Camilla Wibrand†,1,4, Monica M. Yang5, Dmitry Rychkov1, Emily Flynn5,6, Johanne Nititham5, Whitney Tamaki1, Umair Khan1, Alex Carvidi5, Michele Krueger7, Erenne Niemi5, Yang Sun8, Gabriela Fragiadakis5,6, Jérémie Sellam3, Encarnita Mariotti-Ferrandiz2, David Klatzmann2, Andrew Gross5, Chun Jimmie Ye8, Atul J. Butte1,9, Lindsey A. Criswell5,10, Mary C. Nakamura5,11, Marina Sirota*1,9

Affiliations:
1. Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco; CA, USA.
2. Immunology Immunopathology Immunotherapy, Pitie Salpetriere Hospital UMRS 959, Sorbonne Université; Paris, France.
3. Department of Rheumatology, Research Center Saint Antoine, UMRS 938, Sorbonne Université, AP-HP, Saint-Antoine Hospital, Inserm UMRS 938; Paris, France.
4. Aarhus University; Aarhus, Denmark.
5. Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, Division of Rheumatology, Department of Medicine, University of California San Francisco; San Francisco, CA, United States.
6. CoLabs, University of California; San Francisco, CA, United States.
7. Department of Medicine, Oregon Health & Science University; Portland, OR, United States.
8. Human Genetics, University of California; San Francisco, CA, USA.
9. Department of Pediatrics, University of California San Francisco; San Francisco, CA, United States.
10. National Human Genome Research Institute (NHGRI), NIH, Bethesda, MD, United States.
11. San Francisco VA Health Care System; San Francisco, CA, USA.

† equal contribution
* corresponding author

Corresponding Author:
Marina Sirota, PhD
Professor, Acting Director
Bakar Computational Health Sciences Institute, (BCHSI)
University of California, San Francisco
marina.sirota@ucsf.edu
(415) 279-8737
Abstract: scRNA-seq profiling of disease activity in RA

Rheumatoid arthritis (RA) management lean toward achieving remission or low-disease activity. In this study, we conducted single-cell RNA sequencing (scRNAseq) of peripheral blood mononuclear cells (PBMCs) from 36 individuals (18 RA patients and 18 matched controls, accounting for age, sex, race, and ethnicity), to identify disease-relevant cell subsets and cell type-specific signatures associated with disease activity. Our analysis revealed 18 distinct PBMC subsets, including an IFITM3 overexpressing Interferon-activated (IFN-activated) monocyte subset. We observed an increase in CD4+ T effector memory cells in patients with moderate to high disease activity (DAS28-CRP ≥ 3.2), and a decrease in non-classical monocytes in patients with low disease activity or remission (DAS28-CRP < 3.2). Pseudobulk analysis by cell type identified 168 differentially expressed genes between RA and matched controls, with a downregulation of pro-inflammatory genes in the gamma-delta T cells subset, alteration of genes associated with RA predisposition in the IFN-activated subset, and non-classical monocytes. Additionally, we identified a gene signature associated with moderate-high disease activity, characterized by upregulation of pro-inflammatory genes such as TNF, JUN, EGR1, IFIT2, MAFB, G0S2, and downregulation of genes including HLA-DQB1, HLA-DRB5, TNFSF13B. Notably, cell-cell communication analysis revealed an upregulation of signaling pathways, including VISTA, in both moderate-high and remission-low disease activity contexts. Our findings provide valuable insights into the systemic cellular and molecular mechanisms underlying RA disease activity.
INTRODUCTION

Rheumatoid arthritis (RA) is a systemic autoimmune disease characterized by chronic inflammation and joint destruction (1). While the prevalence and disease burden vary considerably between geographic regions and populations (2), RA impacts approximately 1.3 million adults in the United States, representing 0.6 to 1% of the country's population (3, 4). RA is a debilitating condition and a major socio-economic burden with a prevalence of work disability around 35% (5). Effective RA management necessitates early diagnosis, a treat-to-target approach, and the attainment of remission or low disease activity (6). Achieving optimal therapeutic success remains the main challenge in RA as only 16% of patients reach sustained remission or low disease activity (7, 8). This has been particularly underscored by the recent recommendations from the European Alliance of Associations for Rheumatology (EULAR), particularly concerning the management of difficult-to-treat RA patients (9).

The understanding of cellular and molecular mechanisms underlying disease activity has garnered substantial attention. Notably, specific cell subsets, such as synovial tissue macrophages, have been associated with both remission and disease activity (10). Additionally, several synovial molecular and pathobiological markers have shown promise in predicting treatment response (11–13). The emergence of bulk transcriptomic data has further revealed that alterations in synovial and blood transcriptomic profiles were closely associated with disease activity and flares (14–16). In the pursuit of comprehensive insights, single-cell RNA sequencing (scRNA-seq) emerges as a powerful tool to simultaneously profile cell subset compositions and cell type-specific transcriptional states, enabling a deeper understanding of mechanisms associated with non-remission. Several studies have utilized single-cell resolution to investigate RA, although the majority of research has focused on synovial tissue and none specifically studied disease activity (17–19).

Recent cross-tissue meta-analyses of transcriptome data, encompassing samples from both human and murine models, have uncovered genes associated with disease activity, and underscored a divergence between synovium and peripheral blood profiles (20, 21). Consequently, it is crucial to recognize that markers identified in the synovium cannot be directly extrapolated to those found in peripheral blood. The accessibility of peripheral blood, compared to the more invasive nature of synovial sampling, emphasizes its practical advantage for both research and potential clinical applications. An additional challenge in studying RA disease activity is the inherent heterogeneity of the condition and the potential influence of demographic factors, such as sex, age, ethnicity and race, on disease activity (22–25). A critical aspect of addressing this challenge involves promoting the establishment of more standardized and diverse cohorts, enabling a better exploration of specific cell subsets and biomarkers that contribute to disease activity.

In this study, we describe a comprehensive analysis of disease activity using scRNA-seq of peripheral blood mononuclear cells (PBMC) in a diverse cohort of RA patients, matched with controls based on age, sex, race, and ethnicity. Our primary objective was to identify specific cell subsets and biomarkers associated with disease activity. Additionally, we aimed to assess the specific RA cell subsets and gene signatures in a diverse population, providing valuable insights into the multifaceted nature of RA pathogenesis.
RESULTS

Experimental study design
We performed a single cell analysis on PBMC samples from 36 participants (18 RA and 18 controls matched on age, sex, ethnicity, and race). Study design and population characteristics are described in Fig. 1a, and table 1. The mean age was 53.75 ± 15.9 (mean ± standard deviation (SD)), the study was composed of 66.7% (N=24) of women, 11.1 % of Asian Americans, 83.3% of Caucasian, 5.6% of Latinx population. There was no significant difference in population characteristics between RA patients and matched controls (table S1). Among the RA patients, 62.5% (N=10) presented a positive Rheumatoid Factor (RF), 87.5% (N=14) presented positive anti-citrullinated antibodies (ACPA). RA disease activity was evaluated using the Disease Activity Score on 28 joints using C-reactive protein (CRP) (DAS28-CRP) (26). Clinical data regarding disease activity was available for 16 out of the 18 RA patients included in our study. The mean DAS28-CRP was 3.3 ± 1.0. Patients were stratified in remission - low disease activity (DAS28-CRP<3.2) (N=9) and moderate-high disease activity (DAS28-CRP ≥ 3.2) (N=7). Erosive disease was present in 62.5% of patients (N=10). The mean time since diagnosis was 4.13 ± 4.41 years. Seven patients were treated by conventional Disease-modifying antirheumatic drugs (DMARDs) (38.9%), only one patient was treated by a biological DMARDs (TNF-inhibitor: etanercept) and 6 (33.3 %) patients were also treated with oral corticosteroids (prednisone ≥ 5 mg)

Identification of 18 PBMC cell subtypes.
PBMCs collected on peripheral blood were pooled, profiled and barcoded in three batches with 12 lanes using the 10x Genomics Chromium Single Cell technology. RA and matched controls were evenly split within each batch and lanes. Cell Ranger v3 was used for demultiplexing and read mapping to the human genome. The mean number of cells for all samples described was 5990.7 (SD:1842.8) before filtering and 3307.8 (SD: 1031.7) after. An average of 1259.2 cells were filtered out per sample as doublets and an average of 1423.7 cells per sample were filtered out due to low quality (Fig. S14, Table S8-9). Following 10X sequencing and preprocessing with Scanpy, our dataset consisted of 125,698 cells and 22,159 genes (Fig. S1, Table S6-7). Leiden community detection was used to group cells into clusters, and annotation using established cell markers showed the presence of all major PBMC cell types (Fig. S2). All major cell types, annotated using established cell markers, were present in our dataset. Further clustering and annotation were used to identify cell subsets, including five CD4+ T cell subsets (CD4+ T central memory, CD4+ T effector memory, CD4+ IFIT+ T cells, CD4+ Naive T cells, gamma-delta T cells), three CD8+ T cell subsets (CD8+ Early T effector memory, CD8+ Naive T cells, Terminally differentiated effector memory (TEMRA)), two Natural Killer (NK) cell subsets (CD56bright NK cells, CD56 low NK cells), three B cell subsets (Naive B cells, Memory B cells and Plasmabasts), and five monocyte subsets (Classical Monocytes, IFITM3+ IFN-activated, IL1B, Myeloid dendritic cells, non-classical monocytes) (Fig. 1b, Table S10). For the control sample that was replicated across batches, we found no statistically significant differences in cell proportions. Each of the cell subsets presented a distinct expression profile (Fig. 1c, Fig. S2). CD4+, CD8+ T cells and NK cells subtypes showed higher similarity profiles, and the expression profiles of gamma-delta T cells exhibited a stronger correlation within the CD8+ T cells (Fig. 1d). Cell subsets and top genes of each cell-subset identified through Wilcoxon-rank sum analysis are included in Fig. 2a-b. We identified two cell subtypes (CD4+ IFIT+ cells and IFITM3+ IFN-activated monocytes) associated with genes related to Interferon (IFN)-pathway activation in RA. The pro-inflammatory CD4+
IFIT+ cell subtype presented significant expression of several genes associated with inflammatory response and immune regulation, including \textit{IFIT2}, \textit{PMAIP1}, \textit{NFKBIZ}, \textit{TNFAIP3} and \textit{ZC3HAV1}. IFITM3+ IFN-activated monocytes had elevated levels of \textit{IFITM3}, \textit{ISG15}, \textit{FTL}, \textit{TYMP}, and \textit{FTH1}. In addition, IFITM3 expression was specific to this monocyte subset (Fig. S4). Non-classical monocytes cell proportions were significantly lower in RA patients compared to controls (Wilcoxon-signed rank analysis p=0.024) (Fig. S3, Fig. S5). There were no significant differences in the proportion of other cell types.

Pseudobulk differential expression analysis revealed a specific down-regulation of pro-inflammatory genes related to gamma-delta T cells in RA patients in comparison to healthy controls

By performing pseudobulk differential expression analysis on 18 cell subtypes, we identified a total of 168 genes that exhibited differential expression between individuals with RA and the control group (FDR ≤ 0.05, |log2FC| ≥ log2(1.6)). The majority of those genes were expressed in monocytes (n = 94) and CD8+ T cells (n = 39), 26 genes were expressed in CD4+ T cells, 6 in B cells, and 3 in NK cells (Fig. 3a, 3b, Fig. 3c, table S2, table S3, Fig. S6). 121 genes were unique, and 47 genes were expressed across multiple cell types. Patients with RA had higher expression of genes associated with inflammation and cardiovascular risk in IL-1β and classical monocytes, including \textit{IFITM2}, \textit{TXNIP}, \textit{EAF1}, \textit{RIT1}, \textit{EGR1}, \textit{TLE3}, and \textit{SLA}. In addition, they showed over-expression of cytotoxic genes \textit{KLRD1}, \textit{GZMH}, and \textit{EBP} in CD8+ T cells. RA patients also displayed significant downregulation of proinflammatory genes such as \textit{IFNG}, \textit{IFIT2}, \textit{TNF}, \textit{GZMA}, \textit{ISG15}, and \textit{S100A4} exclusively in the gamma-delta T cells. Non-classical monocytes showed a specific transcriptomic profile of 19 differential expressed genes not shared with other cell-subsets (Fig. 3b,c), including a downregulation of \textit{ETNK1}, \textit{TNFSF13B}, \textit{DUSP7}, \textit{IGSF6} and an upregulation of \textit{CXCR4} in RA. Interestingly, the IFN-activated subset also presented a cell-type specific downregulation of \textit{HLA-DQB1}, \textit{LRRK2}, \textit{MS4A7}, and \textit{G0S2} in RA.

Over representation analysis found significant upregulation of B cell activation in RA patients

Functional analysis derived from pseudobulk differential expression analysis (FDR ≤ 0.05), identified 25 significantly up-regulated pathways across cell subsets in patients with RA compared to the control group. Among these pathways, 11 were upregulated in B cells, 10 in Monocytes, and 4 in CD4+ T cells. Additionally, 21 pathways were significantly downregulated, with 16 in Monocytes, 13 in CD4+ T cells, one in CD8+ T cells, and one in B cells (FDR ≤ 0.01, Gene ratio ≥ 15, counts ≥ 5) (Fig. 4, Table S4). We observed a noteworthy abundance of up-regulated pathways specifically in B cells of patients with RA as compared to the control group. These pathways primarily encompassed immune response, B cell activation, B cell receptor pathways, antigen-receptor mediated signaling pathways and immune response regulating cell surface receptor signaling pathways. Furthermore, within the gamma-delta T cell population, we observed a significant downregulation of pathways involved in positive regulation of myeloid and leukocyte differentiation as well as cytokine production regulation, in RA patients. Finally, in non-classical monocytes, we observed an up-regulation of cytokine-mediated signaling pathway and a
downregulation of pathways involved in T cell activation, lymphocytes regulation and mononuclear cell proliferation, and leukocyte cell-cell adhesion in RA patients.

**CD4+ central memory cell and non-classical monocytes proportions were associated with disease activity in RA patients**

We performed a stratified analysis based on disease activity comparing individuals with active vs. inactive disease. Information regarding DAS4-28-CRP was available in 16 out of 18 RA patients included in the study. RA patients were divided into two groups: one consisting of individuals in remission or with low disease activity, characterized by a DAS28-CRP <3.2 (N=9), and the other group comprising patients with moderate and high disease activity, indicated by DAS28-CRP ≥3.2 (N=7). There was no significant statistical difference between the two groups in terms of age, sex, race, ethnicity, BMI, proportions of ACPA and RF, erosive disease, treatment strategy, and disease duration (Table 1). We compared differences in cell density and cell subset proportions between those two groups and the control group (Fig. 5a-c). Additionally, we conducted a non-parametric Spearman correlation analysis to evaluate the association between cell proportions and DAS28-CRP as a continuous score (Fig. S7). Although there was no statistical significance in B cell proportion across groups, we observed a clear density shift between patients with remission low disease activity vs moderate and high activity from Naive to activated Memory B cells (Fig. 5b). RA patients with moderate to high disease activity showed a significantly increased proportion of CD4+ central memory cells (p=0.034) (Fig. 5c). Conversely, Non-classical monocytes were significantly lower in patients in the remission-low disease activity group compared to both control and the group with moderate-high disease activity (p=0.022).

**Identification of a gene signature specific to moderate and high disease activity in RA**

Using a gene list consisting of 121 unique genes that exhibited differential expression between individuals with RA and matched controls in our pseudobulk analysis, we conducted a subanalysis focusing on patients with different disease activity levels: those in remission or with low disease activity, and those with moderate to high disease activity. Among the 121 genes, 75 were significantly associated with low disease activity, and 89 associated with high disease activity (FDR ≤ 0.05, \(|\log2(FC)| \geq \log2(1.6), 0.08 \leq \text{base, mean} < 4\)). Interestingly, 52 genes were significantly upregulated in patients within the moderate to high disease activity group. These genes included G0S2, THBS1, DUSP7, IFIT2, IGSF6, MAFB, RIT1, TNF, JUN, CXCR4, and TLE3 (Fig. 6, Fig. S8). Furthermore, we observed a separate set of 37 genes that exhibited a significant downregulation in patients with moderate to high disease activity compared to the control group. These downregulated genes included TRBC1, KLRB1, IL32, HLA-DQB1, HLA-DRB5, TNFSF13B, CCL3, LRRK2, and TMA7. Additionally, we identified a set of 12 genes that were significantly overexpressed only in patients with remission or low disease activity, which included TXNIP, LGALS2, and AREG.

**CD4+ T cells and B cell subsets were associated with the highest levels of cell-cell communication in RA patients**

To gain a comprehensive understanding of immune cell communication, we conducted a cell-cell communication inference analysis using CellChat, which uses a repeated permutation to identify significant cell-cell communications. We found a statistically significant increase in cell-cell communication in RA patients as compared to healthy controls in 35 pairs of cell types. One cell-cell pair showed no difference, and 255 pairs had less communication in RA (Fig 7a) The largest
increase in communication was found in CD4+ naive T and CD4+ T central memory cells along with Naive and Memory B cells both as senders and receivers. In addition, there was an increase between Classical, Non-Classical, and IL-1β Monocytes as senders and Naive CD4+ T cells as receivers. When stratifying patients based on disease activity, we found similar patterns. In patients with low disease activity, 28 pairs had statistically significant increases and 286 pairs had statistically significantly decreased in communication as compared to controls (Fig S9). In patients with moderate and high disease activity, 37 pairs had statistically significant increases in communication and 259 pairs presented with statistically significant decreases in communication, while one pair had a similar amount of communication as compared to controls. For the high activity states, a decrease as both sender and receiver was observed in particular for CD8+ T naïve and early TEM and CD4+ T effector memory as well as CD56 bright NK cells and IFITM3 IFN-activated monocytes, whereas the low disease activity state had a decrease of both sending and receiving communication of IFIT positive CD4+ T cells, myeloid dendritic cells as well as IFITM3 IFN-activated monocytes. On the other hand, myeloid dendritic cells appeared to be more involved in communication in high disease activity as both a sender and receiver, while NK cells and CD4+ naive T cells were more involved in sending and receiving communication in low disease activity.

Cell-cell communication revealed an upregulation of VISTA and Interferon II pathways in RA patients with moderate and high disease activity.

CellChat utilizes a curated database of ligand-receptor pairs, grouped into communication pathways which may contain multiple ligand-receptor pairs. The change in communication pathways between disease states is then calculated as the relative contribution per disease state to the total communication amount for a specific communication pathway. We used a threshold for significant contribution at less than 35% or more than 65% of the total communication and a p-value < 0.05. We found several distinct communication pathways to be up- or downregulated in RA, with 7 pathways showing more communication and 14 showing less communication as compared to controls (Fig 7b, Fig S10). Upregulated pathways included NT, HGF, VISTA, IFN-II and WNT. When comparing moderate-high disease activity to low disease activity, we found the highest number of changed pathways to be in moderate-high (36 vs 22) (Fig 7c, Fig S11a-c). Many pathways from the initial comparison remained significantly and solely expressed in RA, including NT (neurotrophic), HGF, VISTA and IFN-II, although PECAM-1 also became significant. For each pathway, specific ligand-receptor pairs were considered main contributors (Fig S12).

For most pathways, the direction of dysregulation was similar in both high and low disease activity (n = 12). However, the THBS (thrombospondin), CD99 and GH pathways were upregulated in high disease activity but downregulated in low disease activity. On the other hand, there was more communication in TGFb, activin, GRN and IL-2 pathways in low disease activity as compared to high disease activity. Although the direction of dysregulation was alike for high and low disease activity, the cells contributing to communication pathways differed. For IFN-II, naïve CD4+ T cells and CD56 bright NK cells sent signals to primarily IL1B and nonclassical monocytes in high disease activity (Fig 7d). In low disease activity, most signaling came from classical and IL-1b monocytes and was received by IFITM3 IFN-activated monocytes (Fig 7d). In other pathways, such as VEGF and NT, the involved cells remained unchanged across disease states; however, small changes in importance were observed (Fig 7c-d, Fig S13). For IL2, most communication
was between CD4+ cells, although myeloid dendritic cells also played a large role in high disease activity (Fig S13). Another upregulated pathway across disease states was the VISTA pathway, for which we found discrepancies in the involved cells, as non-classical and IFITM3 IFN-activated monocytes were the primary senders in high disease activity as opposed to low disease activity where non-classical monocytes were out shadowed by classical and IL1B-monocytes (Fig 7e). For both disease states, however, the main receivers were within the CD4+ and myeloid subsets, although CD4+ Naive T cells were more prominent receivers in low disease activity, and CD4+ Memory T cells were more prominent in high disease activity.

**DISCUSSION**

Here we describe a dataset of scRNA-Seq of PBMCs from a diverse population of 36 patients with 18 RA and 18 matched controls on age, sex, race, and ethnicity. Of these subsets, we found that non-classical monocyte proportions were significantly lower in RA patients compared to controls. This is in line with other studies of RA biomarkers, which hypothesize that inflammatory cell subsets may migrate from blood to synovial tissue (20,21). Our findings showing increased CD4+ central memory cell proportions and decreased non-classical monocytes specifically in patients with moderate to high disease activity compared to controls provide further evidence aligned with these hypotheses. However, unlike previous studies (78), we did not find statistically significant enrichment of plasmablasts or other cell subsets in RA patient populations, which may be due to low sample sizes or differing patient demographics from previous studies. One other subset of interest identified in our study was an IFITM3+ expressing IFN-activated monocyte subset. IFITM3, also known as IFN induced trans-membrane 3, is associated with type 1 IFN response and viral restriction (27). Associations between IFITM3 haplotypes and rheumatoid arthritis, particularly in the context of a Korean population (28). Additionally, Zhang et al. (17) identified upregulation of IFITM3 in synovial IL1B+ monocytes from a subset of patients in RA, prompting interest in whether differences in IFITM3+ cell proportions could be contributing to these differences in expression. Other interferon induced genes, particularly OAS1, JSG15, IFI44L, and IFI6 (77), have also been shown to be upregulated in whole blood samples from patients with RA, although these studies did not find any association with disease activity. Studies on IFN-activated monocytes have also suggested their role in antigen presentation through upregulation of CD86 and HLA-DR and driving the differentiation of Th17 cells in the synovium of patients with RA (47, 77).

While we did not find differences in IFITM3+ monocyte proportions or upregulation of the IFITM3+ gene specifically, our other findings described below did show significant differences in VISTA signaling in this subset, and upregulation of IFN-induced genes in pseudobulk analysis from patients exhibiting moderate and high disease activity. The differences between our study and previous findings from Zhang et al. may be explained by varying study design, including cohort selection and control populations. While we compare PBMC samples between patients with RA stratified by disease activity and healthy controls, upregulation of IFITM3 shown by Zhang et al. was only shown in synovial samples from leukocyte rich RA compared to control samples obtained from patients with osteoarthritis. Further studies of IFITM3 and other IFN-related genes are needed to better elucidate the global role of interferon and IFN-activated monocytes in RA pathophysiology.
In our pseudobulk differential expression analysis, we observed a specific down regulation of pro-inflammatory genes in the gamma-delta T cells subsets including, \textit{IFNG}, \textit{IFIT2}, \textit{TNF}, \textit{GZMA}, \textit{ISG15}, \textit{S100A4}. Interestingly Mo et al. described that peripheral V\textdelta2T cells were significantly lower in patients with RA and were negatively correlated with disease activity. In addition, they described that V\textdelta2 T cells from RA accumulated in the synovium and produced high levels of pro-inflammatory cytokines including IFN-\gamma and IL-17 and also showed elevated chemotaxis potential (34). Our results could reinforce the potential chemotaxis role of V\textdelta2 T cells in the synovium. However, it is essential to acknowledge that this remains a hypothesis as we lacked matched tissue data in our study to confirm this hypothesis conclusively. Similarly, we observed a lower proportion of non-classical monocytes in PBMCs from patients with remission-low disease activity groups compared to both controls and moderate-high disease activity. Guła et al have reported that the absolute number of circulating non-classical monocytes negatively correlates with DAS28 and swollen joint count in peripheral spondyloarthritis patients (35). Non-classical monocytes have also been associated as key mediators of tissue destruction in osteoclasts in murine models of rheumatoid arthritis (36). In addition we found a specific downregulation in the non-classical monocytes subsets of several genes such as \textit{IGFS6}, \textit{ETNK1}, \textit{DUSP7}, and \textit{TNFSF13B}. \textit{IGFS6} expression has been significantly associated with RA fibroblast cells in humans (37). \textit{Etnk1} has been associated as a candidate gene in collagen induced arthritis (38). \textit{DUSP7} is involved in MAPK signaling and low levels of mRNA of \textit{DUSP7} have been associated with RF ACPA positive RA patients (39). \textit{TNFSF13B} variants have been associated in several studies with RA and also with other auto-immune diseases such as systemic erythematosus lupus (40, 41).

We also identified a gene signature of 89 genes specific to disease activity including an upregulation of proinflammatory genes such as \textit{TNF}, \textit{JUN}, \textit{EGR1}, \textit{IFIT2}, \textit{IGSF6}, \textit{TMX1}, and \textit{MAFB}, but also potential genes associated as therapeutic targets or in treatment response such as \textit{G0S2}, \textit{PTGS2}, and \textit{THBS1}. \textit{Mafb} has been associated with monocytes and macrophage differentiation but also involved in the activation of myeloid cells associated with joint destruction such as RANK+ TLR2- cells in murine models of RA (42, 43). \textit{G0S2} has been associated with anti-TNF response prediction in a meta-analysis of 11 studies (44). Thrombospondin-1 expression has been associated with NR4A2 activity and is modulated by TNF inhibitors (45). We also observed several genes downregulated in patients with high disease activity such as \textit{HLA-DQB1}, \textit{HLA-DRB5}, and \textit{TNFSF13B}. Similarly, Klimenta et al. also found a protective role of HLA-DRB5 in RA (46). In an independent single cell study, Wu et al. showed that HLA-DRB5+ expression was lower in the synovial tissues of ACPA- RA patients (18). A thorough analysis of these gene signatures enables the identification of various functional groups, notably the upregulation of IFN-induced genes in patients exhibiting moderate and high disease activity. These findings align with existing literature that emphasizes the significant role of IFN in both RA and disease activity (47, 48). Furthermore, they underscore the potential of cutting-edge RA treatments, such as JAK inhibitors, as IFN exerts its effects through the JAK/STAT pathway (49).

Cell-cell communication allowed us to confirm several well-known signaling pathways from the RA literature, including type II IFN (IFN-\gamma), TGF\beta and VEGF (48, 50, 51). Additionally, our findings revealed an upregulation of the IL-2 signaling pathway in patients with remission and low disease activity, whereas a downregulation was observed in patients with moderate and high disease activity compared to controls. IL-2 has been correlated with disease activity and severity
in several studies(52, 53). The IL-2 pathway plays a pivotal role in Regulatory T cell response and holds significance in rheumatic diseases (54). IL-7 is known to play a significant role in the activation and proliferation of many cells, and notably T cells, including regulatory T cells (55). IL-7 is also important for the development and differentiation of Tregs, their homeostatic maintenance and their expansion, although to a lesser extent that IL-2 which is indispensable(56, 57). In addition, the IL7R shares the common gamma chain, and thus part of the downstream activation pathway. Our results underscore the intricate interplay between these cytokines and their downstream signaling cascades and may signify that the shared IL-2/IL-7 pathways and genes are particularly relevant in the context of RA pathogenesis.”. However, it should be acknowledged that our study lacked sufficient power to identify specific memory and regulatory T cell subsets and did not identify IL-2 or IL-7 in our differential expression gene list. Thus, further research is warranted to refine and enhance this hypothesis.

We also found an up-regulation of the VISTA signaling pathway in RA patients. VISTA is a negative checkpoint regulator, playing a key role in suppressing T cell-mediated immune responses, and its disruption has been linked to pro-inflammatory phenotypes and a susceptibility to autoimmune diseases(58). In a recent groundbreaking study, ElTanbouly et al. investigated the role of VISTA expression in T cells and found that VISTA expressed on naive T cells was playing a critical role in quiescence and peripheral tolerance, and hence that blocking or knocking out VISTA exacerbates inflammation in mice models (57). However, results from Ceeraz et al have found a significant reduction of arthritis score in VISTA deficient mice in a collagen antibody-induced arthritis model independent of T- and B-cells.(56). Taken together, those results could suggest that VISTA may have immunoregulatory roles on naïve T cells but inflammatory roles on myeloid cells. In our signaling network analysis, however, CD4+ T cells subsets were the main receivers of VISTA signaling. CD4+ naive T cells were the primary receiver population in the remission-low disease activity group, while CD4+ T central memory was the predominant group in moderate-high disease activity. Interestingly, the monocytes cells population were associated as the main sender in the VISTA signaling network, in particular IL-1β Monocytes and classical monocytes in remission-low disease activity and Non-classical and IFITM3+ monocytes in moderate and high disease activity groups. Targeting VISTA shows promising potential as an innovative immunoregulatory therapy (59). Our findings further support the involvement of VISTA in RA and its potential impact on the communication between monocytes and CD4+ T cells, possibly indicating a more significant role of VISTA on monocytes as compared to T cells in RA, which however needs confirmation in laboratory studies. Additional research is necessary to gain a comprehensive understanding of VISTA's role in RA and its apparent multifaceted role in autoimmunity.

Several limitations should be acknowledged in our study. First, it is challenging to use scRNA-seq of unsorted PBMCs to study very small cell subsets, such as B cells subsets or regulatory T cells, which may have led to their role in disease activity being underestimated and not extensively explored. Additionally, the relatively small sample size of our study may have limited the statistical power to detect subtle differences. This limitation also restricted our ability to thoroughly explore potential differences associated with sex, race and ethnicity, emphasizing the need for more inclusive representation in future investigations to ensure a comprehensive understanding of RA across diverse populations.
Finally we acknowledge that RA treatments could be expected to modify the phenotype of PBMCs, and that the small sample size of this study doesn’t allow us to explore these differences analytically. Future studies should be designed specifically to explore the effects of treatments on cell type specific expression at a single cell level. Prospective single cell study and longitudinal data on flare and treatment response could also provide further insight, could reinforce previous findings from the work form Orange et al, and could deepen our understanding of RA disease activity(15). Another limitation is the absence of matched disease synovial tissue, which could have provided more comprehensive insights into how patterns in PBMCs relate to cellular and molecular mechanisms involved in RA synovium. Also, we have looked strictly at PBMCs, but a significant role of neutrophils in RA has been proposed and future studies should explore this potential interaction (61). (https://pubmed.ncbi.nlm.nih.gov/33469455/)

Nonetheless, our study provides valuable insights into the cellular and molecular mechanisms associated with disease activity in RA. We carefully considered matched controls on age, sex, race, and ethnicity, and our work has identified key cell subsets and genes that may be associated with disease activity. These findings have the potential to serve as new biomarkers and therapeutic targets. We are optimistic that our research will contribute to advancing our understanding of RA pathogenesis and lead to the development of more effective treatments for this complex autoimmune disease.

METHODS

Study Design
RA Patients meeting the American College of Rheumatology (ACR) classification criteria (62) were recruited from the University of California San Francisco (UCSF) rheumatology clinic between 2016 and 2020 (63). Healthy controls were recruited through local advertising and through the database ResearchMatch (64). Controls were matched to RA patients by age, sex, race and ethnicity. The use of the word ‘women’ and ‘men’ in this study is utilized as a proxy for biological sex and does not represent gender identity. Blood samples and participant level data were collected at time of enrollment. Clinical data including demographics, medication status, laboratory values, such as erythrocyte sedimentation rate, C-reactive protein (CRP), RF, ACPA antibodies, and clinical measurements of disease activity with the Disease Activity Score in 28 joints using CRP (DAS28-CRP). Patients were stratified in remission or low disease activity (DAS28-CRP <3.2) and moderate and high disease activity, (DAS28-CRP ≥ 3.2) according to the 2019 updated ACR recommendation on disease activity measures (26).

Sample processing and 10X single cell RNA-sequencing
Peripheral Blood Mononuclear cells (PBMCs) were isolated from peripheral blood samples by UCSF Bay Area Center for AIDS Research Specimen Processing and Banking Subcore (previously AIDS specimen bank). Blood samples were collected in EDTA tubes, processed per manufacturer’s guidelines, and cryopreserved in liquid nitrogen. Single cell sequencing was performed using the 10X Chromium microfluidics system (10X Genomic). PBMCs from 18 RA patients and 18 healthy controls were thawed, counted, and pooled and profiled in three batches and 12 lanes, using 10x Genomics Chromium Single Cell 3’V3. Barcoded cDNA libraries were prepared using the single cell 3’mRNA kit. Cell Ranger v3 (3.1.0) was used to demultiplex cellular
barcodes and map reads to the human (GRCh38-3.0.0) genome(65). Sample deconvolution and doublet identification was performed using demuxlet(66). RA and matched control samples were evenly split within each batch to limit technical and biological bias in our analysis. The first batch consisted of 14 individuals (seven RA and seven controls), the second batch included eight individuals (three RA and five controls), and the third batch included 16 individuals (eight RA and eight controls). One control was sequenced across three batches as a technical replicate to control for batch effect.

**Preprocessing**
Preprocessing was performed using Scanpy (1.9.1)(67) following previously published single cell workflows (68). Additional details on these methods can be found in our supplemental materials and published code. Genes found in less than three cells were filtered out, as well as cells containing fewer than 100 genes or more than 1000 genes. Cells with platelet or megakaryocyte gene markers (PF4, GNG11, PPBP, SDPR) were also removed. Additionally, cells containing greater than 20% mitochondrial genes or less than 3% ribosomal genes were removed. Following filtering, the data were normalized to counts per million and log transformed. Technical variation from sequencing depth, mitochondrial percentage, ribosomal percentage were regressed out during scaling. Cell cycle scoring was performed using Scanpy using standard genes (69) and also regressed out. Batch correction was then performed using HarmonyPy(70), and samples were clustered in an unsupervised manner using leiden clustering with a resolution of 3.0. Each cluster was assigned as CD4+ T cells, CD8+ T cells, Monocytes, NK cells, or B cells using manual annotation with predefined marker genes according to the human protein atlas. Clusters of platelets, erythrocytes and suspected doublets were removed from further analysis based on the presence of marker genes. Sub-clustering was repeated for each cell type to allow for fine annotations of cell subsets, again based on referenced marker genes and leiden clustering. No size cutoff was used to remove small clusters. Highly expressed genes within each subset were identified using Wilcoxon rank testing implemented in Scanpy.

**Compositional analysis**
Cell densities in each subset were calculated and plotted for RA samples versus controls using Scanpy embedding density functions. The proportion of each cell type within a sample relative to the total number of annotated cells for that sample was also calculated. Cells proportions were compared using Wilcoxon signed rank tests between RA and their matched controls. Additionally, Mann-Whitney U tests were used for cell proportions comparison among controls, RA patients with remission-low disease activity, and moderate-high disease activity. Correlation between cell-type proportion and DAS28-CRP was also performed by calculating Spearman rank-order coefficient. Two sided p-value <0.05 was the threshold for statistical significance.

**Differential gene expression analysis**
Differential gene expression analysis was performed between RA and matched controls using a pseudobulk approach using the bulk RNAseq tool DESeq2 (1.38.3)(71). Pseudobulk methods outperforms mixed models and limits pseudoreplication bias (72, 73). For each cell subtype, read counts were summed across each sample to create a pseudobulk count matrix. DESeq2 was applied, using a Likelihood ratio test corrected on batch effect with an additional fit of a Gamma-Poisson Generalized Linear Model (GLM)(74). P-values were adjusted using Benjamini-Hochberg method and genes with a false discovery rate (FDR) ≤ 0.05 were selected. Additional filtering was
applied with an absolute log2Fold change $\geq \log2(1.6)$ and a base mean expression between 0.08 and 4.

**Over-representation analysis**
Functional and over-representation analysis was performed on differentially expressed genes for each cell subtype using clusterprofiler (4.6.2) (75) and Gene Ontology (GO) database (76). We selected up and down regulated pathways related to biological processes, with a Gene Ratio $\geq 0.15$, count $\geq 5$ and (FDR) $\leq 0.01$.

**Cell-cell communication analysis using CellChat**
Cell-cell communication inference and visualization was performed using the CellChat R package (version 1.6.0) (77). CellChat uses the log-normalized expression matrix as input and predicts cell-cell communication based on ligand-receptor pairs in a curated database. For each pair of ligands and receptors, the communication probability is calculated based on the average expression of the ligand in one cell type and the average expression of the receptor in another cell type, taking into account the law of mass action. CellChat also considers other important signaling factors such as heteromeric complexes and cell type proportion in the estimation of the strength of interactions. We followed the standardized CellChat workflow, including the ‘projectData’-function, which allows for projecting the gene expression onto a validated protein-protein interaction network to impute the data. Cell-cell pairwise communication was visualized as the relative number of communication between groups of interest (RA v. control, low disease activity v. controls and high disease activity v. controls). Statistical significance (FDR adjusted $p \leq 0.05$) of cell type sender and receivers was assessed by performing 50 permutations and comparing the results using a student’s t-test. Only pathways that were statistically significant ($p \leq 0.05$) and with a relative contribution in the RankNet-function of either more than 0.65 or less than 0.35 were considered.

**Study approval**
The study was conducted in accordance with the principles outlined in the Declaration of Helsinki and was granted ethical approval by the Human Research Protection Program and the Institutional Review Board of UCSF (IRB Project No. 15-17175). All participants provided written informed consent.

**Data availability**
Data presented in this study are deposited in the CellxGene Discover resource at https://cellxgene.cziscience.com/collections/e1a9ca56-f2ee-435d-980a-4f49ab7a952b upon publication. The code used for this analysis will be made publicly available on Github at https://github.com/BMiao10/RASingleCell. Values for all data points in graphs are reported in the Supporting Data Values file.
Supplementary Materials:
Fig. S1 to S13
Table S1 to S10
Data file S3
Data file S4
Data file S5
Supporting data values
REFERENCES


https://doi.org/10.1016/j.jbspin.2014.01.013.


Acknowledgments: We would like to thank the patients who contributed to the study and the administrative and medical team of UCSF rheumatology clinics. We would like to thank the PREMIER center as well as members of the Sirota Lab for useful discussion. This work is in part supported by Pfizer and NIAMS P30 AR070155.

Funding: This work was supported by the Rheumatology Research Foundation Targeted Research Pilot Grant p30 from the National Institutes of Health (NIH) , National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) (P30-AR070155) and additional funding from Pfizer.

Author’s contribution: AG and LAC recruited the patients for the study, MN, EN performed the single cell experiments, D.R. and W.T. performed the alignment and the samples deconvolution. M.B. and B.Y.M. performed the quality controls, the pre-processing and the cell proportion analyses. M.B., B.Y.M. and C.W. contributed to the cell-type annotations. M.B. performed the pseudobulk analysis and the analysis associated with disease activity. M.B. and C.W. performed the functional analysis. C.W. performed the cell-cell communication analysis. E.F. and U.K. peer-reviewed and provide essential feedback on the code and the analysis strategy. M.Y. and M.N. provided complementary information on the study metadata. M.B., B.Y.M., C.W., and M.Y. wrote the manuscript and M.S. revised the manuscript. M.S, L.A.C and M.N conceptualized the study and M.S was involved in the funding acquisition. M.S was involved in the project administration. Y.S, G.F, J.N, J.S, D.K. E.M.F A.J.B, A.G, J.Y, L.A.C M.N provided critical feedback and edits to the manuscript.

Competing interests: MB's research is funded by a grant from the French Society of Rheumatology, the Osteoarthritis Foundation grant, and a doctoral fellowship from Sorbonne University. BYM is a paid consultant for SandboxAQ. U.K. is a paid consultant for Vevo Therapeutics. DR's research is supported by a grant from Pfizer Aspire, NIH, and the Rheumatology Research Foundation. EF's research is funded by PREMIER and by grants from SICCA and Eli Lilly. GKF's research is funded by grants from NIH, Gerson Baar Foundation, ImmunoX Computational Biology Initiative, Eli Lilly, Bill & Melinda Gates. JS received research grants from Pfizer, MDS, Schwaedico, BMS and received lecture fees from MDS, Pfizer, Abbvie, Fresenius, Kabi, BMS, Roche, Chugai, Sandoz, Lilly, Gilead, Novartis, Janssen. EMF is receiving lecture fees from Sanofi France and expert testimony payment from Takara Bio. AJB's research is funded by NIH and Chan Zuckerberg Initiative grants. AJB is a paid consultant for Maze Therapeutic, HiBio Santa Ana, Trex Bio, ImYoo, DeciBio, and received lecture fees from Boston Children’s Hospital, Johns Hopkins University, Endocrine Society, Alliance for Academic Internal Medicine, Roche, Children’s Hospital of Philadelphia, University of Pittsburgh Medical Center, Cleveland Clinic, University of Utah, Society of Toxicology, Mayo Clinic, Pfizer, Cerner, Johnson and Johson, the Transplantation society; consultant to Samsung, Mango Tree Corporation,
Dartmouth, Gladstone Institute, Boston Children’s Hospital; and has received honoraria and travel reimbursement for invited talks from Alliance for Academic Internal Medicine, Cleveland Clinic, University of Utah, Society of Toxicology, Mayo Clinic, Children’s Hospital of Philadelphia, American Association of Clinical Chemistry, Analytica, Life Science & Diagnostics Association. AJB receives royalty payments through Stanford University, for several patents and other disclosures licensed to NuMedii, Personalis and Progenity. AJB’s research has been funded by NIH, Genentech, Johnson and Johnson, Merck, Peraton (as prime for an NIH contract), Priscilla Chan and Mark Zuckerberg Bakar Family foundation. None of these entities had any role in the design, execution, evaluation, or writing of this manuscript.

**Data and materials availability:** Upon manuscript acceptance, the data and code will be made publicly accessible on GO, dbGap, and GitHub. Supporting data values for figures are also included in supplemental files.
FIGURE LEGENDS

Figure 1 | Identification of 18 PBMCs cell subsets. A. UMAP embeddings and subset annotations of single cell RNAseq dataset from patients with rheumatoid arthritis (n=18) and healthy controls (n=18) matched on age, sex, and ethnicity. B. Normalized expression of the top 40 ranked genes in different cell subsets (Wilcoxon rank test, FDR≤0.05). C. Correlation heatmap of genes expression across cells subsets (Spearman correlation) CD, Cluster differentiation; DCs, Dendritics cells; IFIT, Interferon Induced proteins with Tetratricopeptide repeats; IFITM, interferon-induced transmembrane; Tem: T effector memory, TEMRA: Terminally differentiated effector memory. RA Rheumatoid Arthritis

Figure 2 | Cell subsets and top marker genes identified in Wilcoxon rank sum. A. UMAP embedding for cell subsets in B cells, Monocytes CD4 T cells, CD8 T cells, NK cells. B. Dot plot of top ranking genes in each cells subsets CD, Cluster differentiation; DCs, Dendritics cells; IFIT, Interferon Induced proteins with Tetratricopeptide repeats; IFITM, interferon-induced transmembrane; Tem: T effector memory, TEMRA: Terminally differentiated effector memory. RA Rheumatoid Arthritis

Figure 3 | Pseudobulk analysis between RA patients and matched controls for each subsets. A. Single cell UMAP projection of patients with rheumatoid arthritis and matched controls. B Differentially expressed genes between patients with RA and matched controls. C. Upset plots of upregulated and downregulated genes across different cell subsets. CD, Cluster differentiation; DCs, Dendritics cells; IFIT, Interferon Induced proteins with Tetratricopeptide repeats; IFITM, interferon-induced transmembrane; Tem: T effector memory, TEMRA: Terminally differentiated effector memory. RA Rheumatoid Arthritis

Figure 4 | Functional analysis and between RA and matched controls. A Pathways and over-representation analysis, for each cell subtypes (Gene ratio > 0.15, FDR ≤ 0.05, 0.08 < base mean < 4). CD, Cluster differentiation; DCs, Dendritics cells; IFIT, Interferon Induced proteins with Tetratricopeptide repeats; IFITM, interferon-induced transmembrane; Tem, T effector memory; TEMRA, Terminally differentiated effector memory; RA, Rheumatoid Arthritis

Figure 5 | Cell proportion and cell density between patients with remission-low and moderate-high disease activity and matched controls A Umap representation of cell subsets B Compositional and density analysis between control patients with low disease and high activity. C. Cell proportion analysis between controls and RA patients with low disease and high activity. Each point represents the cell subset proportion of each patient normalized to the total number of cells for that patient. (Mann-Whitney - Wilcoxon p≤ 0.05). CD, Cluster differentiation; DCs, Dendritics cells; IFIT, Interferon Induced proteins with Tetratricopeptide repeats; IFITM, interferon-induced transmembrane; Tem, T effector memory, TEMRA, Terminally differentiated effector memory; RA, Rheumatoid Arthritis

Figure 6 | Gene signature associated with disease activity and percentage of expression across cell subsets. Gene expression heatmap between controls and RA with low and high disease activity and average expression across cell subtypes. CD, Cluster differentiation; DCs, Dendritics cells;
IFIT, Interferon Induced proteins with Tetratricopeptide repeats; IFITM, interferon-induced transmembrane; Tem, T effector memory; TEMRA, Terminally differentiated effector memory; RA, Rheumatoid Arthritis

**Figure 7 | Cell-cell communications between patients with remission-low and moderate-high disease activity and matched controls.**

A. Heatmap representing the relative number of interactions between RA and matched controls. B. Barplot illustrates statistically significant communication pathways based on the weight of interactions between RA patients and controls. C Dot-plot of the relative contribution of communication pathways based on weight of interactions between high and low disease activity compared to controls. D. Heatmaps of the relative importance of cells as senders and receivers for the IFN-II, the VEGF and the NT signaling pathway network in high and low disease activity. E. Circle plots representing the relative importance of cells as senders and receivers for the VISTA signaling pathway network in high and low disease activity and overall. CD, Cluster differentiation; DA: disease activity; DCs, Dendritic cells; IFIT, Interferon Induced proteins with Tetratricopeptide repeats; IFITM, interferon-induced transmembrane; Tem, T effector memory; TEMRA, Terminally differentiated effector memory; RA, Rheumatoid Arthritis
Table 1 | Clinical characteristics of RA patients and patients with remission-low and moderate-high disease activity. The table presents demographic data for a total of 18 RA patients. Disease activity information is available for 16 of them. A subanalysis was conducted, stratifying patients into two groups, those in remission or with low disease activity (N=7) and those with moderate to high disease activity (N=9) using a DAS28-CRP threshold of 3.2. A p-value threshold for significance was set at <0.05. Student's t-test was conducted to analyze continuous variables. For categorical variables, a Chi-square test was performed.

ACPA, anti-citrullinated body mass index; bDMARDs: biological disease modifying antirheumatic drugs; cDMARDS, conventional disease modifying antirheumatic drugs; DAS: Disease activity score; RA, Rheumatoid Arthritis; RF, Rheumatoid Factor

<table>
<thead>
<tr>
<th></th>
<th>RA remission and low disease activity (DAS28-CRP &lt; 3.2) (N = 9)</th>
<th>RA moderate and high disease activity (DAS28-CRP ≥ 3.2) (N = 7)</th>
<th>p-value</th>
<th>Missing values (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>12 (66.7 %)</td>
<td>4 (57.1 %)</td>
<td>ns</td>
<td>0</td>
</tr>
<tr>
<td>Male</td>
<td>6 (33.33)</td>
<td>3 (42.9 %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years) (mean (SD))</td>
<td>51.7 (15.3)</td>
<td>59.2 (14.3)</td>
<td>ns</td>
<td>0</td>
</tr>
<tr>
<td>Race (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2 (11.1 %)</td>
<td>1 (14.3 %)</td>
<td>ns</td>
<td>0</td>
</tr>
<tr>
<td>Caucasian</td>
<td>15 (83.3 %)</td>
<td>6 (85.7 %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1 (5.6 %)</td>
<td>0 (0.0 %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>1 (5.6 %)</td>
<td>0 (0.0 %)</td>
<td>ns</td>
<td>0</td>
</tr>
<tr>
<td>Not Hispanic</td>
<td>17 (94.4 %)</td>
<td>7 (100.0 %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI kg/m² (mean (SD))</td>
<td>26.7 (10.3)</td>
<td>24.1 (5.4 %)</td>
<td></td>
<td>5.6</td>
</tr>
<tr>
<td>DAS4 28 CRP (mean (SD))</td>
<td>3.3 (1.0)</td>
<td>4.2 (0.6)</td>
<td>&lt;0.001</td>
<td>11.1</td>
</tr>
<tr>
<td>RF (%)</td>
<td>10 (62.5)</td>
<td>5 (71.4 %)</td>
<td>ns</td>
<td>11.1</td>
</tr>
<tr>
<td>ACPA (%)</td>
<td>14 (87.50)</td>
<td>6 (85.8 %)</td>
<td>ns</td>
<td>11.1</td>
</tr>
<tr>
<td>Erosive status (%)</td>
<td>10 (62.5)</td>
<td>5 (71.4 %)</td>
<td>ns</td>
<td>11.1</td>
</tr>
<tr>
<td>Oral Corticosteroids (%)</td>
<td>6 (27.8 %)</td>
<td>2 (28.6 %)</td>
<td>ns</td>
<td>5.6</td>
</tr>
</tbody>
</table>
(Prednisone ≥ 5mg) (%)

<table>
<thead>
<tr>
<th>Category</th>
<th>Value 1 (Percentage)</th>
<th>Value 2 (Percentage)</th>
<th>Value 3 (Percentage)</th>
<th>Stat</th>
<th>Value 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDMARDs (%)</td>
<td>7 (38.9 %)</td>
<td>5 (55.6 %)</td>
<td>2 (33.3 %)</td>
<td>ns</td>
<td>5.6</td>
</tr>
<tr>
<td>bDMARD (%)</td>
<td>1 (5.6 %)</td>
<td>0 (0 %)</td>
<td>1 (14.3 %)</td>
<td>ns</td>
<td>11.1</td>
</tr>
<tr>
<td>Disease duration (years)</td>
<td>4.13 (4.41)</td>
<td>3.61 (4.31)</td>
<td>4.80 (5.22)</td>
<td>ns</td>
<td>5.3</td>
</tr>
</tbody>
</table>
B cells
CD4 T cells
CD8 T cells
Monocytes

Regulatory pathways:
- Positive regulation of myeloid cell differentiation
- Positive regulation of small molecule metabolic process
- Positive regulation of leukocyte differentiation
- Positive regulation of hemopoiesis
- Regulation of myeloid cell differentiation
- Alpha-beta T cell activation
- RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
- mRNA splicing, via spliceosome
- RNA splicing, via transesterification reactions
- Negative regulation of leukocyte activation
- Negative regulation of cell activation
- Regulation of lymphocyte proliferation
- Regulation of mononuclear cell proliferation
- Lymphocyte proliferation
- Lymphocyte differentiation
- Regulation of hemopoiesis
- Cytokine–mediated signaling pathway
- Leukocyte proliferation
- T cell differentiation
- Regulation of leukocyte proliferation
- Regulation of cell–cell adhesion
- Regulation of T cell activation
- Leukocyte cell–cell adhesion
- Regulation of leukocyte cell–cell adhesion
- Positive regulation of cytokine production
- Response to virus
- Maintenance of location
- Regulation of translation
- Mononuclear cell differentiation
- Regulation of mRNA metabolic process
- Response to mechanical stimulus
- Activation of immune response
- Antigen receptor–mediated signaling pathway
- Immune response–activating cell surface receptor signaling pathway
- Immune response–activating signal transduction
- Immune response–regulating cell surface receptor signaling pathway
- B cell activation
- Negative regulation of immune system process
- Lymphocyte mediated immunity
- Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
- B cell receptor signaling pathway

Legend
- **FDR < 0.01**
- **Gene ratio**
  - 0.20
  - 0.25
  - 0.30
- **Regulation**
  - Down regulated pathway
  - Up regulated pathway
RA high disease activity signature
RA low activity
Control
Naive B cells
Memory B cells
Plasmablasts
Classical Monocytes
IL1b–Monocytes
Non–classical Monocytes
Myeloid DCs
IFITM3 Monocytes
CD4 T central memory
CD4 T Naive
CD4 T effector memory
CD4 T IFIT
CD8 T Naive
CD8 TEMRA
CD8 T early Tem
NKCD56bright
NKCD56low

Figure Legend