ORIGINAL RESEARCH

Role of synovial fibroblast subsets across synovial pathotypes in rheumatoid arthritis: a deconvolution analysis

Raphael Micheroli [1], Muriel Elhai,[1] Sam Edalat,[1] Mojca Frank-Bertoncelj,[1] Kristina Bürki,[1] Adrian Ciurea [1], Lucy MacDonald,[2] Mariola Kurowska-Stolarska,[2] Myles J Lewis,[3] Katriona Goldmann,[3] Cankut Cubuk,[3] Tadeja Kuret,[4] Oliver Distler,[1] Costantino Pitzalis,[3] Caroline Ospelt [1]

ABSTRACT

Objectives To integrate published single-cell RNA sequencing (scRNA-seq) data and assess the contribution of synovial fibroblast (SF) subsets to synovial pathotypes and respective clinical characteristics in treatment-naïve early arthritis.

Methods In this in silico study, we integrated scRNA-seq data from published studies with additional unpublished in-house data. Standard Seurat, Harmony and Liger workflow was performed for integration and differential gene expression analysis. We estimated single cell type proportions in bulk RNA-seq data (deconvolution) from synovial tissue from 87 treatment-naïve early arthritis patients in the Pathobiology of Early Arthritis Cohort using MuSiC. SF proportions across synovial pathotypes (fibroid, lymphoid and myeloid) and relationship of disease activity measurements across different synovial pathotypes were assessed.

Results We identified four SF clusters with respective marker genes: $PRG4^+$ SF (*CD55, MMP3, PRG4, THY1*^neg); $CXCL12^+$ SF (*CXCL12, CCL2, ADAMTS1, THY1*^low); $POSTN^+$ SF (*POSTN, collagen genes, THY1*); $CXCL14^+$ SF (*CXCL14, C3, CD34, ASPN, THY1*) that correspond to lining ($PRG4^+$ SF) and sublining ($CXCL12^+$ SF, $POSTN^+$ + and $CXCL14^+$ SF) SF subsets. $CXCL12^+$ SF and $POSTN^+$ + were most prominent in the fibroid while $PRG4^+$ SF appeared highest in the myeloid pathotype. Corresponding, lining assessed by histology (assessed by Krenn-Score) was thicker in the myeloid, but also in the lymphoid pathotype + the fibroid pathotype. $PRG4^+$ SF correlated positively with disease severity parameters in the fibroid, $POSTN^+$ SF in the lymphoid pathotype whereas $CXCL14^+$ SF showed negative association with disease severity in all pathotypes.

Conclusion This study shows a so far unexplored association between distinct synovial pathologies and SF subtypes defined by scRNA-seq. The knowledge of the diverse interplay of SF with immune cells will advance opportunities for tailored targeted treatments.

Key messages

What is already known about this subject?
► Different studies were able to show the presence of diverse synovial fibroblast (SF) subsets in the rheumatoid arthritis (RA) synovium using single-cell RNA sequencing.
► In RA, three distinct synovial pathotypes based on the presence of immune cells have been identified which could be connected with specific clinical parameters and outcome measures.

What does this study add?
► We could confirm the presence of specific SF subsets in the synovium and showed the feasibility of integration of single-cell RNA sequencing data from different resources.
► By deconvolution of bulk RNAseq, we show differences in the proportion of SF subtypes within different pathotypes and specific correlations of the various SF subtypes with disease activity/severity dependent on the histological pathotype.

How might this impact on clinical practice or further developments?
► This so far unexplored connection between the histological pathotype and SF subtypes guides further research to understand the impact of the presence of immune cells on SF subtypes, and thus opens new avenues for targeted treatment according to the synovial composition.

INTRODUCTION

During the development of rheumatoid arthritis (RA), synovial architecture and cellular content change dramatically. The thin membrane lining the joint synovium becomes an inflamed, hyperplastic and invasive tissue mass of infiltrating cells—most prominent synovial fibroblasts (SF)—that ultimately lead to joint destruction.[1] In the sublining, the presence of infiltrating immune cells, expanded SF and increased vascularity are characteristic for RA synovium.

Previous research revealed the presence of three distinct synovial pathotypes based on

cellular and molecular analysis of synovial tissue:[2] (1) lymphomyeloid dominated by the presence of B-cells in addition to myeloid cells (hereafter lymphoid); (2) diffuse-myeloid with myeloid lineage predominance but poor in B cells (hereafter myeloid) and (3) pauci-immune characterised by scanty immune cells and prevalent stromal cells (hereafter fibroid). Correlation analysis demonstrated that elevation of myeloid-associated and lymphoid-associated gene expression strongly correlated with disease activity, acute phase reactants and response to disease-modifying antirheumatic drugs at six months.[3] Patients with predominant fibroid pathology showed less severe disease activity and radiographic progression, but poor treatment response.[3]

New single cell sequencing tools, including single-cell RNA sequencing (scRNA-seq), provide deep insight into tissue biology at the cell-state level, uncovering a diversity of synovial lymphoid, myeloid and stromal cell populations.[4–7]

Despite this expanded knowledge and the fact that already early studies showed the presence of different synovitis subtypes[8] and the potential of SF to adapt to the inflammatory status in the synovium,[9] the relationship between SF subsets and the diversity of tissue pathology that is, different synovial pathotypes is unknown. Therefore, we first aimed to confirm specific subsets of SF in synovial tissue by integration of published and unpublished scRNA-seq data. We then assessed the contribution of the defined SF subsets to synovial pathotypes and clinical characteristics using deconvolution analysis of bulk RNA-seq in treatment-naïve early arthritis synovial tissue.

MATERIAL/METHODS

ScRNA-seq of synovial biopsies

The used publicly available datasets are described in online supplemental text 1. Following, we report the additionally used in-house datasets. Synovial biopsies were obtained from one wrist and one metacarpophalangeal (MCP) joint of patients with active RA by ultrasound guided fine-needle biopsy after obtaining informed consent. Both female patients fulfilled the ACR/EULAR classification criteria[10] for RA and had no concomitant therapy during the time of the biopsy. Synovial biopsies were washed with phosphate-buffered saline, mechanically minced and enzymatically digested using Liberase TL (100 µg/mL; Roche) and DNAse I (100 µg/mL; Roche) in RPMI 1640 cell culture medium (Thermo Fisher) for 30 min at 37°C. After stopping the digestion process with fetal calf serum, erythrocytes were lysed with Red Blood Cell Lysis solution (Milteny Biotec). Cells were washed and counted on a LUNA automated cell counter (Logos Biosystems). A total of 15 000 unsorted synovial cells per patient were prepared for single cell analysis using the Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3, the Chromium Chip B Single Cell Kit (10× Genomics) and the Chromium controller (all 10× Genomic). Libraries were sequenced on the

Illumina NovaSeq instrument to a sequence depth of 20 000–70 000 reads per cell. CellRanger (V.2.0.2) from 10× Genomics was used to demultiplex, align the reads to Ensembl reference build GRCh38.p13 and collapse unique molecular identifiers.

Integration of scRNA-seq datasets

Standard Seurat (V.3.3) for R (V.3.6) protocol—a very well established package to analyse scRNA-seq data—was used for the integration.[11] In a sensitivity analysis, the integration step was additionally performed with Liger (V.1.0)[12] as well as Harmony (V.0.1),[13] since these three methods have been found to be superior compared with other integration protocols.[14] If not otherwise stated, default settings were used for each integration method. Online supplementary text 1 gives further information about the integration process.

To identify differentially expressed genes between groups of cells, we used the Wilcoxon rank-sum test. We used a minimum log2 FC of 0.25 for average expression of genes in a cluster relative to the average expression in all other clusters combined. To take multiple testing into account, p values were adjusted by false discovery rate (FDR) controlling. Significant Marker genes (adjusted p<0.05) were sorted by average log2 FC. KEGG Pathways of the Marker Genes were estimated with clusterProfiler R package.

Pseudotime trajectory analysis

Monocle 2 R package (V.2.2) was used to perform pseudotime trajectory analysis.[15] Monocle applies advanced machine learning methods to find transcriptomic changes each cell goes through as part of a dynamic biological process. Once the overall 'trajectory' of gene expression changes has been identified, each cell gets placed at its proper position in the trajectory. Furthermore, Monocle tracks changes as a function of progress along the trajectory, which is termed 'pseudotime''. After ordering of cells along the trajectory, genes that change as a function of pseudotime will be identified. Most significant genes with similar trends over pseudotime were grouped together with the function 'plot_pseudotime_heatmap'. All tasks were performed with default settings.[16]

Deconvolution analysis of SF subsets across synovial pathotypes

Bulk RNA-seq data from synovial tissue as well as corresponding cellular (immunohistology with pathotype grading) and clinical characteristics were available from 87 treatment-naïve early RA patients (<12 months symptoms duration) in the Pathobiology of Early Arthritis Cohort (PEAC).[17] Information about the PEAC-Cohort was previously described.[2 3] From the 87 patients, 16 had a fibroid, 45 a lymphoid and 20 a myeloid pathotype; in 6 patients the grading was not possible.

Fastq files were downloaded fromthe European Bioinformatics Institute with the accession code E-MTAB-6141 and were mapped to hg19 and sequence reads assigned
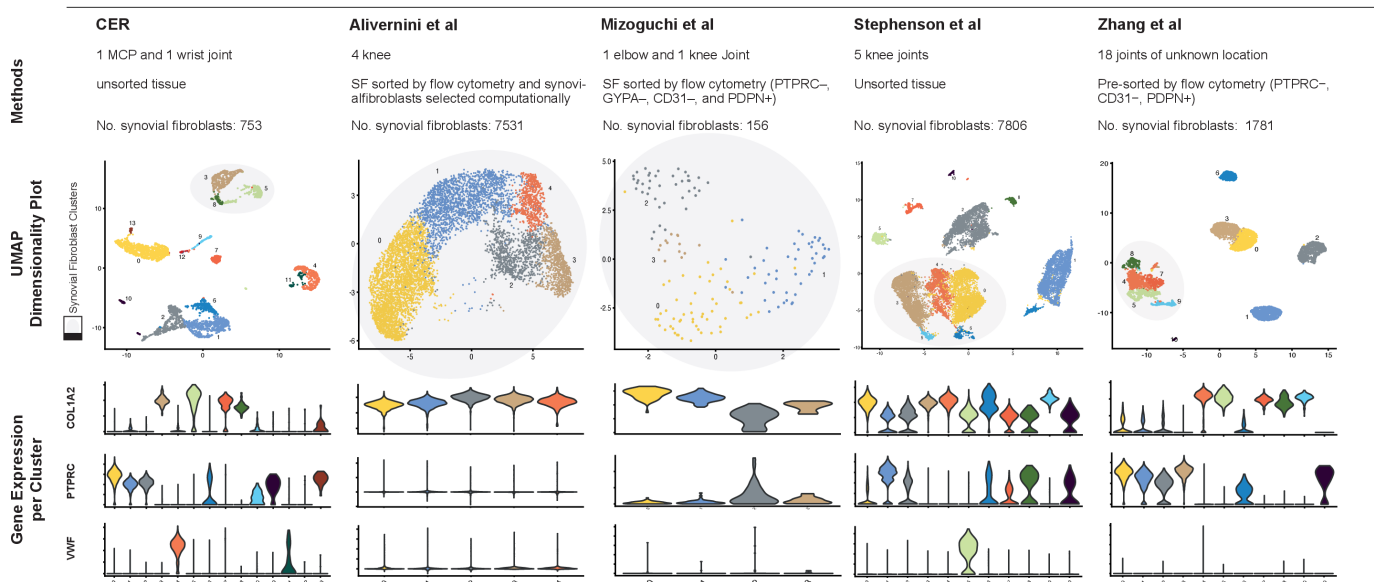
DATASETS

Figure 1 Individual synovial tissue datasets with respective used methods and number of synovial fibroblasts (SF), UMAP visualisation and SF positive (COL1A2) and negative (*PTPRC*, *vWF*) marker genes. Selected SF clusters in unsorted datasets are marked in grey in the UMAP plot. CER = Center of Experimental Rheumatology Zurich, UMAP = Uniform Manifold Approximation and Projection for Dimension Reduction. UMAP; Uniform Manifold Approximation and Projection for Dimension Reduction.

to genomic features using STAR[18] and featureCounts,[19] respectively.

MuSiC R package was used to perform the deconvolution analysis.[20] MuSiC uses cell-type specific gene expression from scRNA-seq data to characterise cell type compositions in bulk RNA-seq. A standard protocol was used to estimate cell type proportions in bulk tissue.[21]

Pairwise-Wilcox test was used to calculate pairwise comparisons between group levels. SF cluster proportions were analysed for correlation with disease activity measures across different synovial pathotypes.

P values were adjusted for multiple testing by FDR controlling.

Histological analysis
The pathotype and Krenn lining score[22] were assessed in a subset of 69 synovial samples from patients with RA as previously described.[2]

RESULTS
Integration of five scRNA-seq datasets
We first collected scRNA-seq data from 31 RA synovial tissues of four published studies and of an in-house dataset (figure 1 and online supplementary text 1). SFs were selected based on the presence of *COL1A2* and absence of *PTPRC* (CD45, leucocyte marker) and *VWF* (von Willebrand factor, endothelial cell marker) gene expression. Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) visualisation showed that the expression of these genes was uniformly present in neighbouring clusters apart from cluster 7 in the samples from the Center of Experimental

Rheumatology Zurich (CER). The main marker genes of cluster 7 were suggestive of a smooth muscle origin (eg, *MYH11*, *ACTA1* and *ACTA2*) and thus the cluster was excluded from further analysis. The studies of Alivernini *et al*[7] and Stephenson *et al*[4] had a substantially higher number of SF cells compared with the other datasets (figure 1). To correct for this imbalance, we randomly reduced the number of identified SF of these studies to 3000 and used the remaining 8693 SF for further analysis.

Generation SF signatures
Integration and clustering of the identified SF cells using Seurat, Liger or Harmony identified similar clusters of SF illustrated by the expression of marker genes determined in previous studies: *CD55, THY1, CD34, POSTN* and *HLA-DRA* (figure 2A). In all three approaches, four SF subtypes were distinguishable (figure 2A). The top 20 marker genes of the Seurat, Liger and Harmony determined SF clusters are presented in online supplemental tables 1-3.

The first subtype (*PRG4*+ SF) expressed high levels of *CD55, MMP3, PRG4 and FN1*, and was *THY1* negative. This subtype was previously suggested to reside in the lining layer of the synovium.[4–7] The second subtype (*CXCL12*+ SF) expressed *CXCL12, CCL2 and ADAMTS1* and had the lowest *THY1* expression of the three *THY1* positive subtypes. The third subtype (*POSTN*+ SF) showed high expression of *POSTN* and collagen genes. This subtype expressed intermediate levels of *THY1*. The fourth subtype (*CXCL14*+ SF) expressed high levels of *CXCL14, C3, ASPN, THY1* and *CD34*. Liger identified an additional fifth subtype. This subtype contained *CD55*
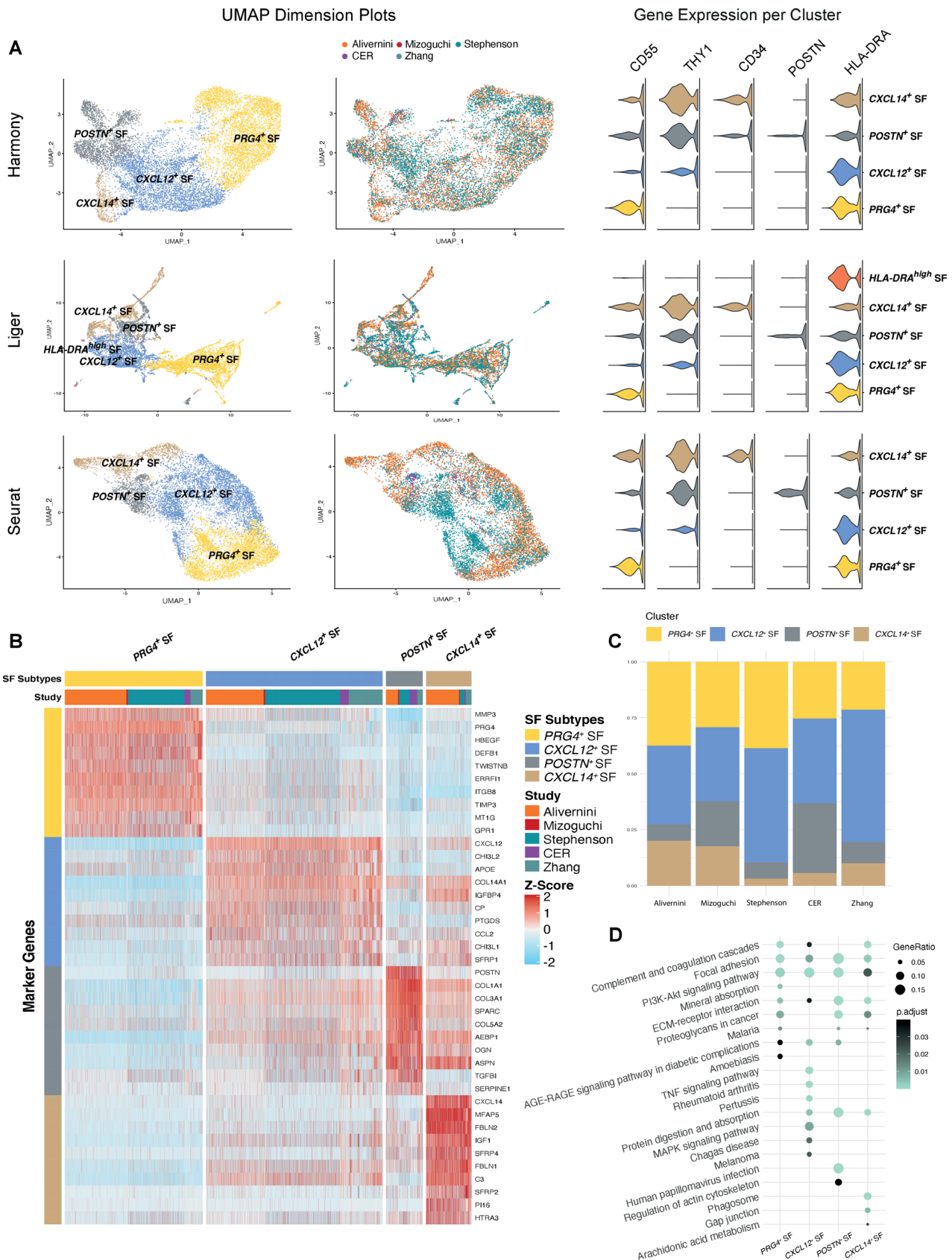
Figure 2 Identification of synovial fibroblast (SF) subsets (A) different integration methods with UMAP visualisation, grouping of cells according to datasets and violinplots showing the gene expression per cluster of *CD55, THY1, CD34, POSTN* and *HLA-DRA*. (B) Heatmap of the 10 most significant marker genes of each SF cluster (via Seurat). (C) Distribution of seurat clusters across the different datasets. (D) KEGG pathways enrichment analysis across SF clusters. Top 20 pathways are shown. UMAP; Uniform Manifold Approximation and Projection for Dimension Reduction.

and *THY1* negative cells with high expression of *HLA-DRA* (figure 2A). The cluster represented a minor population of cells, some of which were also scattered across the other four subtypes. Therefore, this subtype was not considered in the further analysis.

UMAP with colour-coding according to the data source confirmed that the four subtypes were derived from data across all data sets (figure 2A). However, the distribution of the SF subsets differed between the studies (figure 2C). The variation in distribution of the subtypes may be due to the differences in material and methods used. But even in the two studies with the most similar approach (Mizoguchi *et al*[5] and Zhang *et al*[6]) the distribution of the subtypes varied and two studies with a very different approach (CER and Mizoguchi *et al*[5]) showed similar patterns of SF subtype distribution. Differences in patient selection (disease stage and activity), previous and current treatment and joint location might contribute to variation in subtype distribution between the studies.

We then performed KEGG pathway analysis across the SF subtypes (figure 2D, online supplementary table 4). *PRG4*[+] SF genes were enriched for terms such as +'focal adhesion', 'ECM-receptor interaction' and 'mineral absorption', which supported the assignment of these SF to the synovial lining layer. *CXCL12*[+] SF genes were enriched for different pathways associated with proinflammatory states (+ 'RA', 'TNF signalling pathway', 'MAPK signalling pathway'). Pathways that were unique for *POSTN*[+] + were 'human papillomavirus infection' (included genes: *COL1A1, COL3A1, COL1A2, ACTN1, LAMB1*) and 'regulation of actin cytoskeleton' (included genes: *ITGA10, MYLK, ACTN1, MYH10, PDGFRB, MYL9, ENAH, ITGA11, ITGB5*). Most significant pathways of *CXCL14*[+] + were 'phagosome', 'gap-junction' and 'arachidonic acid metabolism'.

Assessment of dynamic relationships between SF subtypes

After the investigation of different gene expression profiles of the SF clusters, we assessed the relationships between the different clusters using monocle 2 (figure 3). A continuum over five different cell states was found by advanced machine learning technique (figure 3A). When overlaid with the SF subtypes, it appeared that the cells dispersed along the trajectory beginning in *PRG4*[+] + over *CXCL12*[+] + to *CXCL14*[+] SF (figure 3B and C). *CXCL12*[+] SF additionally formed two small branches of cell states. The state at one side of the trajectory - with mostly containing PRG4[+] SF – was defined as starting point for further analysis. Differential expression analysis over pseudotime identified several genes that changed significantly during the transition (online supplementary table 5). Figure 3D shows the 50 most significant changing genes over the pseudotime trajectory in a heatmap. Included in these genes are *THY1* and *PRG4,* genes that were previously described as relevant positional markers of SF in the synovium[23] in sublining and lining SF, respectively (figure 3D and E).

In summary, we successfully integrated five datasets with which we could recapitulate previously identified SF subtypes as well as the plasticity of these cellular states according to their location within the synovium.

Distribution of SF phenotypes across synovial pathotypes

We then integrated and clustered all synovial cells to be able to perform the deconvolution analysis of bulk transcriptomics data of synovial tissues (online supplementary table 6 and supplementary figure 1). Deconvolution analysis revealed a distribution of the different cell types as the respective pathotypes would suggest (figure 4A). The proportion of SF was highest in the fibroid pathotype, while the lymphoid pathotype showed a clear enrichment in myeloid cells, plasma cells, B cells and T cells (figure 4A). The myeloid pathotype was denoted by high percentage of myeloid cells but absence of T and B cells. Overall, these results confirmed the accuracy of our approach.

Analysis of SF subtype enrichment within the various pathotypes suggested that *POSTN*[+] + and *CXCL14*[+] + are most prominent in the fibroid pathotype, while the proportion of *PRG4*[+] SF expanded in the myeloid pathotype (+). Statistical analysis confirmed that *PRG4*[+] + had higher proportions in the myeloid pathotype compared with the fibroid pathotype (figure 4B). Accordingly, lining thickness assessed by histology was more pronounced in the myeloid, but also in the lymphoid pathotype compared with the fibroid pathotype (online supplementary figure 2). *CXCL12*[+] + and *POSTN*[+] + had the highest proportions in the fibroid pathotype, followed by the lymphoid pathotype in *CXCL12*[+] + and the myeloid pathotype in *POSTN*[+] SF, respectively (+).

In summary, these data suggest an expansion of lining *PRG4*[+] SF fibroblasts in the myeloid pathotype and increased proportions of *CXCL12*[+] SF and *POSTN*[+] + in the fibroid pathotype.

Correlation of SF subtypes with clinical characteristics

To assess the connection of the different subtypes with pathological processes and clinical symptoms in RA, we correlated demographic and clinical data with the presence of the SF subtypes (figure 4C,D, online supplementary table 7). Proportions of SF were neither different between men and women nor between seropositive and seronegative patients or different age groups. Without differentiating the patients according to the synovial pathotype, there was a significant negative relationship between *CXCL14*[+] + and swollen joint count (R=−0.61, adjusted p=value 0.047) as well as Disease Activity Score-28 (DAS28 (R=−0.64, adjusted p<0.001). Also, within the different pathotypes, proportions of *CXCL14*[+] + were generally negatively correlated with clinical parameters. Specifically in the myeloid pathotype, *CXCL14*+ SF showed the strongest negative association with DAS28 (R=−0.98; p=0.0025, adjusted p=0.047).

Even though *CXCL12*[+] + was characterised by inflammatory signalling pathways, its presence did not correlate
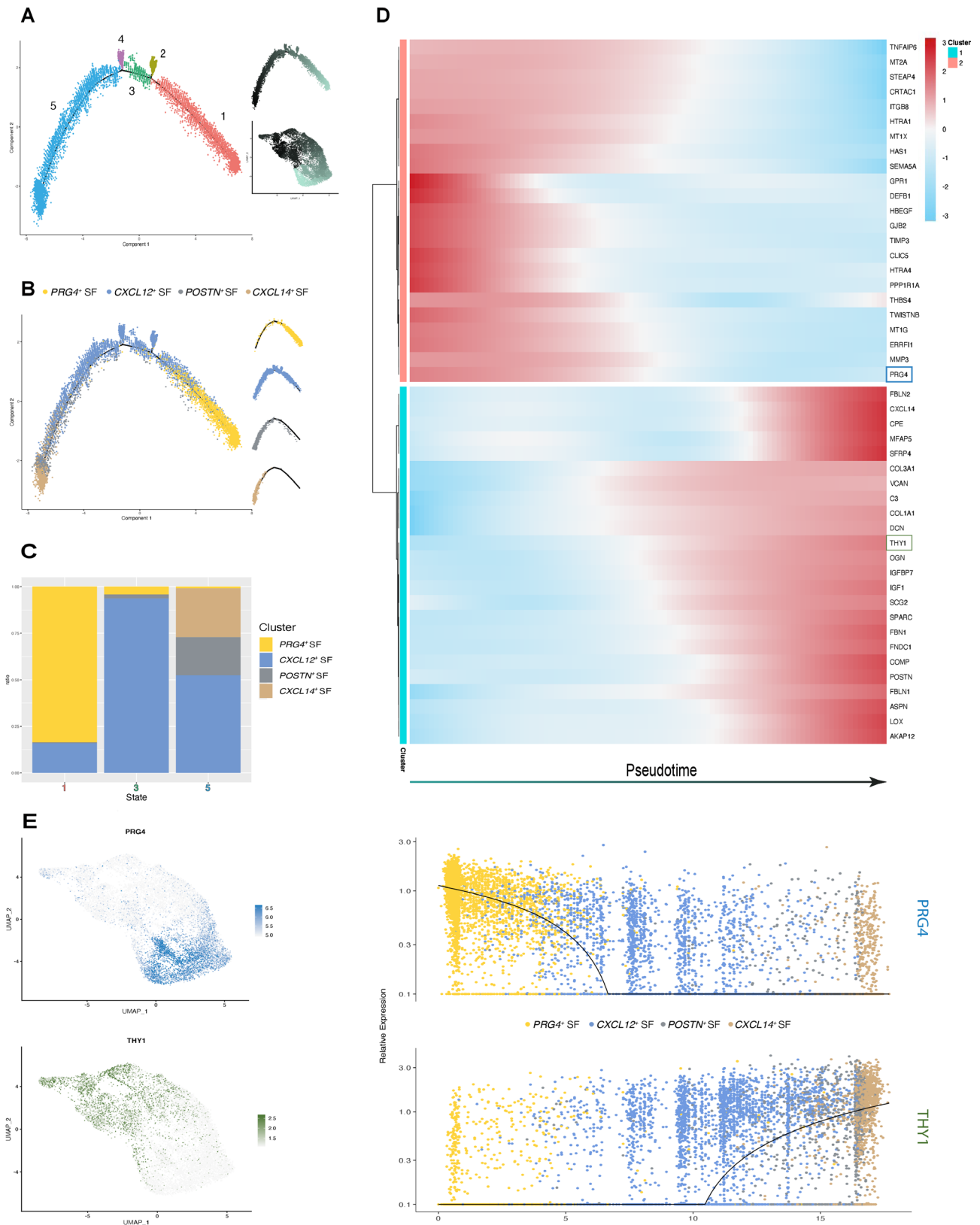
Figure 3 Pseudotime analysis of synovial fibroblast (SF) subsets. (A) Visualisation of the trajectory states of all SF. Top and middle right shows the pseudotime trajectory in the reduced dimension and UMAP visualisation. (B) Distribution of the SF cluster along the trajectory (right split by cluster). (C) Distribution of SF clusters in the main trajectory states 1, 3 and 5. (D) Most significant genes that covary across pseudotime split in two clusters (state one left, state five right). (E) Plot of gene expression levels of *PRG4* and *THY1* in the UMAP of the SF clusters on the left and along the pseudotime trajectory on the right. UMAP = Uniform Manifold Approximation and Projection for Dimension Reduction.
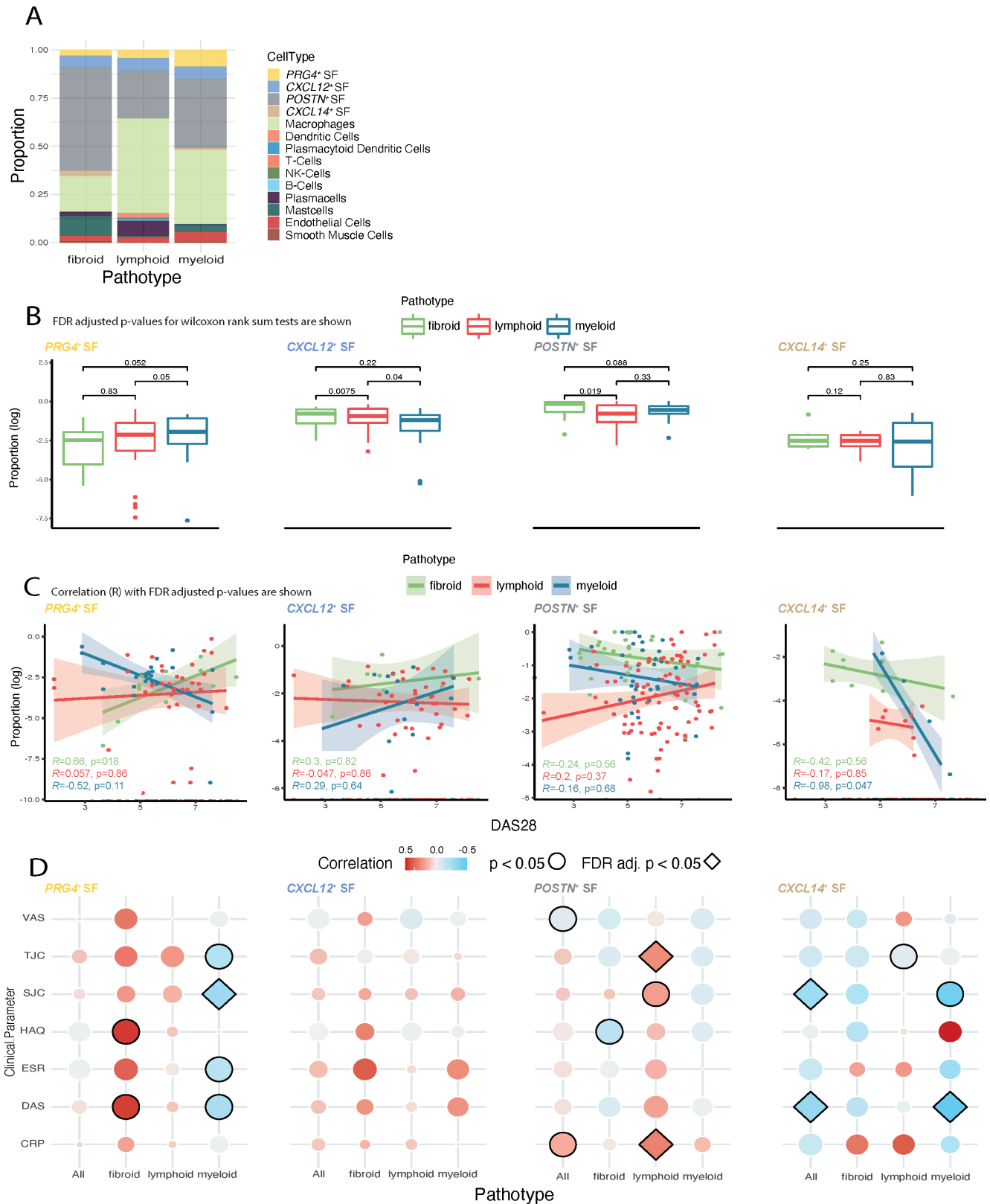
Figure 4

Figure 4 Synovial fibroblasts (SF) across synovial pathotypes and comparison with disease activity measurements. (A) Distribution of different cell types between the different synovial pathotypes. (B) Proportion of the different SF subtypes within the pathotypes. (C) Correlation between disease activity measurements and SF proportions across synovial pathotypes. (D) Dotplot with Pearson correlation of different clinical parameters with SF subtypes across pathotypes. DAS; disease activity score. ESR; erythrocyte sedimentation rate. FDR; false discovery rate. HAQ; health assessment questionnaire. SJC; swollen joint count. TJC; tender joint count. VAS; visual analog scale.

with any clinical symptoms. Strikingly for *PRG4*⁺ + and *POSTN*⁺ SF, correlations were opposite in the different pathotypes. In the lymphoid pathotype, increasing *POSTN*⁺ SF proportions were related with higher tender and swollen joint counts and CRP levels, but in the fibroid and lymphoid pathotype these correlations tended to be reversed (figure 4D and online supplementary table 7). Similarly for *PRG4*⁺ SF, in the fibroid pathotype a positive relationship with disease activity (+=0.66, p=0.02, adjusted p=0.18) and with the Health Assessment Questionnaire (R=0.67, p=0.017, adjusted p=0.18) was seen, but in the myeloid pathotype a negative correlation with disease activity (R=−0.52; p=0.007, adjusted p=0.11), particularly a negative association with the swollen tender joint count (R=−0.63; p<0.001; adjusted p=0.023) was measured (figure 4D). Together these data point to a substantial influence of immune cells on shaping gene expression of *PRG4*⁺ + and *POSTN*⁺ + in arthritic synovium.

We also looked at the relationship of other synovial cell types with DAS28 (online supplementary figure 3); however, without further division of these additional cell types in subtypes, no significant relationship was found.

DISCUSSION

In this study, we were able to integrate synovial single-cell data from different sources and protocols and confirmed the presence of four distinct SF subpopulations in RA synovium.[4–7] Furthermore, we could show that the different SF subtypes vary in their distribution within pathotypes and correlate with distinct clinical disease characteristics dependent on the pathotype. Thus, with our data, we revealed a presently unexplored connection between the SF subtypes and the type of immune cell infiltration in RA synovium.

The SF subtypes that we defined with the integrated dataset are largely in line with previously identified SF subtypes. All the previously published datasets obtained a subset of SF with high expression of *CD55* and absence of *THY1* expression. Further analysis—using staining of CD55 in the synovium[4] and high expression of known SF lining genes[5]—concluded that these cells most likely represent lining SF. In the fibroid pathotype, the proportions of *PRG4*⁺ SF showed the most positive correlations with clinical parameters, leading to the hypothesis that in low presence of immune cells the lining *PRG4*⁺ SF subset might play a stronger role in driving RA symptoms as compared with the myeloid pathotype, where proportions of *PRG4*+ SF mostly negatively correlated with clinical parameters. It has been postulated that lining SF mediate tissue damage and sublining SF coordinate inflammatory responses.[24] Serum *MMP3*—the most significant marker gene of *PRG4*⁺ SF—has a known association with disease activity and joint damage.[25 26] However, previous studies from the PEAC cohort found that the fibroid pathotype was associated with less radiographic progression compared with the other pathotypes,[2] implicating a role of additional cell types in mediating tissue

destruction. Notably, *PRG4*⁺ SF proportions were higher in the myeloid pathotype + the others. This is in concordance with a previous study showing that lining SFs are positively correlated with macrophage density.[27] Surprisingly, *PRG4*⁺ SF proportions negatively correlated with clinical parameters in the myeloid pathotype suggesting that the increased presence of macrophages might have a regulatory effect on this SF subtype.

CXCL12⁺ SF subset of integrated data set showed high expression of *CXCL12* and *HLA-DRA* and no expression of *CD34*, consistent with an SF subtype described by Alivernini[7] and Zhang[6] +, in which the corresponding associated genes were more highly expressed in leucocyte-rich RA than in OA.[6]

The *POSTN*⁺ SF subset was + by high expression of *POSTN*. Also, Mizoguchi *et al*[5] found high expression of *POSTN* in the *CD34/THY*⁺ subset and an association of the respective SF subset with proportion of infiltrated leucocytes, histological synovitis, and synovial hypertrophy by ultrasound. In accordance, we found a positive correlation of this SF subtype with clinical parameters of active RA, but only in the lymphoid pathotype suggesting that interaction with lymphocytes may influence the pathogenic role of this SF subtype.

The *CXCL14*⁺ SF subset expressed *CD34*, as well as *CXCL14*. Based on the expression of inflammatory cytokines and higher number of recruited peripheral blood monocytes in a transwell + assay, Mizoguchi *et al*[5] suggested that *CD34*^pos SF are mostly responsible for monocyte recruitment in inflamed synovial tissue. However, the latter analysis was done with cultured SF subsets, which might alter their behaviour. Alivernini *et al*[7] separated a *CD34*⁺ + a *CXCL14*⁺ SF subsets. The *CXCL14*⁺ SF in Alivernini *et al*[7] was further characterised by high expression of *GAS6*, regulating the function of synovial tissue macrophages in remission (MerTK^pos CD206^pos). *GAS6* expression in *CXCL14*⁺ + was higher in remission compared with active RA in Alivernini *et al*,[7] which points to a possible anti-inflammatory role of this SF subtype. In accordance, in our analysis the presence of the *CXCL14*⁺ SF subtype consistently negatively correlated with clinical symptoms of RA. Thus, *CXCL14*⁺ SF might represent a regulatory SF subtype attenuating inflammation. Besides this assumed influence of immune cells on SF, the other way around is possible as well; SF could be the initial driver and attract immune cells via different mechanism. via different mechanisms.

Using a deconvolution method, we were able to estimate proportions of various cell types within specific synovial pathotypes in the PEAC cohort. Most interestingly, the fibroid pathotype was characterised not only by a majority of SF but also by a high proportion of mast cells, which has not been studied in this context so far.

The limitation of this study includes that deconvolution is only a computational model, and further single cell analysis within the pathotypes will have to be performed to validate our data. Furthermore, functional analyses of SF subtypes are needed to investigate the relation

between different SF subsets within pathotypes in detail. Nonetheless, our study has brought important insights for further research in this direction.

In conclusion, our study shows a so far unexplored association between the type of immune cell infiltration and the formation of SF subtypes. Knowledge of the influence of different SF subtypes on joint inflammation will enhance our understanding of the pathogenesis of RA, open new avenues for tailored targeted treatment according to the pathotype and reveal therapeutic targets for influencing the activated stroma in RA.

Author affiliations
[1]Center of Experimental Rheumatology, Department of Rheumatology, University Hospital Zurich, Zurich, Switzerland
[2]Research Into Inflammatory Arthritis Centre Versus Arthritis (RACE), Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, UK
[3]Centre for Experimental Medicine and Rheumatology, Barts and The London School of Medicine and Dentistry, William Harvey Research Institute, London, UK
[4]Department of Rheumatology, University Medical Center Ljubljana, Ljubljana, Slovenia

Twitter Raphael Micheroli @ramicheroli and Caroline Ospelt @CarolineOspelt

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study as approved by National Research Ethics Service Committee London- Dulwich, Reference number 05/Q0703/198.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. All data relevant to the study are included in the article or uploaded as online supplemental information. All data relevant to the study are included in the article, uploaded as online supplemental information or already publicly available.

ORCID iDs
Raphael Micheroli http://orcid.org/0000-0002-8918-7304
Adrian Ciurea http://orcid.org/0000-0002-7870-7132
Caroline Ospelt http://orcid.org/0000-0002-9151-4650

REFERENCES
1 Turner JD, Filer A. The role of the synovial fibroblast in rheumatoid arthritis pathogenesis. *Curr Opin Rheumatol* 2015;27:175–82.
2 Humby F, Lewis M, Ramamoorthi N, *et al*. Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Ann Rheum Dis* 2019;78:761–72.
3 Lewis MJ, Barnes MR, Blighe K, *et al*. Molecular portraits of early rheumatoid arthritis identify clinical and treatment response phenotypes. *Cell Rep* 2019;28:2455–70.
4 Stephenson W, Donlin LT, Butler A, *et al*. Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation. *Nat Commun* 2018;9:791.
5 Mizoguchi F, Slowikowski K, Wei K, *et al*. Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. *Nat Commun* 2018;9:789.
6 Zhang F, Wei K, Slowikowski K, *et al*. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol* 2019;20:928–42.
7 Alivernini S, MacDonald L, Elmesmari A, *et al*. Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis. *Nat Med* 2020;26:1295–306.
8 Holweg CTJ, Holweg CTJ, Kummerfeld SK, *et al*. Synovial phenotypes in rheumatoid arthritis correlate with response to biologic therapeutics. *Arthritis Res Ther* 2014;16:R90.
9 Timmer TCG, Timmer TCG, Smeets TJ, *et al*. Fibroblast-like synoviocytes derived from patients with rheumatoid arthritis show the imprint of synovial tissue heterogeneity: evidence of a link between an increased myofibroblast-like phenotype and high-inflammation synovitis. *Arthritis Rheum* 2005;52:430–41.
10 Aletaha D, Neogi T, Silman AJ, *et al*. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/ European League against rheumatism collaborative initiative. *Ann Rheum Dis* 2010;69:1580–8.
11 Stuart T, Butler A, Hoffman P, *et al*. Comprehensive integration of single-cell data. *Cell* 2019;177:1888–902.
12 Welch JD, Kozareva V, Ferreira A, *et al*. Single-Cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* 2019;177:1873–87.
13 Korsunsky I, Millard N, Fan J, *et al*. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods* 2019;16:1289–96.
14 Tran HTN, Ang KS, Chevrier M, *et al*. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol* 2020;21:12.
15 Trapnell C, Cacchiarelli D, Grimsby J, *et al*. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32:381–6.
16 Monocle. Available: http://cole-trapnell-lab.github.io/monocle-release/
17 Array Express. E-MTAB-6141 - Comprehensive transcriptome analysis of synovium and peripheral blood reveals major disease heterogeneity early in the RA disease process prior to therapeutic intervention. Available: https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6141/samples/
18 Dobin A, Davis CA, Schlesinger F, *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
19 Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
20 Wang X, Park J, Susztak K, *et al*. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* 2019;10:380.
21 GitHub. xuranw / MuSiC. Available: https://github.com/xuranw/ MuSiC
22 Krenn V, Rüther W, Rüther W, *et al*. 15 years of the histopathological synovitis score, further development and review: a diagnostic score for rheumatology and orthopaedics. *Pathol Res Pract* 2017;213:874–81.
23 Wei K, Korsunsky I, Marshall JL, *et al*. Notch signalling drives synovial fibroblast identity and arthritis pathology. *Nature* 2020;582:259–64.
24 Ospelt C. Synovial fibroblasts in 2017. *RMD Open* 2017;3:e000471.
25 Galil SMA, El-Shafey AM, Hagrass HA, *et al*. Baseline serum level of matrix metalloproteinase-3 as a biomarker of progressive joint damage in rheumatoid arthritis patients. *Int J Rheum Dis* 2016;19:377–84.
26 Mahmoud RK, El-Ansary AK, El-Eishi HH, *et al*. Matrix metalloproteinases MMP-3 and MMP-1 levels in sera and synovial fluids in patients with rheumatoid arthritis and osteoarthritis. *Ital J Biochem* 2005;;54:248–57.
27 Izquierdo E, Cañete JD, Celis R, *et al*. Synovial fibroblast hyperplasia in rheumatoid arthritis: clinicopathologic correlations and partial reversal by anti-tumor necrosis factor therapy. *Arthritis Rheum* 2011;63:2575–83.