

1 Cellular deconstruction of inflamed synovium defines diverse 2 inflammatory phenotypes in rheumatoid arthritis

3

4 Authors

5 Fan Zhang^{1,2,3,4,5,6,28}, Anna Helena Jonsson^{1,28}, Aparna Nathan^{1,2,3,4,5,28}, Kevin Wei^{1,28}, Nghia
6 Millard^{1,2,3,4,5,28}, Qian Xiao^{1,2,3,4,5}, Maria Gutierrez-Arcelus^{1,2,3,4,5,7}, William Apruzzese⁸, Gerald F.
7 M. Watts¹, Dana Weisenfeld¹, Joyce B. Kang^{1,2,3,4,5}, Laurie Rumker^{1,2,3,4,5}, Joseph Mears^{1,2,3,4,5},
8 Kamil Slowikowski^{4,5,9,10}, Kathryn Weinand^{1,2,3,4,5}, Dana E. Orange^{11,12}, Javier Rangel-Moreno¹³,
9 Laura Geraldino-Pardilla¹⁴, Kevin D. Deane¹⁵, Darren Tabechian¹³, Arnold Ceponis¹⁶, Gary S.
10 Firestein¹⁶, Mark Maybury¹⁷, Ilfita Sahbudin¹⁷, Ami Ben-Artzi¹⁸, Arthur M. Mandelin II¹⁹,
11 Alessandra Nerviani²⁰, Felice Rivellesse²⁰, Costantino Pitzalis²⁰, Laura B. Hughes²¹, Diane
12 Horowitz²², Edward DiCarlo²³, Ellen M. Gravallesse¹, Brendan F. Boyce²⁴, Accelerating
13 Medicines Partnership Program: Rheumatoid Arthritis and Systemic Lupus Erythematosus
14 (AMP RA/SLE) Network, Larry W. Moreland^{15,25}, Susan M. Goodman^{11,26}, Harris Perlman¹⁹, V.
15 Michael Holers¹⁵, Katherine P. Liao^{1,4}, Andrew Filer¹⁷, Vivian P. Bykerk^{11,26}, Deepak A. Rao^{1,*},
16 Laura T. Donlin^{11,26,*}, Jennifer H. Anolik^{13,*}, Michael B. Brenner^{1,*}, Soumya
17 Raychaudhuri^{1,2,3,4,5,27,*}

18

19

20 Affiliations

21 ¹Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Brigham and
22 Women's Hospital and Harvard Medical School, Boston, MA, USA.

23 ²Center for Data Sciences, Brigham and Women's Hospital, Boston, MA, USA.

24 ³Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard
25 Medical School, Boston, MA, USA.

26 ⁴Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.

27 ⁵Broad Institute of MIT and Harvard, Cambridge, MA, USA.

28 ⁶Division of Rheumatology and the Center for Health Artificial Intelligence, University of
29 Colorado School of Medicine, Aurora, CO, USA.

30 ⁷Division of Immunology, Department of Pediatrics, Boston Children's Hospital and Harvard
31 Medical School, Boston, MA. US.

32 ⁸Accelerating Medicines Partnership® Program: Rheumatoid Arthritis and Systemic Lupus
33 Erythematosus (AMP® RA/SLE) Network

34 ⁹Center for Immunology and Inflammatory Diseases, Department of Medicine, Massachusetts
35 General Hospital (MGH), Boston, MA, USA

36 ¹⁰MGH Cancer Center, Boston, MA, USA

37 ¹¹Hospital for Special Surgery, New York, NY, USA.

38 ¹²Laboratory of Molecular Neuro-Oncology, The Rockefeller University, New York, NY, USA.

39 ¹³Division of Allergy, Immunology and Rheumatology; Department of Medicine, University of
40 Rochester Medical Center, Rochester, NY, USA.

41 ¹⁴Division of Rheumatology, Columbia University College of Physicians and Surgeons, New
42 York, NY, USA.

43 ¹⁵Division of Rheumatology, University of Colorado School of Medicine, Aurora, CO, USA.

44 ¹⁶Division of Rheumatology, Allergy and Immunology, University of California, San Diego, La
45 Jolla, CA, USA.

46 ¹⁷Rheumatology Research Group, Institute for Inflammation and Ageing, University of
47 Birmingham, NIHR Birmingham Biomedical Research Center and Clinical Research Facility,
48 University of Birmingham, Queen Elizabeth Hospital, Birmingham, UK.

49 ¹⁸Division of Rheumatology, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

50 ¹⁹Division of Rheumatology, Department of Medicine, Northwestern University Feinberg School
51 of Medicine, Chicago, IL, USA.

52 ²⁰Centre for Experimental Medicine & Rheumatology, William Harvey Research Institute, Queen
53 Mary University of London; London, UK.

54 ²¹Division of Clinical Immunology and Rheumatology, Department of Medicine, University of
55 Alabama at Birmingham, Birmingham, AL, USA.

56 ²²Feinstein Institute for Medical Research, Northwell Health, Manhasset, New York, NY, USA.

57 ²³Department of Pathology and Laboratory Medicine, Hospital for Special Surgery; New York,
58 NY, USA.

59 ²⁴Department of Pathology and Laboratory Medicine, University of Rochester Medical Center,
60 Rochester, NY, USA

61 ²⁵Division of Rheumatology and Clinical Immunology, University of Pittsburgh School of
62 Medicine; Pittsburgh, PA, USA.

63 ²⁶Weill Cornell Medicine, New York, NY, USA

64 ²⁷Versus Arthritis Centre for Genetics and Genomics, Centre for Musculoskeletal Research,
65 Manchester Academic Health Science Centre, The University of Manchester, Manchester,
66 UK.

67 ²⁸These authors contributed equally

68 *These authors jointly supervised this work

69 **Summary**

70 Rheumatoid arthritis (RA) is a prototypical autoimmune disease that causes destructive tissue
71 inflammation in joints and elsewhere. Clinical challenges in RA include the empirical selection of
72 drugs to treat patients, inadequate responders with incomplete disease remission, and lack of a
73 cure. We profiled the full spectrum of cells in inflamed synovium from patients with RA with the
74 goal of deconstructing the cell states and pathways characterizing pathogenic heterogeneity in
75 RA. Our multicenter consortium effort used multi-modal CITE-seq, RNA-seq, and histology of
76 synovial tissue from 79 donors to build a >314,000 single-cell RA synovial cell atlas with 77 cell
77 states from T, B/plasma, natural killer, myeloid, stromal, and endothelial cells. We stratified
78 tissue samples into six distinct cell type abundance phenotypes (CTAPs) individually enriched
79 for specific cell states. These CTAPs demonstrate the striking diversity of RA synovial
80 inflammation, ranging from marked enrichment of T and B cells (CTAP-TB) to a congregation of
81 specific myeloid, fibroblast, and endothelial cells largely lacking lymphocytes (CTAP-EFM).
82 Disease-relevant cytokines, histology, and serology metrics are associated with certain CTAPs.
83 This comprehensive RA synovial atlas and molecular, tissue-based CTAP stratification reveal
84 new insights into RA pathology and heterogeneity, which could lead to novel targeted-treatment
85 approaches in RA.

86

87 **Introduction**

88 Rheumatoid arthritis (RA) is a systemic autoimmune disease affecting up to 1% of the
89 population¹. It causes synovial joint tissue inflammation and extra-articular manifestations that
90 lead to pain, damage, disability²⁻⁵. The clinical course of RA has been transformed by targeted
91 therapeutics, including those aimed at TNF, IL-1, IL-6, B cells, T cell co-stimulation, and the
92 JAK-STAT pathway^{2,6}. Unfortunately, many patients are refractory to these therapies and do not
93 achieve remission. Less than 25% of patients achieve an ACR70 response to any subsequent

94 treatment after failing first-line therapies⁷⁻⁹. While current treatments can partially ameliorate
95 disease activity, there is no cure. Thus, there is a clinical need for new RA treatment targets and
96 an improved ability to predict patient-specific responses to treatment.

97
98 Genetic diversity and highly variable responses to targeted therapeutics suggest that RA may
99 be a heterogeneous disease¹⁰⁻¹³. For example, patients who produce antibodies specific for
100 cyclic citrullinated peptides (CCP) have different HLA and non-HLA susceptibility factors
101 compared to CCP-negative patients¹⁴. However, genetic differences and clinical differences in
102 disease duration or activity have not reliably predicted treatment response or druggable targets
103 thus far¹⁵⁻¹⁸.

104
105 A more granular understanding of tissue inflammation and cell states may reveal synovial
106 phenotypes that could inform prognosis and potentially identify new treatment targets.
107 Encouragingly, preliminary clinical trials using histological or bulk RNA-seq analysis of tissue
108 suggest treatment response may depend on tissue cellular composition^{19,20}. We and others
109 previously identified specific effector cell states in RA pathophysiology that represent promising
110 treatment targets including pro-inflammatory *HBEGF*⁺*IL1B*⁺ macrophages, *MERTK*⁺
111 macrophages, *ITGAX*⁺*TBX21*⁺ autoimmune-associated B cells (ABCs), *PDCD1*⁺ peripheral
112 helper T (T_{PH}) cells, and *NOTCH3*⁺ synovial fibroblasts²¹⁻²⁷. We do not yet know whether this is
113 a comprehensive list of disease-associated populations and if these disease-associated
114 populations are present in every patient with RA.

115
116 To deconstruct the inflammatory cellular components of RA synovium, we analyzed cell-state
117 composition in a diverse set of patients with clinically active RA. We sought to determine
118 whether certain states are enriched only in certain subsets of patients. Since RA is a
119 prototypical autoimmune disease that shares disease-associated tissue cell states^{23,28-32} and

120 risk loci with other autoimmune diseases^{33,34}, these analyses may offer insights into other
121 diseases in which tissue inflammation is a hallmark.

122

123 **Results**

124 To characterize RA patient heterogeneity, we utilized a multimodal single-cell synovial tissue
125 pipeline to stratify tissue samples into distinct subgroups, characterize their associated cell
126 states, and identify their clinical and histologic associations (**Figure 1A-D**).

127

128 **Collection of synovial samples from RA patients.**

129 We recruited patients exhibiting moderate to high disease activity (100% with CDAI \geq 10; 80.6%
130 with DAS28-CRP \geq 3.2) and obtained synovial tissue biopsies. To capture the full diversity of
131 RA, we recruited treatment-naive patients (n=28) early in their disease course (mean 2.64
132 years), methotrexate-inadequate (MTX) responders (n=27), and anti-TNF agent inadequate
133 responders (n=15). The patients were similar in age, sex, disease activity, and other clinical
134 parameters across the three treatment groups (**Supplementary Table 1**). For comparison, we
135 obtained tissues from patients with osteoarthritis (OA, n=9). We assayed a total of 82 synovial
136 tissue samples, including three pairs of samples from RA patients biopsied at two separate
137 times. Three pathologists independently scored each sample for lining layer hyperplasia, cell
138 density, and aggregates³⁵, and observed that only cell density was different among patient
139 groups ($p=0.005$, **Supplementary Table 1**).

140

141 **Multimodal single-cell integration defines major cell types**

142 We used CITE-seq to simultaneously characterize the full transcriptome and surface expression
143 of 58 proteins, for which we developed and optimized an oligo-conjugated antibody CITE-seq
144 panel spanning key immune and stromal cell lineage and functional markers (**Supplementary**

145 **Table 2**). We titrated 58 oligo-conjugated antibodies to maximize signal-to-noise (**Methods**).
146 After disaggregating synovial tissue samples, we sorted viable cells for sequencing. A total of
147 314,030 cells (~3,800 per sample) passed stringent RNA QC, protein QC, and doublet
148 detection. We also excluded cells with inconsistent cell-type identities based on protein and
149 mRNA (**Supplementary Figure 1A-G, Methods**). The proportion of cells within 15 lineage
150 gates in CITE-seq and in flow cytometry correlated across samples (median Pearson $r=0.88$,
151 **Supplementary Figure 1G-H, Supplementary Table 3**). We integrated surface marker and
152 RNA data using canonical correlation analysis (CCA), corrected batch effects with Harmony³⁶,
153 and defined six major cell types: T, B/plasma, natural killer (NK), myeloid, stromal, and
154 endothelial cells (**Figure 1E-F, Supplementary Figure 2A-G, Methods**).

155

156 **Clustering samples on major cell-type abundance to define CTAPs**

157 We quantified the frequency of the six major cell types in each synovial tissue sample (**Figure**
158 **1G**). We used these six major cell types instead of finer-grained cell states to create a broad
159 categorization scheme that generalizes easily to many technologies (e.g. flow cytometry) for
160 wide clinical use. We then used hierarchical clustering to classify the spectrum of patient
161 samples into six different synovial cell-type abundance phenotypes (CTAPs). We arrived at six
162 groups because they demonstrated robust in-group similarity with bootstrapping and revealed
163 biological heterogeneity (**Figure 1G-H, Supplementary Figure 2H**, Jaccard index=0.727). We
164 named CTAPs based on dominant cell type(s): 1) endothelial, fibroblast, and myeloid cells
165 (EFM), 2) fibroblasts (F), 3) T cells and fibroblasts (TF), 4) T and B cells (TB), 5) T and myeloid
166 cells (TM), and 6) myeloid cells (M) (**Figure 1I, Supplementary Table 4, Methods**). CTAPs
167 reflect a spectrum of cell-type abundances apparent in principal component analysis (PCA) of
168 cell-type frequencies (**Figure 1J**).

169

170 **Characterizing a comprehensive RA synovial cell state atlas**

171 We defined finer-grained cell states and quantified sample abundances within cell types (**Figure**
172 **2**). Surface proteins were informative for cell-state delineation in T and B cells (**Supplementary**
173 **Figure 3A-C**), so we clustered cells on CCA canonical variates (CVs) capturing both RNA and
174 protein data (**Supplementary Figure 3D-F, Supplementary Figure 4, Methods**). For other cell
175 types, proteins were less informative, so we defined clusters from mRNA alone. In total we
176 defined 77 cell states: 24 T cell clusters (n=94,056 cells), 9 B/plasma cell clusters (n=30,697),
177 14 NK clusters (n=8,497), 15 myeloid clusters (n=76,181), 5 endothelial clusters (n=25,044),
178 and 10 stromal clusters (n=79,555) (**Figure 2A**). Using Symphony³⁷, we mapped cell states
179 from our prior study of 5,000 synovial cells²¹ onto these fine clusters²¹; coarse cell states
180 previously identified as associated with RA versus OA were also associated in this data set
181 (**Supplementary Figure 5, Supplementary Table 5**).

182
183 The 24 T cell clusters spanned innate-like states and CD4+ and CD8+ adaptive lineages
184 (**Figure 2A, Supplementary Figure 6A-C**). These included states implicated in autoimmunity,
185 such as regulatory CD4+ T cells (T_{reg}; T-8 and T-9) and T_{PH} and T_{FH} cells (T-3, T-7)^{24,28–30,38–41}. T-
186 3 and T-7 both expressed B cell-helper factors CXCL13 and *IL21*. T-7 comprised exclusively
187 T_{PH} cells and expressed more *ICOS*, *IFNG*, and *GZMA*, while T-3 contained T_{PH} and T_{FH} cells
188 expressing the lymphoid homing marker *CCR7* (**Supplementary Figure 6A-D**). T_{PH} cells are
189 known to be expanded in RA compared to OA^{21,24}. CD8+ subsets expressed different
190 combinations of *GZMB* and *GZMK* (T-13, T-14, T-15), reflecting differential cytotoxic potential.
191 With surface protein data we resolved T cell clusters that were not observed in our earlier
192 study²¹. This included *GNLY*+CD4+ (T-12), two double-negative (CD4-CD8-) gamma-delta T cell
193 clusters expressing *TRDC* (T-22 and T-23), and a cluster containing double-negative and CD8+
194 T cells expressing *ZBTB16* (PLZF) that resemble innate-like T cells such as natural killer T cells
195 and mucosal-associated innate T (MAIT) cells (T-21).

196

197 We found distinct separation between CD20⁺ (*MS4A1*⁺) B cells and CD138⁺ (*SDC1*⁺) plasma
198 cells (**Figure 2B, Supplementary Figure 7A-D**). CD20⁺ B cells comprised six clusters,
199 including *IGHM*⁺*IGHD*⁺*TCL1a*⁺ naive (B-2) and two CD27⁺ memory B cell clusters:
200 CD24⁺CD27⁺CD11b⁺ switched memory B cells (B-0) and CD24⁺⁺CD27⁺*IGHM*⁺ unswitched
201 memory B cells (B-1). We also identified CD11C⁺CXCR5^{low} ABCs (B-5)⁴²⁻⁴⁴, previously noted to
202 be associated with RA relative to OA²¹. B-5 expresses LAMP1, a lysosomal-associated
203 membrane protein that may play a role in B cell antigen-presentation⁴⁵. Additional B-5 genes
204 suggest ABC antigen-presentation⁴⁶ including *HLA-DR* and *CIITA*⁴⁷. We unexpectedly observed
205 CD1c⁺ B cells (B-3) with *CD27* and *IGHD* expression consistent with recirculating extrasplenic
206 marginal zone (MZ) B cells⁴⁸⁻⁵¹. CD1c⁺ MZ-like B cells (B-3) and other non-plasma B cells were
207 high producers of *IL6* and *TNF* (**Supplementary Figure 7D**). We identified *AICDA*⁺*BCL6*⁺ GC-
208 like B cells (B-4) consistent with ectopic germinal center (GC) formation in the synovium^{52,53}.
209 Plasma cells were surprisingly diverse and included *HLA-DRA*⁺*MKI67*⁺ plasmablasts (B-7),
210 *IGHM*⁺ plasma cells (B-6), and more mature *IGHG1*⁺*IGHG3*⁺ plasma cells (B-8). Plasma cell
211 heterogeneity may reflect both *in situ* generation and circulation from the periphery.
212
213 We also captured innate lymphocytes, including CD56^{br}CD16⁻ NK (8 clusters), CD56^{dim}CD16⁺
214 NK (4 clusters), and CD56^{dim}CD16⁻IL7R⁺ innate lymphoid cells (ILCs, 2 clusters) (**Figure 2C,**
215 **Supplementary Figure 8A-C**). CD56^{br}CD16⁻ NK cells were more abundant (mean 47.6% per
216 donor) than CD56^{dim}CD16⁺ NK cells (35.7%) and ILCs (12.9%), consistent with previous
217 observations in gut and lymph nodes⁵⁴. CD56^{br}CD16⁻ NK clusters were the only innate
218 lymphocytes expressing *GZMK*, and they variably expressed other genes encoding cytotoxic
219 molecules such as *GZMB* and *GNLV*. CD56^{dim}CD16⁺ NK cells had universally high expression
220 of *GZMB*, *GNLV*, and *PRF1*. *IFNG* was expressed highly in two CD56^{dim}CD16⁺ clusters (NK-1,
221 and NK-2) but was also expressed in NK-5 and NK-10. Some activating and inhibitory NK cell
222 receptors were differentially expressed, including *KLRK1* (NKG2D), predominantly expressed by

223 CD56^{br}CD16⁻ cells, and *KLRF1* (NKp80) and *FCRL6*, predominantly expressed by
224 CD56^{dim}CD16⁺ cells (**Supplementary Figure 8D**). We identified ILCs based on absence of
225 CD56 and CD16 and high expression of CD127 (IL-7Ra) protein⁵⁵. The larger ILC cluster
226 resembled group 3 ILCs (*RORC*⁺ NK-12), the functional analog of T_H17 T cells^{55,56}. The smaller
227 CD161⁺ population resembled group 2 ILCs (*GATA3*⁺ NK-12)⁵⁵⁻⁵⁷, analogous to T_H2. We did not
228 see a discrete cluster of T_H1-analogous group 1 ILCs, which may have co-clustered with NK
229 cells.

230

231 We identified 15 myeloid clusters spanning tissue macrophages, infiltrating monocytes,
232 conventional and plasmacytoid dendritic cells (**Figure 2D**). CD68 and CCR2 protein expression
233 discriminate tissue macrophages from infiltrating monocytes (**Supplementary Figure 9A-C**).
234 Three tissue macrophage clusters (M-0, M-1, M-2) in RA synovium were also found at high
235 frequencies in OA synovium and display a phagocytic phenotype with high expression of CD206
236 (*FOLR2*), CD163, *MERTK* and *MARCO* (**Supplementary Figure 9B,D**), suggesting
237 homeostatic debris-clearing function^{58,59}. *LYVE1* expression on tissue macrophages (M-0) may
238 indicate a perivascular function^{25,60}. Infiltrating monocytes included a sizable
239 *IL1B*⁺*FCN1*⁺*HBEGF*⁺ pro-inflammatory subset (M-7), likely derived from classical CD14^{high}
240 monocytes, which we previously described^{21,25}. A *STAT1*⁺*CXCL10*⁺ subset (M-6) likely derives
241 from non-classical CD14^{low}CD16^{high} monocytes and expresses interferon-response gene
242 signatures; these cells are enriched in the inflamed lung from COVID-19 pneumonia, colon from
243 Crohn's disease, and tumors^{23,61,62}. *MERTK*⁺*HBEGF*⁺ (M-3) and *SPP1*⁺ (M-4) bridged infiltrating
244 monocytes and tissue macrophages; both expressed high levels of *SPP1*, a marker of bone-
245 marrow-derived macrophages^{63,64} suggesting a transition from an inflammatory monocyte to a
246 more phagocytic phenotype of tissue macrophages. We identified four DC populations
247 corresponding to subsets described by Villani *et al*⁶⁵. Reflecting their respective antigen
248 presentation capacities, DC1 (M-12) expressing *CLEC9A* and *THBD* (CD141) cross-present

249 extracellular antigens to CD8 T cells, while DC2 and DC3 (M-10, 9) are *CLEC10A*^{high} cells that
250 activate and polarize CD4 T cells⁶⁵ (**Supplementary Figure 9D**). DC4 (M-11) expresses genes
251 found in CD14⁺ monocytes such as *IL1B* while also displaying a strong IFN signature. Lastly, we
252 identified a fifth DC subset (M-14) with high expression of endosomal marker *LAMP3*⁶⁶.

253

254 In the stroma, fibroblasts were divided broadly into lining (*PRG4*^{high}) and sublining (*THY1*⁺
255 *PRG4*^{low}) (**Figure 2E, Supplementary Figure 10A-F**). As previously described, lining fibroblasts
256 (F-0, F-1) were relatively depleted in RA and enriched in OA synovium, while sublining
257 fibroblasts separated into *HLA-DRA*⁺, *CD34*⁺, and *DKK3*⁺ groups^{21,67,68} (**Supplementary Table**
258 **6**). Lining fibroblasts subdivided into *PRG4*⁺*CLIC5*⁺ (F-0), *PRG4*⁺ (F-1), and an *RSPO3*⁺
259 population (F-8) with an intermediate lining/sublining phenotype. The *CD34*⁺ sublining fibroblast
260 cluster (F-2) highly expressed *PI16* and *DPP4* (CD26), suggesting they may be fibroblast
261 progenitors⁶⁹. *CXCL12*⁺ fibroblasts included an inflammatory *CD74*^{high}*HLA*^{high} cluster (F-5) with
262 high HLA expression, and a *CXCL12*⁺*SFRP1*⁺ cluster (F-6) with the highest levels of *IL6*, a
263 proven drug target in RA⁷⁰⁻⁷². The inflammatory signature in F-5 and F-6 suggest an
264 inflammatory phenotype driven by cytokine activation by infiltrating immune cells⁷³. The stromal
265 compartment also included a small cluster of *NOTCH3*⁺*MCAM* (CD146)⁺ mural cells (Mu-0).

266

267 Endothelial cells separated into *NOTCH4*⁺ arteriolar (E-3), *SPARC*⁺ capillary (E-0), *CLU*⁺
268 venular (E-1, E-2), and *LYVE1*⁺*PROX1*⁺ lymphatic endothelial cells (LEC, E-4) (**Figure 2F,**
269 **Supplementary Figure 10G-K**). The majority (53%) were venular and further subdivided into
270 *LIFR*⁺ (E-1) and *ICAM1*⁺ (E-2); these cells had high expression of inflammatory genes such as
271 *IL6* and *HLA*, along with genes that facilitate the transmigration of leukocytes into tissue such as
272 *ICAM1* and *SELE* (E-selectin)(**Supplementary Figure 10I**)⁷⁴. Arteriolar cells expressed high
273 levels of *CXCL12*, *LTBP4*, *NOTCH4*, and NOTCH ligand *DLL4*. *SPARC*⁺ capillary cells

274 expressed collagen and extracellular matrix genes. LECs represented a small number of cells
275 (n=324) with high expression of *CCL21* and *FLT4*^{75,76}.

276

277 For each sample, we calculated the proportion of each cell cluster within each cell type. Then,
278 we calculated the average of these cluster proportions within each RA CTAP and OA (**Figure**
279 **2**). These values are independent of cell-type abundance differences since they are calculated
280 relative to each cell type. For example, these values may reflect the relative abundance of *IL1B*⁺
281 macrophages among all myeloid cells, regardless of the total number of myeloid cells in a
282 sample. We observed reported differences in RA compared to OA, including an expansion of
283 sublining fibroblasts relative to lining fibroblasts, and expansion of *IL1B*⁺ macrophages relative
284 to *MERTK*⁺ macrophages.

285

286 **CTAPs are characterized by specific cell states**

287 We next set out to quantify how the composition of fine-grained cell states differed between
288 CTAPs. To accurately identify cell-states associated with individual CTAPs within each given
289 cell type, we used co-varying neighborhood analysis (CNA)⁷⁷. CNA tests highly granular
290 “neighborhoods”—small groups of phenotypically similar cells—rather than larger clusters and
291 accounts for age, sex, and cell count per sample. CNA associations suggest that certain single-
292 cell-resolution states within each cell type are more likely to be found in samples from one
293 CTAP than others. After identifying CTAP-associated neighborhoods, we defined the canonical
294 cell states that contain those neighborhoods to infer biologic meaning. In these analyses, we
295 use “expanded” and “depleted” to refer to changes in relative abundance within a cell type, but
296 notably these changes may not reflect a change in absolute cell numbers relative to total
297 number of cells.

298

299 We observed skewed T cell neighborhoods in CTAP-TB (permutation $p=0.046$) (**Methods,**
300 **Figure 3A-B, Supplementary Figure 6E, Supplementary Table 6**). T cell neighborhoods
301 among $CD4^+ T_{FH}/T_{PH}$ (T-3) and $CD4^+ T_{PH}$ (T-7) cells were expanded, while T cell neighborhoods
302 among cytotoxic $CD4^+ GNLY^+$ (T-12) and $CD8^+ GZMB^+$ cells (T-15) were depleted. Recognizing
303 that T_{FH} and T_{PH} cells differentiate B cells towards antibody production^{24,78}, we tested B cells for
304 association to CTAP-TB (permutation $p=0.03$). We observed expanded neighborhoods in
305 memory B (B-0 and B-1) and ABC (B-5) clusters, while $IgG1^+ IgG3^+$ and IgM^+ plasma cells (B-8,
306 B-6) were relatively depleted (**Figure 3C-D, Supplementary Figure 7E, Supplementary Table**
307 **6**). We note that though plasma cells are depleted among B/plasma cells in CTAP-TB, B and
308 plasma cells overall are enriched among total cells in CTAP-TB (23% compared to 1-10% in
309 other CTAPs) (**Figure 1I, Supplementary Table 4**). While T_{PH} (T-7), T_{FH}/T_{PH} (T-3), and ABC (B-
310 5) cells are enriched in CTAP-TB, they are present in all six CTAPs (**Supplementary Figures**
311 **6E and 7G, Supplementary Table 6**). In contrast, GC cells (B-4) were almost exclusively found
312 in CTAP-TB (**Supplementary Figure 7G**). Consistent with a role for T_{FH}/T_{PH} and IL21 in ABC
313 generation⁴³ and plasma cell differentiation, the frequency of ABCs (B-5) amongst B/plasma
314 cells correlated with the proportion of T_{PH} (T-7) (Pearson $r=0.50$, $p=3.7e-6$, **Figure 3E**) and
315 T_{FH}/T_{PH} (T-3) amongst T cells (Pearson $r=0.24$, $p=0.034$, **Supplementary Figure 7F**).
316
317 T cell neighborhoods enriched in CTAP-TF (permutation $p=0.036$) mainly consisted of cytotoxic
318 $CD4^+ GNLY^+$ (T-12) and $CD8^+ GZMB^+$ cells (T-15) (**Figure 3A, Supplementary Figure 6E,**
319 **Supplementary Table 6**). Similarly, NK cell neighborhoods were altered in CTAP-TF
320 (permutation $p=1e-4$), and these neighborhoods contained $GZMB$ -expressing $CD56^{dim}CD16^+$
321 NK cells (NK-0-3) (**Figure 3G-H, Supplementary Figure 8E**). The $GZMB^+$ (NK-0-3) proportion
322 of NK cells correlated with the $GZMB^+$ (T-15) proportion of T cells (Pearson $r=0.63$, $p=4.87 \times 10^{-$
323 10 , **Figure 3F**). This suggests that a subset of RA samples is enriched in $GZMB^+$ NK and T cells
324 expressing high *IFNG* (**Supplementary Figure 6D, Supplementary Figure 7D**). Conversely,

325 we observed that CD8⁺ T cells expressing *GZMK* (T-13/14) correlated with NK cells expressing
326 *GZMK* (NK-4-8, Pearson $r=0.51$, $p=1.41 \times 10^{-6}$, **Figure 3F**), suggesting that *GZMK*-expressing
327 CD8 T and NK cells share a transcriptional program that may result from their tissue
328 environments.

329
330 CTAP-TF also exhibited specific expansions of fibroblast subpopulations (permutation $p=0.048$,
331 **Figure 4A-B**). Specifically, *CXCL12*⁺*SFRP1*⁺ sublining fibroblasts (F-6) were disproportionately
332 expanded in CTAP-TF. These *CXCL12*⁺*SFRP1*⁺ sublining fibroblasts highly expressed *IL6* but
333 did not express HLA-DR genes.

334
335 Myeloid populations were different in CTAP-M compared to other CTAPs (permutation $p=1e-3$).
336 Cell neighborhoods within *SPP1*⁺ (M-4) and *MERTK*⁺*HBEGF*⁺ (M-3) bone marrow-derived
337 macrophages were enriched in CTAP-M suggesting recruitment of inflammatory monocytes and
338 transition to macrophage function (**Figure 4C-D**). Furthermore, in CTAP-M, *CD74*^{high}*HLA*^{high}
339 sublining fibroblast neighborhoods (F-5) were expanded relative to stromal cells (permutation
340 $p=1e-3$) and *SPARC*⁺ capillary cells (E-0) were expanded relative to endothelial cells
341 (permutation $p=7e-3$, **Figure 4A-B, E-F**). Interestingly, the neighborhoods expanded in CTAP-M
342 were depleted in CTAP-F, while neighborhoods depleted in CTAP-M were enriched in CTAP-F.
343 Specifically, subpopulations like lining (F-0 and F-1) and CD34⁺ sublining (F-2) fibroblasts
344 (permutation $p=3e-3$), *MERTK*⁺*LYVE1*⁺ (M-0) and *MERTK*⁺*S100A8*⁺ (M-2) macrophages
345 (permutation $p=1e-3$), and *LIFR*⁺ venular (E-1) and *ICAM1*⁺ venular (E-2) endothelial cells were
346 expanded in CTAP-F (permutation $p=3e-3$) and depleted in CTAP-M. Notably, the pro-
347 inflammatory *IL1B*⁺ macrophages²¹ (M-7), known to be expanded in RA patients in general²¹,
348 were lower in frequency in CTAP-EFM relative to other CTAPs (**Figure 4C**).

349

350

351 **Cell states and CTAPs associated with histology and clinical metrics**

352 In addition to association with CTAPs (**Figure 5A**), cell neighborhoods may also be associated
353 with histologic features of RA synovium, which are useful in clinical practice and reflect disease
354 pathogenesis^{79–81}. Using CNA, we identified transcriptional neighborhoods associated with
355 histology, accounting for age and sex (**Methods**). We scored samples for Krenn histologic
356 inflammation and lining layer domains, in addition to discrete histologic cell density and
357 aggregate abundance, reflecting inflammatory cell infiltration and organization respectively
358 (**Supplementary Figure 11A**). T cells were associated with aggregate scores (permutation
359 $p=0.0088$), driven by expanded T cell neighborhoods in $CD4^+ T_{FH}/T_{PH}$ (T-3), consistent with their
360 role in organizing secondary lymphoid structures^{82,83} (**Supplementary Figure 11B, Figure 5A**).
361 IgM+ plasma cells (B-6), plasmablasts (B-7), and ABCs (B-5) were also positively associated
362 with aggregates (permutation $p=0.007$) (**Supplementary Figure 11B, Figure 5A**). In similar
363 analysis of NK cell neighborhoods, $CD56^{br}CD16^{-}GZMA^+CD160^+$ cells (NK-4) were positively
364 associated with density and aggregate scores (permutation $p=3e-04$ and $1e-04$, respectively)
365 (**Supplementary Figure 11B**); this population also contained cell neighborhoods relatively
366 enriched in CTAP-TB (**Figure 2**), although the functional role of these cells in follicle-rich
367 synovium is less clear. Inflammatory myeloid neighborhoods within $STAT1^+CXCL10^+$ (M-6),
368 $SPP1^+$ (M-4) and inflammatory DC3 (M-9) (**Supplementary Figure 11B, Figure 5A**) were
369 associated with density (permutation $p=0.005$).

370

371 We wanted to understand if histologic and clinical measures are explained by CTAPs, taking
372 age, sex, cell count, and clinical collection site into account (**Methods**). CTAPs account for 18%
373 variance of histologic density ($p=0.0035$) and 18% of variance for aggregates ($p=0.0059$), with
374 CTAP-TB and CTAP-TF having the highest scores for both (**Figure 5B, Supplementary Figure**
375 **12A**). Consistent with these observations, CTAPs are associated with Krenn inflammation
376 scores ($p=4e-04$), but not with Krenn lining scores ($p=0.11$) (**Figure 5B, Supplementary Figure**

377 **12B)**. CTAP-F, CTAP-EFM, and CTAP-M have the lowest scores for all histological parameters
378 **(Figure 5B)**.

379 The presence of CCP autoantibodies and rheumatoid factor subcategorize RA patients as
380 seropositive or seronegative. Patients with positive CCP have more severe disease and
381 radiographic progression^{84,85}. CCP titer values differed across CTAPs ($p=0.023$, 18% variance),
382 with CTAP-TB having the highest CCP (mean=292) **(Figure 5B)**, even after restricting the
383 analysis to seropositive patients ($p=0.0047$) **(Supplementary Figure 12C)**.

384

385 Intriguingly, CTAPs were independent of most clinical variables including disease activity,
386 clinical inflammatory markers, smoking history, total swollen joint counts, and sex **(Figure 5C,**
387 **Supplementary Table 10, Supplementary Figure 12D-L)**. CTAPs were also mostly
388 independent of anatomic category and clinical sites **(Supplementary Figure 12H-I)**. Patients in
389 CTAP-EFM tended to be older and have longer-standing RA than patients in other CTAPs and
390 were mostly TNFi-inadequate responders **(Supplementary Figure 12J-K)**, although these
391 associations were not statistically significant.

392

393 **CTAPs feature disease-relevant cytokine profiles**

394 We recognized that cell states differentially expressed specific effector molecules, such as
395 cytokines and their receptors **(Supplementary Figure 13)**. Most cytokines and chemokines are
396 produced predominantly by one cell type **(Figure 6A)**. For key cytokines produced by multiple
397 cell types, we quantified the relative contributions of each cell type. For example, we found that
398 roughly equal numbers of T cells and myeloid cells express *TNF* while stromal, endothelial, and
399 B cells dominate among *IL-6*-expressing cells **(Figure 6B)**.

400

401 Next, we linked these key effector molecules to CTAPs to complement the previous analyses
402 where we identified clusters overlapping with associated cell neighborhoods. To do this, we
403 correlated CTAP neighborhood association scores with expression of key cytokines and
404 receptors to identify soluble factors produced by CTAP-associated cell states.
405
406 CTAP-TB T cell neighborhood association scores correlated with expression of T_{FH}/T_{PH} -marker
407 *CXCL13* in T cells (**Figure 6C-D**), consistent with the observation that associated T cell
408 neighborhoods were in the T_{FH}/T_{PH} clusters (**Figure 3A**). In contrast, CTAP-TF-associated T cell
409 neighborhoods were associated with expression of *IFNG* and *TNF*, expressed by cytotoxic
410 (*GZMB*⁺ or *GNLY*⁺) $CD8^{+}$ and $CD4^{+}$ T cell populations (**Figure 3A, Figure 6D**). NK cell
411 populations enriched in CTAP-TF also expressed high *IFNG* and *TNF* (**Figure 3G, Figure 6D**).
412 These results suggest that *TNF* and *IFNG* may be intrinsic to the molecular environment of
413 CTAP-TF.
414
415 Analysis of myeloid cell neighborhoods in CTAP-EFM, CTAP-F, and CTAP-M also highlighted
416 key cytokines (**Figure 6D**). CTAP-M myeloid neighborhood association scores correlated with
417 expression of chemokines that related to activity of myeloid cells and neutrophils, *CXCL10* and
418 *CCL2* (**Figure 6D**), and angiogenic factors *CXCL16* and *VEGFA*. In CTAP-M, endothelial cell
419 neighborhood association scores correlated with *KDR* (VEGF receptor 2) (**Figure 6D**),
420 consistent with the prevalence of capillary cells in CTAP-M⁸⁶. In contrast, in CTAP-F, *LIFR*⁺ and
421 *ICAM1*⁺ venous endothelial cells expressed high levels of *CCL14*, whose cognate receptor
422 *CCR1* was highly expressed by *MERTK*⁺ macrophages, offering a potential mechanism for the
423 enrichment of this macrophage subset (**Figure 4C-D, Figure 6D**).
424
425 **CTAPs serve as a reference to map data from other patients and cohorts**

426 Our study included three patients with replicate biopsies obtained from the same joint (98, 105,
427 and 190 days) after the initial biopsy. We assessed the stability of CTAP phenotypes over time
428 between repeated and baseline samples. We found that the cell-type composition of repeat
429 biopsies was similar to the initial biopsy (mean Mahalanobis distance=1.55, permutation
430 $p=0.073$) (**Supplementary Figure 14A-B**), though a larger study would be necessary to
431 understand how dynamic CTAPs are in a given patient.

432

433 Given the potential benefits of categorizing synovial tissues from future RA studies into CTAPs,
434 we next examined whether samples can be classified into CTAPs using a lower-resolution
435 technology such as flow cytometry. We built a Mahalanobis-distance-based nearest-neighbor
436 classifier, and we were able to accurately replicate CITE-seq-based CTAP assignments based
437 on flow cytometry data (accuracy=87%, **Figure 6E, Supplementary Figure 14C-D, Methods**).
438 Since CTAPs appear to correlate with known drug targets (**Figure 6D**) and can be assigned
439 even with flow cytometry, we expect that CTAPs can be used to systematically query RA
440 heterogeneity across technologies to improve the granularity of clinical studies and trials and
441 potentially to guide therapy selection.

442

443

444 **Discussion**

445 We constructed a comprehensive synovial tissue inflammation reference of >314,000 single
446 cells. This clinically phenotyped RA atlas can be used to classify single-cell data from other RA
447 patients, identify shared pathways across diseases, and identify novel drug targets. We
448 observed that inflamed tissue samples from RA patients have diverse cellular composition that
449 is captured in six CTAPs.

450

451 CTAPs represent categories of RA characterized by the presence of certain cell states and the
452 absence of others. We observed that some previously identified pathogenic cell states in RA are
453 expanded in specific CTAPs. For example, CD4+ T_{FH} and T_{PH} cells, generally observed to be
454 enriched among T cells in RA compared to OA²⁴, are present in all CTAPs but are most
455 expanded in CTAP-TB. These T cell states are enriched along with ABCs and memory B cells,
456 consistent with the formation of B and T cell aggregates. Independent work has shown that B
457 cell activation pathways are active in the setting of human autoimmunity⁴⁴. Importantly, prior
458 work has focused on peripheral blood and normal secondary lymphoid tissue⁸⁷⁻⁸⁹, and it
459 remains unknown how this translates to B cell activation and ectopic lymphoid reactions in RA
460 synovium. Our work suggests the presence of two synovial B cell activation pathways, including
461 conventional germinal center responses and extra-follicular pathways, the latter characterized
462 by CXCR5- ABCs and T_{PH}. The rarity of GC dark-zone B cells and abundance of ABCs suggest
463 the prominence of extra-follicular activation pathways in RA synovium. Other novel findings from
464 the B-cell analyses include genes associated with antigen presentation in the ABCs, the
465 presence of CD1c- MZ-like B cells previously described in Sjogren's disease salivary glands⁹⁰,
466 and the heterogeneity of plasma cells.

467
468 In other cell types as well, CTAPs delineate RA subsets where established cell states of interest
469 may be more or less prominent. For example, among fibroblasts, prior studies have found that
470 inflammatory sublining cells expressing *HLA-DR*, *CXCL9*, *CXCL12*, and *IL6* are known to be
471 enriched in RA compared to OA^{21,68}. Here, we find that these inflammatory sublining cells are
472 composed of several subpopulations—some of which, specifically *CXCL12*⁺ and
473 *CD74*^{high}*HLA*^{high} cells, were enriched in CTAP-TF and CTAP-M, respectively. These findings
474 may reflect different axes of inflammatory fibroblast phenotypes, likely involving signals
475 exchanged with surrounding leukocytes. Interestingly, CTAP-M, where *CD74*^{high}*HLA*^{high}
476 fibroblasts are enriched, also exhibits specific enrichment of *MERTK*⁺*HBEGF*⁺ and *SPP1*⁺

477 (osteopontin) macrophages, and several other myeloid populations (e.g. *IL1B+* *FCN1+*) are also
478 prominent. These and other instances of co-enriched populations (e.g. *GZMK+* versus *GZMB+*
479 CD8 T cells and NK cell subsets) inspire new questions about cell-cell interactions underlying
480 inflammatory phenotypes in RA synovium and potentially other tissues and diseases.

481
482 CTAPs are associated with histologic and serologic (CCP) parameters but not with drug history
483 and clinical disease activity metrics in this study. We argue that CTAPs from biopsies offer
484 independent information from what physician assessments offer. Our study does not address
485 the evolution of CTAPs in patients over time. We anticipate future longitudinal studies to
486 investigate CTAP changes over time along with treatment effects. In our limited assessment of
487 three patients, we noted minimal evolution of CTAPs despite treatment changes.

488
489 Targeting the specific cell subsets enriched in a given CTAP may be key in personalized RA
490 treatment. For example, abrogating T-B cell communication with B cell-depleting antibodies
491 (e.g. rituximab) or blocking costimulation (e.g. abatacept) in CTAP-TB may break the
492 pathogenic mechanisms that drive inflammation in these patients^{18,82}. Conversely, patients with
493 CTAP-TF and CTAP-M feature fibroblast populations with high IL-6, an established target of
494 current FDA-approved treatments of RA (e.g. tocilizumab). CTAP-TF and CTAP-M feature
495 abundant *IFNG*-expressing cells or IFN-associated gene signatures, suggesting that these
496 patients may respond effectively to JAK inhibitors (e.g. tofacitinib, upadacitinib). Lastly, other
497 CTAPs, such as CTAP-EFM and CTAP-F, currently have no obvious targets of currently
498 available treatments and warrant further focused study. Thus, CTAPs represent valuable
499 molecular classifications of RA that may drive the search for new treatments.

500
501 The CTAP paradigm provides a tissue classification system that captures coarse cell-type and
502 fine cell-state heterogeneity. Importantly, CTAPs use global cell-type frequencies and are

503 thereby an accessible tool to categorize heterogeneity of tissue inflammation using multiple
504 technologies. The model presented here may serve as a powerful prototype to classify other
505 types of tissue inflammation, including in other immune-mediated diseases. A deeper
506 understanding of the heterogeneity of tissue inflammation in RA and other autoimmune
507 diseases may shed new light on disease pathogenesis and reveal new treatment targets.

508

509

510 **Accelerating Medicines Partnership Program: Rheumatoid Arthritis and Systemic Lupus**

511 **Erythematosus (AMP RA/SLE) Network includes:**

512 Jennifer Albrecht¹³, Jennifer L. Barnas¹³, Joan M. Bathon¹⁴, David L. Boyle¹⁶, S. Louis Bridges
513 Jr.^{11,26}, Debbie Campbell¹³, Hayley L. Carr¹⁷, Adam Chicoine¹, Andrew Cordle²⁹, Michelle
514 Curtis^{1,2,3,4,5}, Patrick Dunn^{30,31}, Lindsay Forbess¹⁷, Peter K. Gregersen²², Joel M. Guthridge³²,
515 Lionel B. Ivashkiv^{11,26}, Kazuyoshi Ishigaki^{1,2,3,4,5,33}, Judith A. James³², Gregory Keras¹, Ilya
516 Korsunsky^{1,2,3,4,5}, Amit Lakhanpal^{11,26}, James A. Lederer³⁴, Zhihan J. Li¹, Yuhong Li¹, Andrew
517 McDavid³⁵, Nida Meednu¹³, Ian Mantel^{11,26}, Mandy J. McGeachy²⁵, Karim Raza¹⁷, Yakir
518 Reshef^{1,2,3,4,5}, Christopher Ritchlin¹³, William H. Robinson³⁶, Saori Sakaue^{1,2,3,4,5}, Jennifer A.
519 Seifert¹⁵, Melanie H. Smith¹¹, Dagmar Scheel-Toellner¹⁷, Paul J. Utz³⁶, Michael H. Weisman^{18,36},
520 Zhu Zhu¹

521

522 ²⁹Department of Radiology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA.

523 ³⁰Division of Allergy, Immunology, and Transplantation, National Institute of Allergy and
524 Infectious Diseases, National Institutes of Health, Bethesda, MD, USA.

525 ³¹Northrop Grumman Health Solutions, Rockville, MD, USA.

526 ³²Department of Arthritis & Clinical Immunology, Oklahoma Medical Research Foundation;
527 Oklahoma City, OK, USA.

528 ³³Laboratory for Human Immunogenetics, RIKEN Center for Integrative Medical Sciences,
529 Yokohama, Japan.

530 ³⁴Department of Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston,
531 MA, USA.

532 ³⁵Department of Biostatistics and Computational Biology, University of Rochester School of
533 Medicine and Dentistry; Rochester, NY, USA.

534 ³⁶Division of Immunology and Rheumatology, Institute for Immunity, Transplantation and
535 Infection, Stanford University School of Medicine, Stanford, CA, USA.

536

537

538

539

540

541

542

543

544

545

546

547

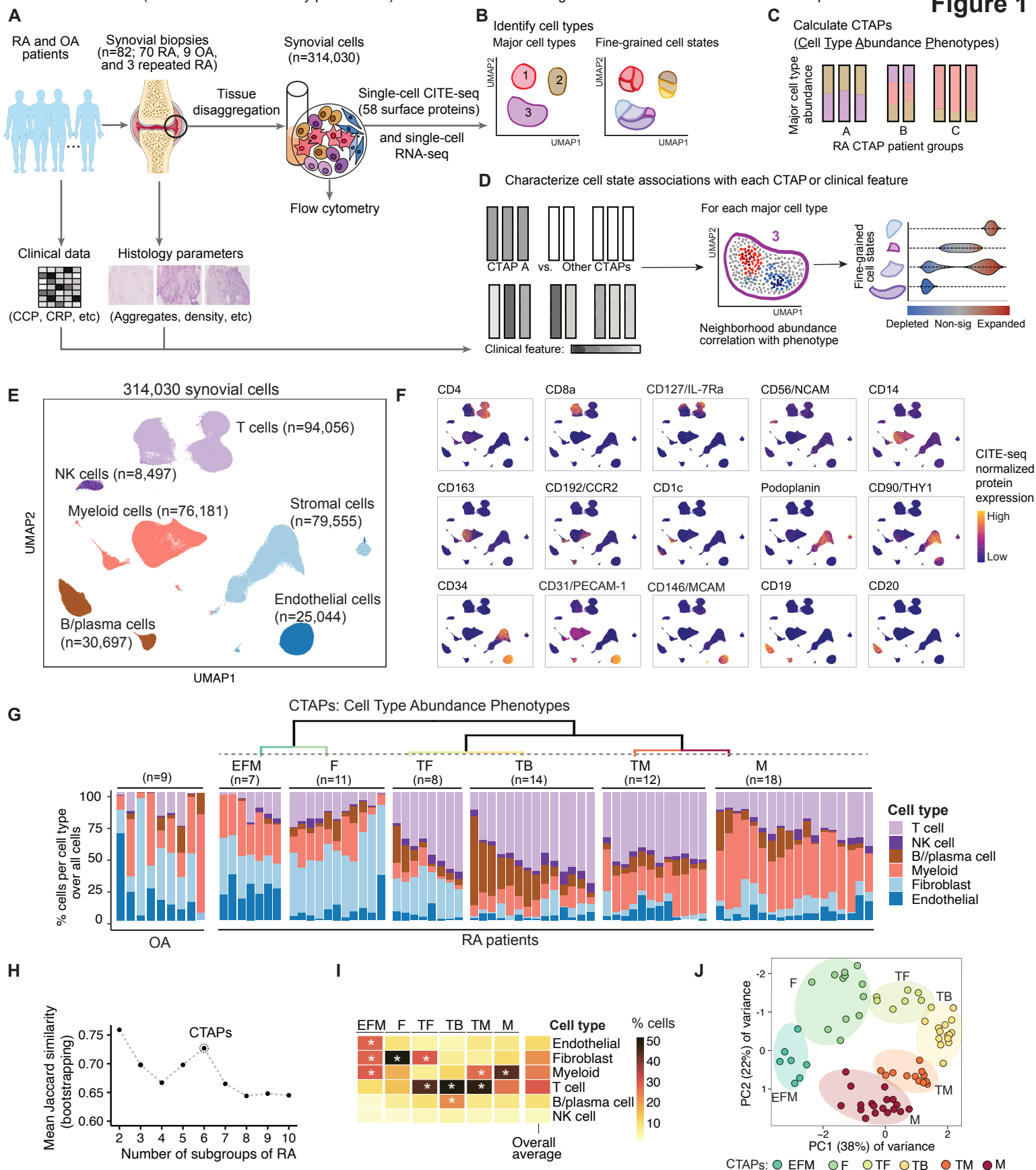


Figure 1. Overview of multimodal single-cell synovial tissue pipeline and cell type abundance analysis reveals distinct RA cell type abundance phenotypes (CTAPs). **A.** Description of patient recruitment, clinical and histologic metrics, synovial sample processing pipeline, and computational analysis strategy, including **B.** identifying major cell types and fine-grained cell states, **C.** definition of distinct RA CTAPs, and **D.** cell neighborhoods associations with each CTAP or with clinical or histologic parameters for each major cell type, **E.** Integrative UMAP based on mRNA and protein discriminated major cell types, **F.** UMAPs of CITE-seq antibody-based expression of cell type lineage protein markers. Cells are colored based on expression from blue (low) to yellow (high), **G.** Hierarchical clustering of cell type abundances captures six RA subgroups, referred to as cell type abundance phenotypes (CTAPs). The nine OA samples are shown as a comparison. Each bar represents one synovial sample, colored by the proportion of each major cell type, **H.** Mean Jaccard similarity coefficient to test CTAP stability by bootstrapping 10,000 times for each tested number of patient subgroups ranging from 2 to 10, **I.** Average proportions of each major cell type among samples in each CTAP. Overall average proportions across all samples are shown as a comparator. Asterisk represents the proportion that is greater than the overall average for that cell type, **J.** PCA of major cell type abundances. Each dot represents a sample colored by CTAPs.

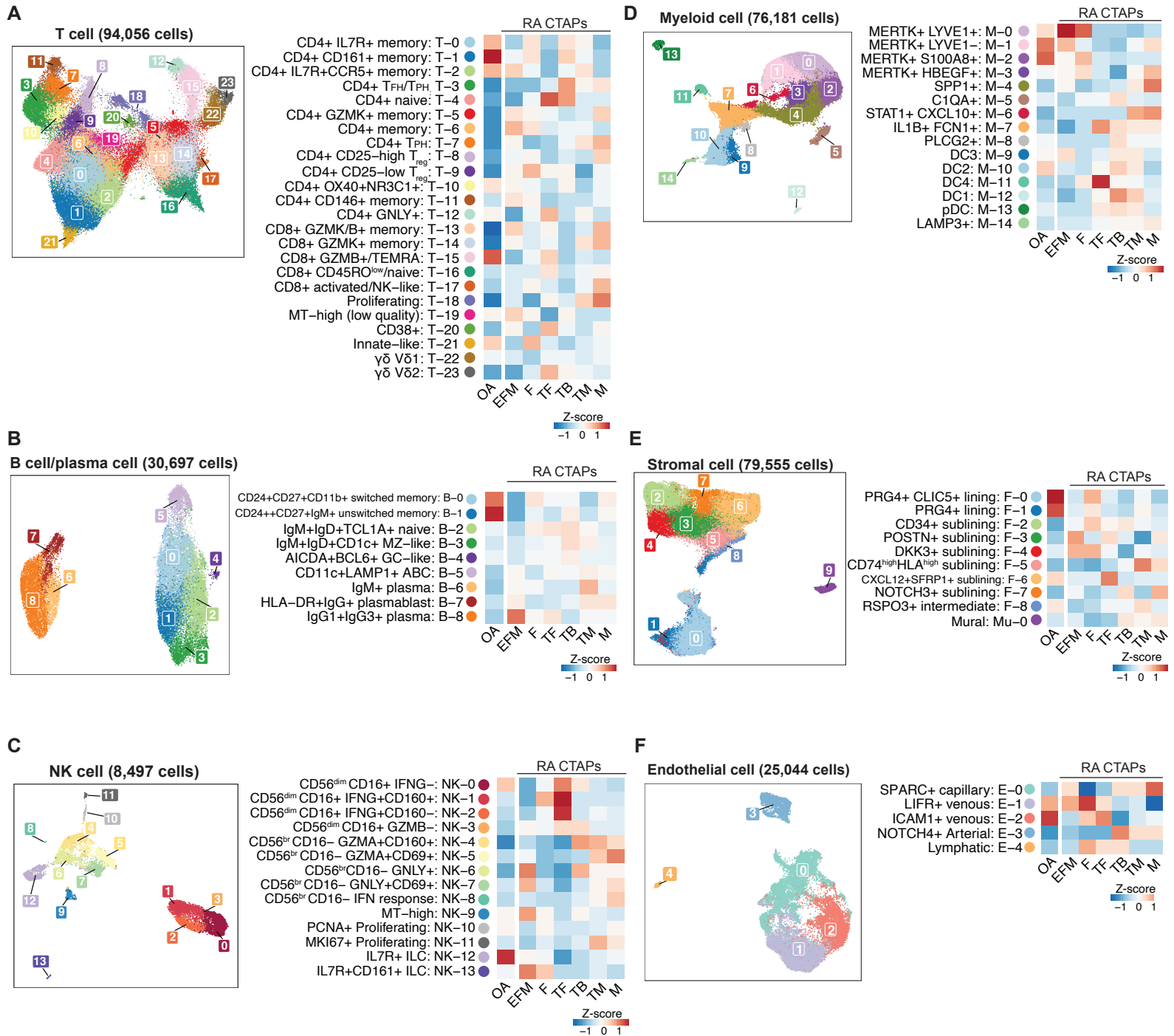


Figure 2. Cell-type-specific single-cell analysis captures 77 distinct cell states in RA synovium. A-F. Six cell-type-specific reference UMAPs colored by fine-grained cell states. For each cell type, the heatmap shows the average proportions of each cluster across patient samples in each RA CTAP and OA, scaled within each cluster.

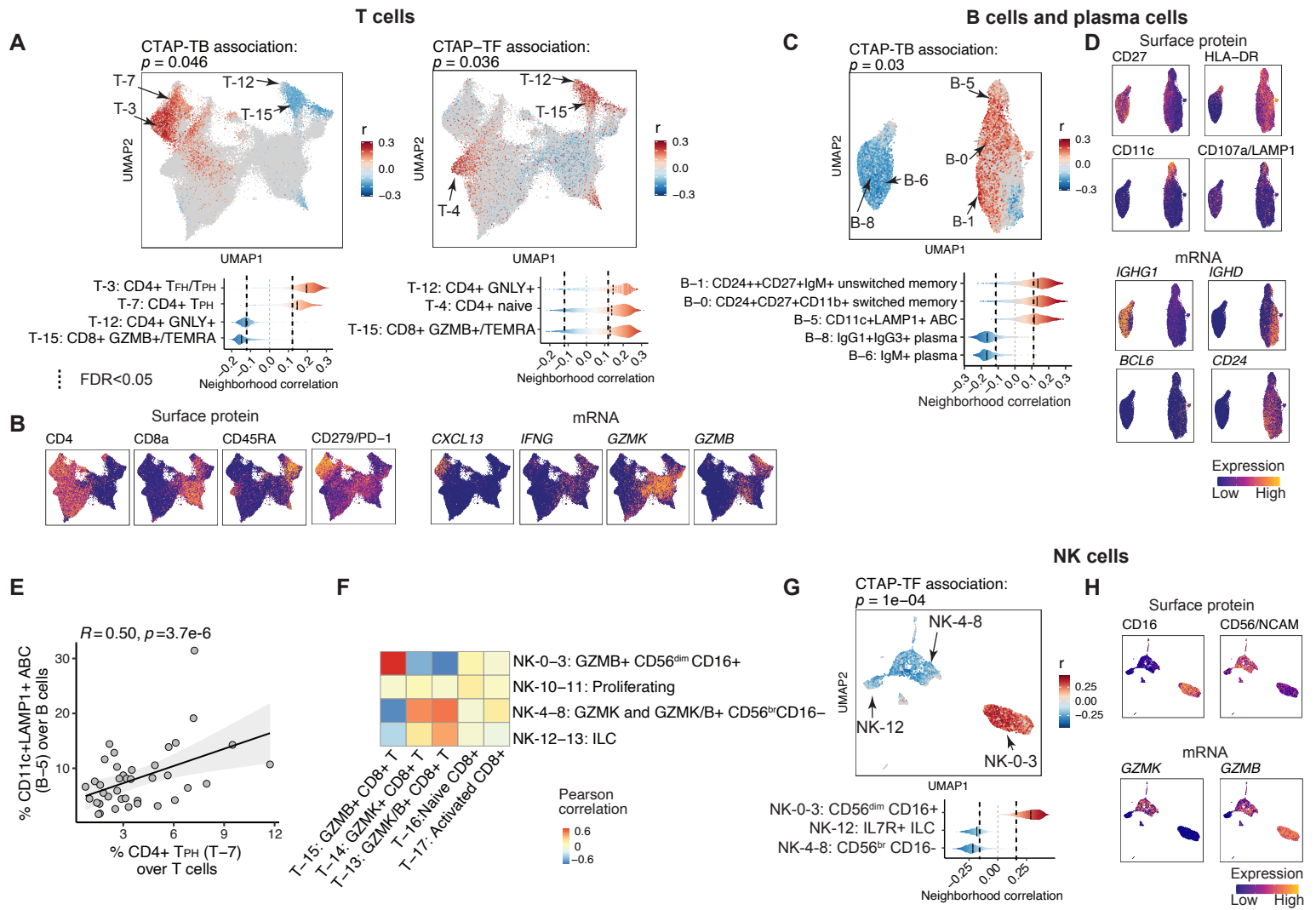


Figure 3. Different T cell, B cell, and NK cell populations are associated with RA CTAPs. **A**, Associations of T cell neighborhoods with CTAP-TB and CTAP-TF. P-values are from the CNA test for each CTAP within T cells. For all CNA results, cells in UMAP are colored in red (positive) or blue (negative) if their neighborhood is significantly associated with the CTAP (FDR < 0.05), and gray otherwise. Distributions of neighborhood correlations are shown for clusters with >50% of neighborhoods correlated with the CTAP at FDR > 0.05. **B**, Expression of selected surface proteins and transcripts among T cells. For all expression UMAPs, cells are colored from blue (low) to yellow (high). **C**, Associations of B/plasma cell neighborhoods with CTAP-TB. **D**, Expression of selected surface proteins and transcripts among B/plasma cells. **E**, Percentage of TPH (T-7) out of T cells and CD11c+ LAMP1+ ABCs (B-5) out of B/plasma cells for each donor sample, represented by points. R and p-value are calculated from Pearson correlation. **F**, Heatmap colored by Pearson correlation between per-donor CD8 T cell and NK cell cluster abundances. **G**, Associations of NK cell neighborhoods with CTAP-TF. **H**, Expression of selected surface proteins and transcripts in NK cells.

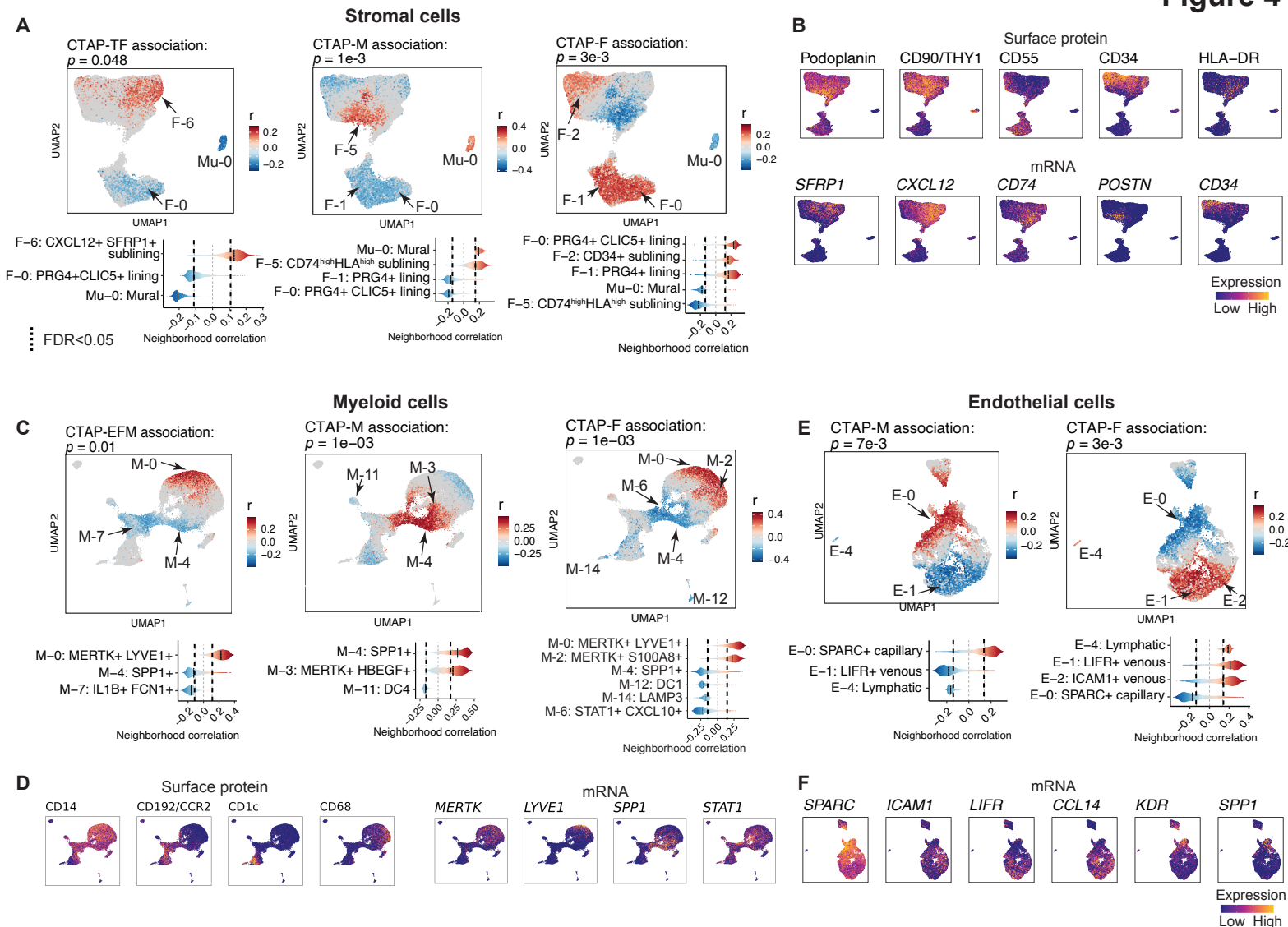
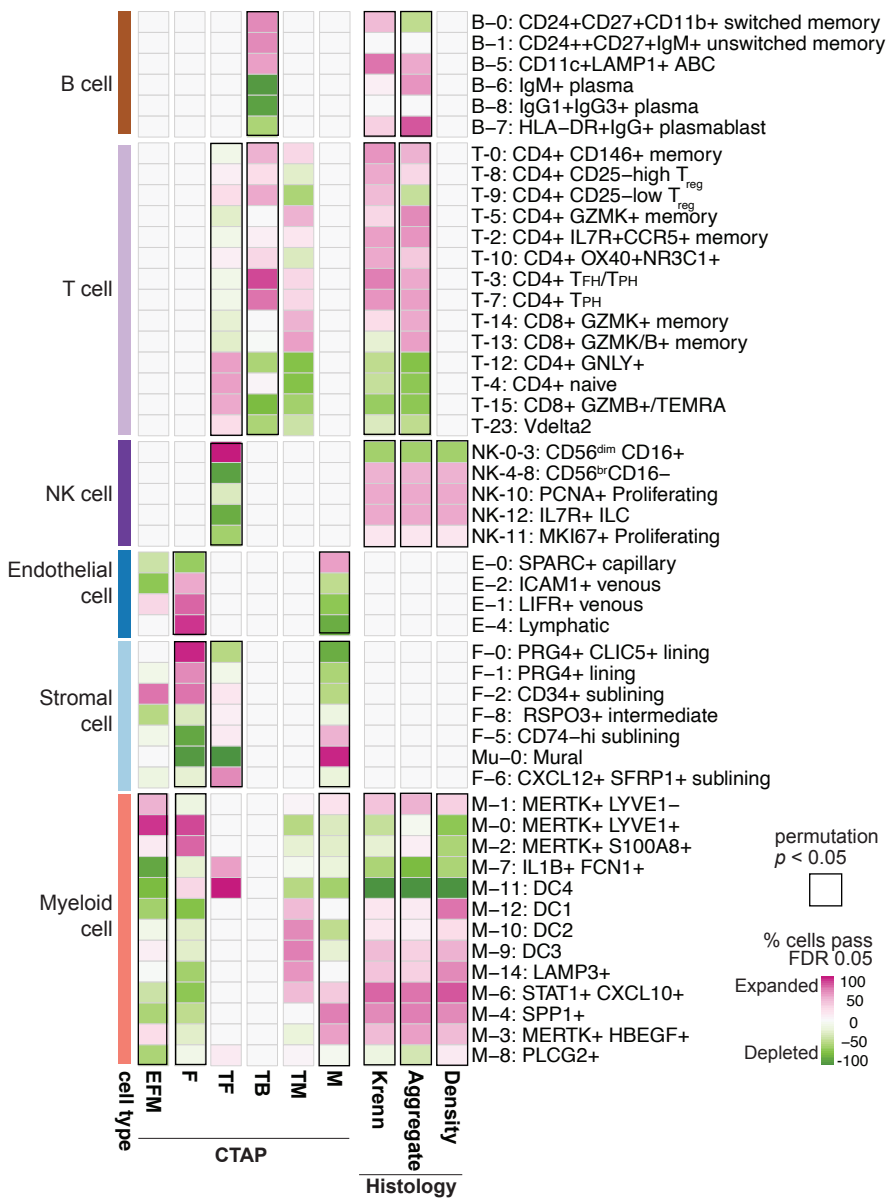
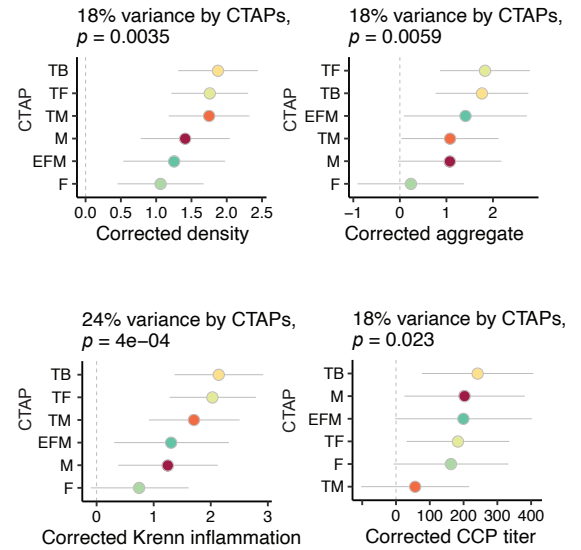


Figure 4. Different stromal, myeloid, endothelial cell populations are associated with RA CTAPs. **A.** Association of stromal cell neighborhoods with CTAP-TF, CTAP-M, and CTAP-F. For all CNA results, cells in UMAPs are colored in red (positive) or blue (negative) if their neighborhood is significantly associated with the CTAP (FDR < 0.05), and gray otherwise. Distributions of neighborhood correlations are shown for clusters with >50% of neighborhoods correlated with the CTAP at FDR > 0.05. **B.** Expression of selected surface proteins and transcripts among stromal cells. For all expression UMAPs, cells are colored from blue (low) to yellow (high). **C.** Association of myeloid cell neighborhoods with CTAP-EFM, CTAP-M, and CTAP-F. **D.** Expression of selected surface proteins and transcripts among myeloid cells. **E.** Association of endothelial cell neighborhoods with CTAP-M and CTAP-F. **F.** Expression of selected surface proteins and transcripts among endothelial cells.

A Covarying neighborhood analysis reveals specific cell states associated with each CTAP and histologic metrics



B



C

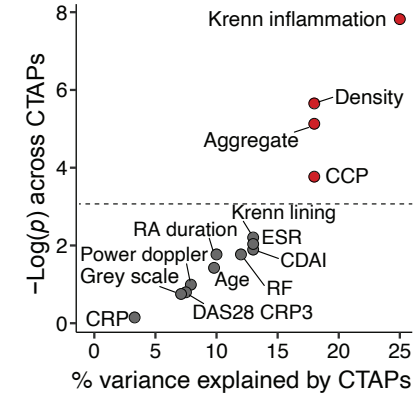


Figure 5. Single-cell covarying neighborhood analysis reveals significant association of cell states with disease indicators. **A.** Heatmap of CNA associations of specific cell states with each RA CTAP. Colors represent % cell neighborhoods from each cell state with local (neighborhood-level) phenotype correlations passing FDR < 0.05 significance from white to pink (expanded) or green (depleted). Cell types significantly associated globally (cell-type-level) with a phenotype at permutation $p < 0.05$ are boxed in black. **B.** Association between clinical features and CTAPs, adjusting covariates for age, sex, cell number, and clinical collection site. Percentage of variance explained by CTAPs alone and p-value are calculated with ANOVA tests. 95% confidence intervals are shown. **C.** Clinical, demographic, and histologic metrics plotted by percentage of variance explained by CTAPs and the ANOVA p-value for its association with CTAPs. Features in red are significant at ANOVA $p < 0.05$.

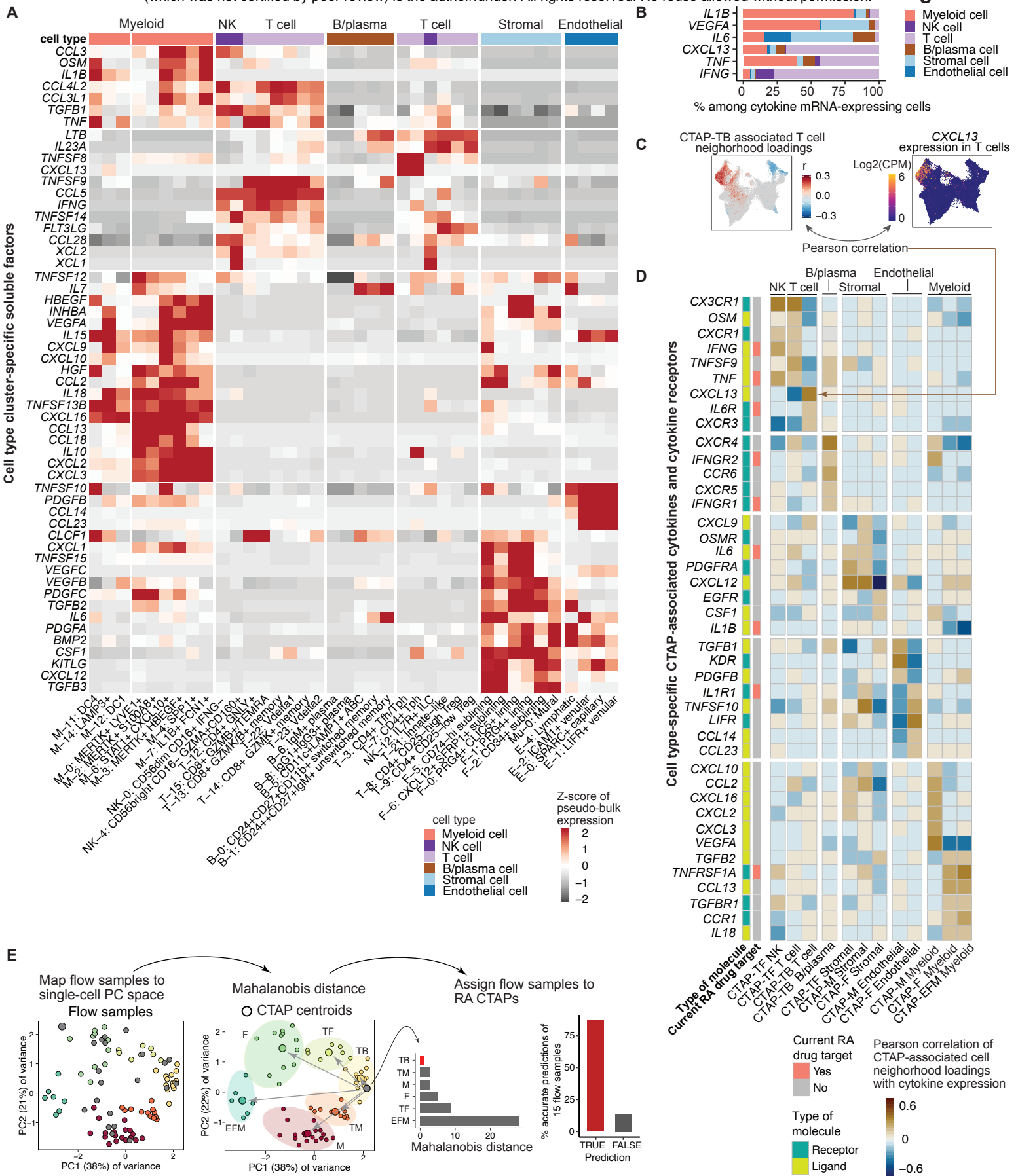


Figure 6. Cell type clusters and CTAPs feature distinct disease-relevant soluble factor and receptor profiles. **A.** Expression profiles of cell type cluster-specific soluble factors. **B.** Percent contribution among cytokine mRNA-expressing cells from each major cell type. **C.** Expression of representative cytokine, CXCL13, that is significantly correlated with CTAP-associated cell neighborhoods. Cells in UMAPs of CTAP associations are colored in red (positive) or blue (negative) if their neighborhood is significantly associated with the CTAP (FDR < 0.05), and gray otherwise. Cells in expression UMAPs are colored from blue (low) to yellow (high). **D.** With a heatmap, we visualized the cytokines and receptors whose expressions are significantly correlated ($r > 0.5$) with CTAP-associated cells; we then hierarchically clustered them based on cell type-specific CTAPs. For each gene, receptor/ligand designation and current RA drug target status are labeled. **E.** Pipeline and results to map and classify flow cytometry samples by single-cell RA CTAPs. Bar plot shows accuracy of flow sample classification (i.e., assigned to the same CTAP as a single-cell sample from the same patient).

548 **Figure Legends**

549 **Figure 1. Overview of multimodal single-cell synovial tissue pipeline and cell type**

550 **abundance analysis reveals distinct RA cell type abundance phenotypes (CTAPs). A.**

551 Description of patient recruitment, clinical and histologic metrics, synovial sample processing
552 pipeline, and computational analysis strategy, including **B.** identifying major cell types and fine-
553 grained cell states, **C.** definition of distinct RA CTAPs, and **D.** cell neighborhoods associations
554 with each CTAP or with clinical or histologic parameters for each major cell type, **E.** Integrative
555 UMAP based on mRNA and protein discriminated major cell types, **F.** UMAPs of CITE-seq
556 antibody-based expression of cell type lineage protein markers. Cells are colored based on
557 expression from blue (low) to yellow (high), **G.** Hierarchical clustering of cell type abundances
558 captures six RA subgroups, referred to as cell type abundance phenotypes (CTAPs). The nine
559 OA samples are shown as a comparison. Each bar represents one synovial sample, colored by
560 the proportion of each major cell type, **H.** Mean Jaccard similarity coefficient to test CTAP
561 stability by bootstrapping 10,000 times for each tested number of patient subgroups ranging
562 from 2 to 10, **I.** Average proportions of each major cell type among samples in each CTAP.
563 Overall average proportions across all the samples are shown as a comparator. Asterisk
564 represents the proportion that is greater than the overall average for that cell type, **J.** PCA of
565 major cell type abundances. Each dot represents a sample, plotted based on its PC1 and PC2
566 projections and colored by CTAPs.

567

568 **Figure 2. Cell-type-specific single-cell analysis captures 77 distinct cell states in RA**

569 **synovium. A-F.** Six cell-type-specific reference UMAPs colored by fine-grained cell state

570 clusters. For each cell type, the heatmap shows the average proportions of each cluster across
571 patient samples in each RA CTAP and OA, scaled within each cluster.

572

573 **Figure 3. Different T cell, B cell, and NK cell populations are associated with RA CTAPs.**
574 **A**, Associations of T cell neighborhoods with CTAP-TB and CTAP-TF. P-values are from the
575 CNA test for each CTAP within T cells. For all CNA results, cells in UMAP are colored in red
576 (positive) or blue (negative) if their neighborhood is significantly associated with the CTAP (FDR
577 < 0.05), and gray otherwise. Distributions of neighborhood correlations are shown for clusters
578 with $>50\%$ of neighborhoods correlated with the CTAP at $FDR>0.05$, **B**. Expression of selected
579 surface proteins and transcripts among T cells. For all expression UMAPs, cells are colored
580 from blue (low) to yellow (high), **C**. Associations of B/plasma cell neighborhoods with CTAP-TB,
581 **D**. Expression of selected surface proteins and transcripts among B/plasma cells, **E**. Percentage
582 of T_{PH} (T-7) out of T cells and CD11c⁺ LAMP1⁺ ABCs (B-5) out of B/plasma cells for each
583 donor sample, represented by points. R and p-value are calculated from Pearson correlation, **F**.
584 Heatmap colored by Pearson correlation between per-donor CD8 T cell and NK cell cluster
585 abundances, **G**. Associations of NK cell neighborhoods with CTAP-TF. **H**. Expression of
586 selected surface proteins and transcripts in NK cells.

587

588 **Figure 4. Different stromal, myeloid, endothelial cell populations are associated with RA**
589 **CTAPs. A**. Association of stromal cell neighborhoods with CTAP-TF, CTAP-M, and CTAP-F.
590 For all CNA results, cells in UMAPs are colored in red (positive) or blue (negative) if their
591 neighborhood is significantly associated with the CTAP (FDR < 0.05), and gray otherwise.
592 Distributions of neighborhood correlations are shown for clusters with $>50\%$ of neighborhoods
593 correlated with the CTAP at $FDR>0.05$, **B**. Expression of selected surface proteins and
594 transcripts among stromal cells. For all expression UMAPs, cells are colored from blue (low) to
595 yellow (high), **C**. Association of myeloid cell neighborhoods with CTAP-EFM, CTAP-M, and
596 CTAP-F, **D**. Expression of selected surface proteins and transcripts among myeloid cells, **E**.
597 Association of endothelial cell neighborhoods with CTAP-M and CTAP-F, **F**. Expression of
598 selected surface proteins and transcripts among endothelial cells.

599

600 **Figure 5. Single-cell covarying neighborhood analysis reveals significant association of**
601 **cell states with disease indicators. A.** Heatmap of CNA associations of specific cell states
602 with each RA CTAP. Colors represent % cell neighborhoods from each cell state with local
603 (neighborhood-level) phenotype correlations passing FDR < 0.05 significance from white to pink
604 (expanded) or green (depleted). Cell types significantly associated globally (cell-type-level) with
605 a phenotype at permutation $p < 0.05$ are boxed in black, **B.** Association between clinical
606 features and CTAPs, adjusting covariates for age, sex, cell number, and clinical collection site.
607 Percentage of variance explained by CTAPs alone and p-value are calculated with ANOVA
608 tests. 95% confidence intervals are shown. **C.** Clinical, demographic, and histologic metrics
609 plotted by percentage of variance explained by CTAPs and the ANOVA p-value for its
610 association with CTAPs. Features in red are significant at ANOVA $p < 0.05$.

611

612 **Figure 6. Cell type clusters and CTAPs feature distinct disease-relevant soluble factor**
613 **and receptor profiles. A.** Expression profiles of cell type cluster-specific soluble factors, **B.**
614 Percent contribution among cytokine mRNA-expressing cells from each major cell type, **C.**
615 Expression of representative cytokine, *CXCL13*, that is significantly correlated with CTAP-
616 associated cell neighborhoods. Cells in UMAPs of CTAP associations are colored in red
617 (positive) or blue (negative) if their neighborhood is significantly associated with the CTAP (FDR
618 < 0.05), and gray otherwise. Cells in expression UMAPs are colored from blue (low) to yellow
619 (high), **D.** With a heatmap, we visualized the cytokines and receptors whose expressions are
620 significantly correlated ($r > 0.5$) with CTAP-associated cells; we then hierarchically clustered
621 them based on cell type-specific CTAPs. For each gene, receptor/ligand designation and
622 current RA drug target status are labeled, **E.** Pipeline and results to map and classify flow
623 cytometry samples by single-cell RA CTAPs. Bar plot shows accuracy of flow sample
624 classification (i.e., assigned to the same CTAP as a single-cell sample from the same patient).

625 **Supplementary Figures**

626 **Supplementary Figure 1. Detailed single-cell CITE-seq quality control.** **A.** Quality of the
627 cells based on number of genes detected and percent mitochondrial UMIs (%MT), **B.**
628 Percentage of good quality cells for sample-level QC, **C.** Doublet detection using Scrublet, **D.**
629 UMAP of the number of genes and UMIs detected, **E.** Number of cells remaining after each step
630 of QC, **F.** Distributions of cell type lineage antibody staining from CITE-seq determine
631 percentage of major cell types based on the thresholds (red line) including % CD45⁺ cells, % T
632 cells based on CD4 antibody, % B cells based on CD20⁺, % macrophages based on CD14⁺, %
633 endothelial cells based on CD146⁺, and % fibroblasts based on PDPN⁺, **G.** Representative
634 gating of flow cytometry data to quantify selected synovial cell populations, **H.** Concordance of
635 single-cell CITE-seq antibody staining with an analogous gating schema for flow cytometry. For
636 flow gating, we determined % CD45⁺ based on CD45⁺ over all live cells, % T cells based on
637 CD45⁺CD3⁺ over all live cells, % B cells based on CD45⁺CD3⁻CD14⁻CD20⁺, % macrophages
638 based on CD45⁺CD14⁺, % fibroblasts based on CD45⁻CD146⁻CD31⁻, % endothelial cells based
639 on CD45⁻CD146⁺CD31⁺, % CD4 T cells based on CD45⁺CD3⁺CD4⁺, % HLA⁺ CD4 T cells based
640 on CD45⁺CD3⁺CD4⁺HLA-DR⁺, % CD8 T cells based on CD45⁺CD3⁺CD8⁺, % HLA⁺ CD8 T cells
641 based on CD45⁺CD3⁺CD8⁺HLA-DR⁺, % PD1⁺ CD4 T cells based on CD45⁺CD3⁺CD4⁺PD1⁺, %
642 HLA⁺ fibroblasts based on CD45⁻CD146⁻CD31⁻HLA⁺, % sublining fibroblast based on CD45⁻
643 CD146⁻CD31⁻CD90⁺, % CD27⁺ B cells based on CD45⁺CD3⁻CD14⁻CD20⁺CD27⁺, and % CD11c⁺
644 B cells based on CD45⁺CD3⁻CD14⁻CD20⁺CD11c⁺, respectively.

645

646 **Supplementary Figure 2. Single-cell CITE-seq integrative analysis.** **A.** CCA-based pipeline
647 for integrating mRNA and protein expressions, **B.** Concordance between average mRNA
648 expression and the correlations of corresponding protein and mRNA expression. Black line
649 represents the linear best fit line and the shaded region represents the 95% confidence interval,

650 **C.** Sample sources ($n=82$) in the UMAP space, paired with sensitivity analyses of Harmony
651 parameters based on LISI scores to measure mixture levels on **D.** samples and **E.** cell types, **F.**
652 The effect of varying the selected number of antibodies based on each antibody's specificity: KL
653 divergence equals 0.5 (25 proteins), 0.3 (36 proteins), and 0 (58 proteins), while also varying
654 the number of highly variable genes used: 500/sample (3,164 genes in total) and 1,000/sample
655 (5,751 genes in total) on the mRNA and protein integrative analysis. We used the top 1,000
656 most variable genes per sample and 36 most specific proteins because it best recovered major
657 cell types and more clearly identified rare cell types, **G.** Gene expression of cell type lineage
658 signatures, **H.** Jaccard similarity coefficient that assessed the clustering stability of CTAPs.

659

660 **Supplementary Figure 3. Surface protein specificity and selection for integrative**
661 **analysis.** Kullback-Leibler divergence measured the specificity of each protein across **A.** all
662 cells, **B.** T cells, and **C.** B/plasma cells. Proteins to the left of the red line were chosen for the
663 CCA integration of each set of cells. Canonical correlations for each of the top 20 canonical
664 variates (CVs) from canonical correlation analysis of **D.** all cells, **E.** T cells and **F.** B/plasma
665 cells, respectively.

666

667 **Supplementary Figure 4. Gene and protein features that correlated with the top 20 CVs**
668 **for integrative analysis.** Correlation z scores for genes (top) and proteins (bottom) in **A.** T
669 cells, and **B.** B/plasma cells.

670

671 **Supplementary Figure 5. The single-cell CITE-seq RA reference serves as an RA atlas to**
672 **query other cells. A-B.** We used Symphony³⁷ to map synovial cells from the AMP phase I RA
673 dataset (Zhang, *et al.*, 2019)²¹ onto this AMP phase II single-cell RA reference, **C-D.** We are
674 able to accurately map and predict the cells (Zhang, *et al.*, 2019) from the same cell types with
675 the reference, **E.** We further used Symphony to map cells (Zhang, *et al.*, 2019) from each cell

676 type including B/plasma cells (n=1,142), T cells (n=1,529), fibroblasts (n=1,844), and
677 macrophages (n=750) onto the corresponding cell-type specific references from this study
678 (B/plasma cell, T cell, stromal cell, and myeloid cell) to determine correspondence between cell
679 types defined in Zhang, *et al*, 2019 to the cell states in this study. Each heatmap shows results
680 for the major cell type, with rows corresponding to cell states from this study and columns
681 corresponding to cell states from Zhang, *et al*, 2019. Blue-red color scale indicates the log(OR)
682 for a given pair of states (OR is the ratio of odds of mapping a cluster cell in Zhang, *et al*, 2019
683 to a given cluster of this study compared to odds of mapping other cells in Zhang, *et al*, 2019
684 onto the same cluster of this study), with higher values indicating greater correspondence
685 between Zhang, *et al.*, 2019 and the fine-grained cell states in this study.

686

687 **Supplementary Figure 6. T cell-specific analysis. A.** T cell UMAP colored by fine-grained cell
688 state clusters, **B.** Expression of selected surface proteins among T cells. Cells are colored from
689 blue (low) to yellow (high), **C.** Heatmap of surface protein expression in T cell clusters colored
690 according to the average normalized expression across cells in the cluster, **D.** Heatmap of gene
691 expression in T cell clusters colored according to the average normalized expression across
692 cells in the cluster, scaled for each gene across clusters, **E.** Distribution of T cells across
693 clusters, stratified by CTAP. The size of each segment of each bar corresponds to the average
694 proportion of cells in that cluster across donors from that CTAP. **F.** Number of T cells per donor,
695 stratified by CTAP. Points represent donors. Box plots show median (vertical bar), 25th and
696 75th percentiles (lower and upper bounds of the box, respectively) and 1.5 x IQR (or
697 minimum/maximum values; end of whiskers).

698

699 **Supplementary Figure 7. B/plasma cell-specific analysis. A.** B/plasma cell UMAP colored by
700 fine-grained cell state clusters, **B.** Expression of selected surface proteins among B/plasma
701 cells. Cells are colored from blue (low) to yellow (high), **C.** Heatmap of surface protein

702 expression in B/plasma cell clusters colored according to the average normalized expression
703 across cells in the cluster, **D.** Heatmap of gene expression in B/plasma cell clusters colored
704 according to the average normalized expression across cells in the cluster, scaled for each
705 gene across clusters, **E.** Number of B/plasma cells per donor, stratified by CTAP. Points
706 represent donors. Box plots show median (vertical bar), 25th and 75th percentiles (lower and
707 upper bounds of the box, respectively) and 1.5 x IQR (or minimum/maximum values; end of
708 whiskers), **F.** Heatmap of correlations between select T and B cell subsets, colored by Pearson
709 correlation between per-donor proportions, **G.** Distribution of B/plasma cells across clusters,
710 stratified by CTAP. The size of each segment of each bar corresponds to the average
711 proportion of cells in that cluster across donors from that CTAP.

712

713 **Supplementary Figure 8. NK cell-specific analysis. A.** NK cell UMAP colored by fine-grained
714 cell state clusters, **B.** Expression of selected surface proteins among NK cells colored from blue
715 (low) to yellow (high), **C.** Heatmap of surface protein expression in NK cell clusters colored
716 according to the average normalized expression across cells in the cluster, **D.** Heatmap of gene
717 expression in NK cell clusters colored according to the average normalized expression across
718 cells in the cluster, scaled for each gene across clusters, **E.** Distribution of NK cells across
719 clusters, stratified by CTAP. The size of each segment of each bar corresponds to the average
720 proportion of cells in that cluster across donors from that CTAP. **F.** Number of NK cells per
721 donor, stratified by CTAP. Points represent donors. Box plots show median (vertical bar), 25th
722 and 75th percentiles (lower and upper bounds of the box, respectively) and 1.5 x IQR (or
723 minimum/maximum values; end of whiskers).

724

725 **Supplementary Figure 9. Myeloid cell-specific analysis. A.** Myeloid cell UMAP colored by
726 fine-grained cell state clusters, **B.** Expression of selected surface proteins among myeloid cells
727 colored from blue (low) to yellow (high), **C.** Heatmap of surface protein expression in myeloid

728 cell clusters colored according to the average normalized expression across cells in the cluster,
729 **D.** Heatmap of gene expression in myeloid cell clusters colored according to the average
730 normalized expression across cells in the cluster, scaled for each gene across clusters, **E.**
731 Distribution of myeloid cells across clusters, stratified by CTAP. The size of each segment of
732 each bar corresponds to the average proportion of cells in that cluster across donors from that
733 CTAP. **F.** Number of myeloid cells per donor, stratified by CTAP. Points represent donors. Box
734 plots show median (vertical bar), 25th and 75th percentiles (lower and upper bounds of the box,
735 respectively) and 1.5 x IQR (or minimum/maximum values; end of whiskers).

736

737 **Supplementary Figure 10. Stromal- and endothelial-specific analysis. A.** Stromal cell
738 UMAP colored by fine-grained cell state clusters, **B.** Expression of selected surface proteins
739 among stromal cells colored from blue (low) to yellow (high), **C.** Heatmap of surface protein
740 expression in stromal cell clusters colored according to the average normalized expression
741 across cells in the cluster, **D.** Heatmap of gene expression in stromal cell clusters colored
742 according to the average normalized expression across cells in the cluster, scaled for each
743 gene across clusters, **E.** Distribution of stromal cells across clusters, stratified by CTAP. The
744 size of each segment of each bar corresponds to the average proportion of cells in that cluster
745 across donors from that CTAP, **F.** Number of stromal cells per donor, stratified by CTAP. Points
746 represent donors. Box plots show median (vertical bar), 25th and 75th percentiles (lower and
747 upper bounds of the box, respectively) and 1.5 x IQR (or minimum/maximum values; end of
748 whiskers), **G.** Endothelial cell UMAP colored by fine-grained cell state clusters, **H.** Expression of
749 selected surface proteins among endothelial cells colored from blue (low) to yellow (high), **I.**
750 Heatmap of gene expression in endothelial cell clusters colored according to the average
751 normalized expression across cells in the cluster, scaled for each gene across clusters, **J.**
752 Distribution of endothelial cells across clusters, stratified by CTAP. The size of each segment of
753 each bar corresponds to the average proportion of cells in that cluster across donors from that

754 CTAP. **K.** Number of endothelial cells per donor, stratified by CTAP. Points represent donors.
755 Box plots show median (vertical bar), 25th and 75th percentiles (lower and upper bounds of the
756 box, respectively) and 1.5 x IQR (or minimum/maximum values; end of whiskers).

757

758 **Supplementary Figure 11. Clinical and histologic association results using CNA. A.**

759 Representative histologic images illustrating different levels of density and aggregation scores,

760 **B.** For each broad cell type, we identified and presented specific cell populations that were

761 associated with histologic density and aggregation scores by controlling age, sex, and number

762 of cells per sample; cells in red/blue represent positive/negative associations that pass FDR

763 0.05 correlation, and global permutation p-value is also shown for each association testing.

764

765 **Supplementary Figure 12. Association of single-cell RA CTAPs with different clinical**

766 **characteristics. A.** Clinical, histologic, and ultrasound parameters of patients in each CTAP.

767 For all box plots, each dot represents a donor; boxes show median (vertical bar), 25th and 75th

768 percentiles (lower and upper bounds of the box, respectively) and 1.5 x IQR (or

769 minimum/maximum values; end of whiskers), **B** Association of Krenn inflammation and Krenn

770 lining with CTAPs, adjusting covariates for age, sex, cell number, and clinical collection site.

771 Percent of variance explained by CTAPs only and p-value are calculated with ANOVA test, **C.**

772 CCP levels among seropositive patients alone, **C.** CTAP frequency among seropositive (CCP+,

773 RF+, or both) versus seronegative patients, **D.** CTAP frequency by sex, **E.** CTAP frequency by

774 smoking history, **F.** CTAP frequency by anatomic site of synovial biopsy **H.** Number of patient

775 samples for each CTAP between biopsy and synovectomy, **I.** Collection/cryopreservation sites,

776 **J.** Association of age and RA duration with CTAPs, adjusting covariates for age, sex, cell

777 number, and clinical collection site. Percentage of variance explained by CTAPs alone and p-

778 value are calculated with ANOVA test. 95% confidence intervals are shown. **K.** Sample

779 distributions across CTAPs by recruitment cohort, **L.** Overview of clinical variables for patient
780 samples distributed by CTAPs. “X” represents missing data for a particular sample.

781

782 **Supplementary Figure 13. Single-cell cellular sources of cytokines and cytokine**

783 **receptors.** Z-scored pseudo-bulk expression across the identified 77 cell states of a curated
784 cytokine and receptor list from KEGG (M9809)¹⁰⁵ is shown. 138 cytokines and receptors that are
785 expressed in more than 3% of total single cells are shown here.

786

787 **Supplementary Figure 14. Assigning repeated biopsy and flow samples to CTAPs. A.**

788 Mapping three repeated biopsy samples onto CTAP PC space based on the cell type
789 abundance, **B.** We evaluated CTAP stability by randomly selecting 1,0000 samples and
790 measuring the Mahalanobis distance between these random samples to the baseline samples,
791 **C.** Mapping flow cytometry samples onto CTAP PC space, **D.** Mahalanobis distance of each
792 flow sample to each CTAP centroid; the original CTAP of the single-cell samples from the same
793 donors are labeled as red.

794

795

796 **Supplementary Tables**

797 **Supplementary Table 1. Statistics of demographic, clinical, and histology metrics across**
798 **recruitment groups and disease activity levels.**

799 **Supplementary Table 2. Antibodies used in CITE-seq panel.**

800 **Supplementary Table 3. Antibodies used in flow cytometry panels.**

801 **Supplementary Table 4. Proportions of cell types within each CTAP compared with the**
802 **proportions across all samples.** For each identified CTAP, we named it based on the cell

803 types if their average proportions were higher in it compared to their average across all
804 samples.

805 **Supplementary Table 5. Examination of our previously identified RA expanded single-cell**
806 **clusters²¹ in the single-cell dataset from this study.** 95% confidence interval (CI) for the
807 odds ratio (OR) and one-sided MASC (mixed-effects modeling of associations)¹⁰⁶ p-value are
808 shown.

809 **Supplementary Table 6. Each identified single-cell cluster's median proportion across**
810 **samples within each CTAP.**

811 **Supplementary Table 7. Details and parameters of single-cell integration and clustering**
812 **for each cell type.** For each broad cell type, we present the number of variable genes, KL
813 divergence threshold for protein selection, Harmony parameters for batch correction, and
814 clustering resolution.

815 **Supplementary Table 8. Differentially expressed genes and relative statistics per CITE-**
816 **seq cluster.** For each broad cell type, pseudo-bulk differential expression is used with a linear
817 regression model accounting for donor and number of UMIs to identify genes that were more
818 highly expressed inside vs. outside the cluster. Likelihood ratio test p-values and fold change
819 are presented for prioritized markers.

820 **Supplementary Table 9. Statistics of single-cell cell type-specific associations with**
821 **CTAPs and histologic parameters.** We show the statistics for each CTAP-specific association
822 testing and histologic parameter association testing.

823 **Supplementary Table 10. Statistics of demographic, clinical, and histologic metrics**
824 **across RA CTAPs.** We show the statistics of clinical characteristics, demographic variables,
825 medications, and treatment groups across RA CTAPs.

826

827

828 **Methods**

829 **RA patient recruitment and clinical data collection**

830 The Accelerating Medicines Partnership (AMP) Network for RA and SLE constructed a cross-
831 sectional cohort - samples were collected from 13 clinical sites across the United States and 2
832 sites in the United Kingdom. The collection occurred over the course of a 45-month period from
833 October 2016 to February of 2020. The study was performed in accordance with protocols
834 approved by the institutional review board. Demographics, RA clinical data, clinical
835 assessments, and measurements of ESR and CRP were performed at the baseline visit. Data
836 collected include age, sex, RA duration, RF or anti-CCP status, RA treatments, tender and
837 swollen joint counts. ESR and CRP were measured using commercial assays in each
838 institution's clinical laboratory. Disease activity for each subject was calculated using a DAS28-
839 CRP3 validated instrument^{91,92}.

840

841 **Synovial tissue collection and processing**

842 Synovial tissue samples were obtained from ultrasound-guided biopsies or surgical procedures.
843 Of the 82 samples that completed the tissue processing pipeline, 54 samples were biopsies
844 obtained with a Quick-Core needle, 15 samples were biopsies obtained with portal and forceps,
845 10 samples were collected during arthroplasty surgery, and 6 samples were collected during
846 surgical synovectomy procedures. All specimens consisted of a median of 13 samples (range
847 4-36), of which 6-8 fragments were fixed in formalin for subsequent paraffin embedding and
848 processing for histologic analysis. The remaining fragments were cryopreserved in one or more
849 aliquots in Cryostor CS10 (Sigma-Aldrich) cryopreservation media. Samples were shipped to a
850 central biorepository site until sample collection was complete. They were then transited to the
851 central pipeline site, where samples were thawed and processed in batches.

852

853 **Histology assessment, definition of density and aggregation for RA synovium**

854 In order to exclude low-quality synovial tissue samples from our multi-omics tissue processing
855 platform, we analyzed hematoxylin and eosin-stained slides of formalin-fixed, paraffin-
856 embedded synovial tissue from each patient. At least six tissue fragments per patient were
857 included in the analysis to mitigate sampling bias. Synovial tissue was identified as previously
858 described²¹, and samples that lacked any discernible synovial tissue were excluded from further
859 analysis. To separate histologic domains of the density of the infiltrate and the extent of
860 formation of discrete aggregates that are not distinguished by the Krenn inflammatory infiltrate
861 domain, we developed consensus semiquantitative four point scales for density and aggregate
862 radial size with a custom atlas using a test set of tissues from the Birmingham Early Arthritis
863 Cohort⁹³, scored by three pathologists. This approach was validated by scoring tissues from the
864 first AMP RA cohort²¹, achieving an intra-class correlation coefficient score of 0.896 for
865 summary mean density score of fragments for each tissue and kappa 0.862 for the worst case
866 aggregate score achieved in each tissue. Equivalent ICC figures for the summary mean scores
867 of fragments for Krenn inflammatory domain and Lining layer thickness domains were 0.937 and
868 0.646 respectively. Three pathologists independently determined Krenn lining and inflammatory
869 infiltrates scores (0-3 each)⁹⁴, cellular density scores (0-3), and aggregate (0-3) scores for each
870 tissue sample, and the mode of the three scores was used for further analysis.

871

872 **Tissue disaggregation, live cell sorting, and cell allocations**

873 For pipeline analysis, cryopreserved synovial tissue samples were thawed and disaggregated
874 into single-cell suspension as previously described⁹⁵. Briefly, thawed synovial tissue fragments
875 were mechanically and enzymatically separated in digestion buffer (Liberase TL (Roche) 100
876 µg/ml and DNase I (New England Biolabs) 100 µg/ml in RPMI) in 37°C water bath for 30 min.
877 Single-cell suspensions from disaggregated synovial tissues were stained with anti-CD235a
878 antibodies (clone 11E4B-7-6 (KC16), Beckman Coulter) and Fixable Viability Dye eFlour 780

879 (eBioscience/ThermoFisher). Live non-erythrocyte cells (viability dye⁻ CD235⁻) were collected by
880 fluorescence-activated cell sorting (BD FACSAria Fusion). Cells were allocated as follows, in
881 order of priority: (1) 60,000 cells for CITE-seq analysis; (2) 50,000 cells for flow cytometry and
882 bulk RNA-seq analysis; (3) remaining cells re-frozen in aliquots of 70,000 - 100,000 cells in
883 CryoStor CS10 for other analyses (e.g. single-cell ATAC-seq and immune cell repertoire
884 studies). Samples with fewer than 60,000 cells were applied to CITE-seq analysis alone.

885

886 **Flow cytometry and bulk RNA-seq**

887 Up to 50,000 sorted live synovial cells were stained with the following antibodies to define cell
888 subsets: CD3 (UCHT1), CD4 (OKT4), CD8 (SK1), CD11c (3.9), CD14 (M5E2), CD19 (HIB19),
889 CD27 (M-T271), CD31 (WM59), CD45 (HI30), CD90 (5E10), CD146 (P1H12), HLA-DR (L234),
890 PD-1 (EH12.2H7). All antibodies were purchased from Biolegend, and staining was performed
891 in the presence of Fc block (eBioscience/ThermoFisher, True-Stain Monocyte Blocker
892 (Biolegend), and Brilliant Stain Buffer (BD Bioscience). We collected flow cytometry data in
893 conjunction with fluorescence-activated cell sorting of up to 1,000 B cells (CD45⁺CD3⁻CD14⁻
894 CD19⁺), fibroblasts (CD45⁻CD31⁻CD146⁻), macrophages (CD45⁺CD3⁻CD14⁺), and T cells
895 (CD45⁺CD3⁺CD14⁻) on a BD FACSAria Fusion cell sorter.

896

897 **Single-cell CITE-seq antibody staining, RNA library preparation, and sequencing**

898 Antibody staining using TotalSeqTM-A antibodies was performed as per the recommended
899 protocol (BioLegend). Briefly, we first curated a list of surface proteins based on markers of cell
900 states identified in previous RA studies and TotalSeqTM-A antibodies available at the time. To
901 identify optimized concentrations of these antibodies for synovial tissue, we conducted a series
902 of pilot studies where we titrated antibodies and measured their staining quality with mean
903 expression (i.e., intensity) and Kullback-Leibler (K-L) divergence (i.e., specificity). We calculated
904 K-L divergence by comparing the distribution across mRNA-defined clusters of cells expressing

905 the protein highly (>85th percentile) versus the null distribution of all cells. If an antibody had low
906 mean staining and low K-L divergence, we removed it from the panel. If it had high mean
907 staining and low K-L divergence, we titrated it at a lower concentration.

908 After optimizing the panel and final concentrations (**Supplementary Table 2**), we
909 prepared a cocktail of TotalSeq antibodies and centrifuged for 10 min at 14,000G to remove
910 precipitates. Up to 60,000 sorted live synovial cells were pre-incubated with Human TruStain
911 FcX (BioLegend) in Cell Staining Buffer (BioLegend) for 10 minutes prior to the addition of 100
912 uL of the antibody cocktail. Single-cell RNA-seq for all synovial samples was performed by the
913 BWH Single Cell Genomics Core. After a 30-minute incubation at 4°C, cells were washed twice
914 in the Cell Staining Buffer and resuspended in 0.4% BSA/PBS. After performing a live cell count
915 using Trypan blue, cells were resuspended at 1,000 cells per microliter and a maximum of
916 15,000 cells were loaded into a Chromium Next GEM Chip G (10x Genomics). For samples with
917 fewer than 15,000 live cells, all cells were loaded into the chip. cDNA and library generation was
918 done according to the manufacturer's protocol. mRNA libraries were sequenced to an average
919 of 50,000 reads per cell using Illumina Novaseq S4. CITE-seq antibody-derived tag (ADT)
920 libraries were sequenced to an average of 5,000 reads per cell using Illumina Hi-Seq X Ten.

921

922 **Single-cell CITE-seq gene expression and protein expression quantification**

923 We quantified mRNA and antibody-derived tag (ADT) unique molecular identifier (UMI) counts
924 using Cell Ranger v3.1.0. First, raw BCL files were demultiplexed using cellranger mkfastq with
925 default parameters to generate FASTQ files. Then, these FASTQ files were aligned to the
926 GRCh38 human reference genome using Cellranger v3.1.0. Gene and ADT reads were
927 quantified simultaneously using cellranger count.

928

929 **Quality control of single-cell CITE-seq data**

930 We show each QC step in **Supplementary Figure 1**. Specifically, we performed consistent QC
931 to remove cells that expressed fewer than 500 genes or contained more than 20% of their total
932 UMIs mapping to mitochondrial genes, resulting in 403,596 cells. Then, we performed sample-
933 level QC and removed samples with a low percentage (< 40%) of cells passing QC. We
934 removed three lower-quality samples (processed on the same day) with less than 40% of cells
935 passing QC compared to 71% for other good quality samples. In the end, we obtained 393,344
936 cells from 82 samples that passed QC.

937

938 We identified and removed doublets based on a combined strategy:

939

940 1. To detect doublets/multiplets based on gene count, we utilized the Scrublet⁹⁶ framework
941 implemented in Python on each sample. We input the full raw, unnormalized UMI count
942 data into the Scrublet() function with default parameters. We determined the doublet
943 scores and the threshold for doublet detection by using the scrub_doublets() function
944 with the following parameters: min_counts = 2, min_cells = 3, min_gene_variability_pctl
945 = 85, and n_prin_comps = 30. Based on the distribution of modes of simulated doublet
946 gene expression distributions, we set the threshold at 0.66. Based on this threshold, we
947 identified 4.5% of cells as doublets.

948 2. Using protein expression, we trained an LDA (Linear Discriminant Analysis)-based
949 classifier on non-doublet cells and then predicted the posterior probability of doublets
950 using cell-type-specific antibodies (CD45, CD3, CD14, CD19, CD20, CD56, CD1C,
951 PDPN, CD146), which improved the precision of doublet detection in a multimodal
952 fashion. We obtained 314,030 cells after doublet detection.

953

954 To assess the accuracy of protein measurements in CITE-seq, we selected antibodies for
955 surface markers of each cell-type lineage: T cells (CD45 and CD3D), NK cells (CD45, CD56,

956 CD16, and IL17R), B cells and plasma cells (CD45 and CD19), macrophages (CD45 and
957 CD14), classical dendritic cells (cDCs, CD11c), fibroblast (PDPN), mural cells (PDPN and
958 CD146), and endothelial cells (CD146) (**Supplementary Table 2**). For flow cytometry, we used
959 13 antibodies (**Supplementary Table 3**). We measured the Pearson correlation between the
960 per-donor proportion of cells in each gate across donors. We removed surface proteins with low
961 expression overall.

962

963 **mRNA feature normalization, selection, and scaling**

964 *Global:* For each cell, we normalized the expression of each gene with $\log(1 + \text{UMIs for}$
965 $\text{gene}/\text{total UMIs in cell} * 10,000)$. Then, we selected the top 1,000 most highly variable genes in
966 each sample based on a variance stabilizing transformation (VST)⁹⁷, which considers overall
967 variance of the transcript per sample. We excluded cell cycle genes from “Seurat::cc.genes” for
968 downstream analysis. We then pooled the most highly variable genes across all samples for a
969 cell type into a data matrix and performed z-score scaling on each gene to have mean=0 and
970 variance=1 across cells.

971

972 *By cell type:* We carried out the same normalization, feature selection, and scaling steps as
973 described for the global analysis, but on only the cells of each given cell type.

974

975 **Protein feature normalization, selection, and scaling**

976 *Global:* For each cell, we normalized each protein with centered-log ratio (CLR):
977 $\{\ln(x_1/g(x)), \dots, \ln(x_n/g(x))\}$, where x is a vector of protein counts⁹⁸. For each feature, we then
978 performed z-score scaling on each protein to have mean 0 and variance 1 across cells. To
979 improve discrimination of signal and background in visualizations, we corrected for antibody'
980 background staining by fitting a Gaussian mixture model (with the normalmixEM function from
981 the mixtools R package; $k = 2$, $\lambda = 0.5$) to the CLR-normalized expression of each protein

982 in each cell type. Then we calculated the mean of the first (lower) Gaussian in each cell type,
983 identified the lowest mean across cell types, and subtracted this value—representing
984 background—from all cells' expression of the protein (with a lower bound of 0 for any values
985 that would otherwise become negative).

986

987 To select variable proteins, we measured Kullback-Leibler (KL) divergence for each protein by
988 comparing the distribution of cells with normalized expression above the 75th percentile for that
989 protein across broad cell-type clusters, versus the distribution of all cells across broad cell-type
990 clusters. For each feature, we then performed z-score scaling on each protein to have mean=0
991 and variance=1 across cells. We used a KL-divergence threshold of 0.3.

992

993 *By cell type:* We carried out the same normalization as described for global analysis, but only on
994 the cells of each given cell type. For T and B/plasma cells only, we conducted protein feature
995 selection and scaling as described for global analysis. We removed proteins expressed in < 1%
996 of cells and selected variable proteins based on KL divergence (computed as described above
997 except using the 85th percentile to define the distribution of protein-expressing cells). Proteins
998 with KL divergence greater than or equal to 0.025 were considered variable.

999

1000 **A unimodal dimensionality reduction strategy for single-cell gene expression**

1001 For cell-type-specific analysis of myeloid cells, fibroblasts/mural cells, endothelial cells, and
1002 natural killer cells, we used a unimodal pipeline to reduce the dimensionality of the data based
1003 on mRNA expression. For each cell type, we used truncated principal component analysis
1004 (PCA) as implemented in the `prcomp_irlba` function from the `irlba` R package⁹⁹ and calculated
1005 20 principal components (PCs) based on the scaled mRNA data. We then corrected sample-
1006 driven batch effects with the `HarmonyMatrix` function from the `harmony` R package³⁶ with

1007 parameters as specified in **Supplementary Table 7** and projected the cells into two dimensions
1008 with UMAP¹⁰⁰.

1009

1010 **A multi-modal dimensionality reduction strategy for CITE-seq data**

1011 For global analysis of all cell types and cell-type-specific analysis of T and B/plasma cells, we
1012 used a multi-modal pipeline to integrate mRNA and surface protein expression from the same
1013 cells and project the cells into a low dimensional embedding informed by both modalities¹⁰¹.

1014 After scaling the protein features so that their total variance was equal to the total variance of
1015 the mRNA features, we used canonical correlation analysis (CCA) as implemented in the `cc`
1016 function from the CCA R package to calculate canonical variates (CVs)¹⁰² based on the scaled
1017 mRNA and surface protein data; these are projections of cells onto axes defined by maximally
1018 correlated linear combinations of genes and surface proteins that capture the greatest amount
1019 of shared variance. For further analysis, we selected the top 20 CVs with highest canonical
1020 correlations, as defined in the mRNA space. We then corrected sample-driven batch effects with
1021 the HarmonyMatrix function from the harmony R package³⁶ with parameters and projected the
1022 cells into two dimensions with UMAP¹⁰⁰.

1023

1024 **Graph-based clustering, differential gene expression, and cell type annotation**

1025 We then constructed shared nearest neighbor graphs derived from the top 20 CVs/PCs and
1026 applied graph-based Louvain clustering¹⁰³ at various resolution levels (0.2, 0.4, 0.6, 0.8, 1.0).

1027 We selected optimized resolution values for each cell type (1.2 for T cells, 0.8 for NK cells, 0.6
1028 for myeloid cells, 0.6 for B cells, 0.6 for stromal cells, 0.3 for endothelial cells) to gain the
1029 biological interpretations that made the most sense. We incorporated the number of variable
1030 genes chosen per sample and parameters for each cell type's analytical pipeline in

1031 **Supplementary Table 7**. In the end, we identified 24 T cell clusters (94,056 cells), 9 B cell
1032 clusters (30,697 cells), 14 NK clusters (8,497 cells), 15 myeloid clusters (76,181 cells), 5

1033 endothelial clusters (25,044 cells), and 10 stromal cell clusters (79,555 cells), for a total of 77
1034 clusters.

1035
1036 For each major cell type, we identified differentially expressed mRNA features and surface
1037 proteins by comparing cells from one cluster with all the other cells. We collapsed single-cell
1038 mRNA and protein expression profiles into pseudo-bulk count matrices by summing the raw
1039 UMI counts for each gene or surface protein across all cells from the same donor and cluster.
1040 For mRNA, we tested all mRNA features that were detected in more than 100 cells per donor
1041 with non-zero UMI counts. For each feature, we normalized counts in each pseudo-bulk sample
1042 into counts per million (CPM). Using linear models, we estimated the effect of each cluster for
1043 each feature on pseudo-bulk expression accounting for effects from the donor and the number
1044 of UMIs for each pseudo-bulk sample. Next, we used likelihood ratio tests (LRT) between two
1045 models: one that has the cluster variable, and another that doesn't have the cluster variable.
1046 Finally, we selected a feature to be a cluster marker if it had a fold change greater than 2 and p-
1047 value less than FDR 5%, which is $p < 0.05/(\text{number of tested genes} \times \text{number of clusters})$. We
1048 repeated a similar analytical pipeline of normalization and scaling, feature selection, multi-modal
1049 dimensionality reduction, clustering, and differential expression analysis for T cells ($p < 1.5 \times 10^{-6}$),
1050 B cells and plasma cells ($p < 1.9 \times 10^{-6}$), NK cells ($p < 1.6 \times 10^{-6}$), myeloid cells ($p < 1.8 \times 10^{-8}$),
1051 stromal cells ($p < 4.3 \times 10^{-7}$), and endothelial cells ($p < 1.2 \times 10^{-6}$), respectively. Furthermore, we
1052 annotated each cell-type cluster based on literature. We present cluster-specific marker genes
1053 and relative statistics in **Supplementary Table 8**.

1054
1055 **Building and mapping to global and cell-type-specific references**
1056 We used the `buildReferenceFromHarmonyObj()` function from the Symphony³⁷ package to build
1057 integrated reference atlases for the global and cell-type specific atlases from the Harmony
1058 objects. To find concordance between cell types from our previous study²¹ and this study, we

1059 used the Symphony mapQuery() function to map the 5,254 scRNA-seq query cells from Zhang
1060 *et al*, 2019 onto the global and respective cell-type reference atlases. We predicted reference
1061 cell types and states for the query cells using the knnPredict() function with k=5. For the cell-
1062 type specific mapping, we excluded reference dendritic cells or mural cells because they were
1063 absent in the query. Note that because the gene expression matrices for the reference (this
1064 study) and query²¹ datasets were generated using different versions of Gencode (version 19 vs.
1065 version 29, respectively), certain genes were named differently between the two datasets (e.g.
1066 *IL8* and *CXCL8* are synonyms for the same gene ENSG00000169429). Because the mapping
1067 procedure uses overlapping gene names between reference and query, we “synced” the query
1068 gene names to the version 29 names using the shared Ensembl IDs (which do not change
1069 between Gencode versions) using the Gencode .gtf files. This converted 9,663 query gene
1070 names, and the synced expression matrix was used as input to mapping.

1071

1072 **Identification of CTAPs based on single-cell cell-type abundance**

1073 We identified six cell-type abundance phenotypes (CTAPs) based on hierarchical clustering on
1074 cell-type abundances for each CITE-seq patient sample. The differences across CTAPs are
1075 also reflected in the PCA space. We named each CTAP based on the cell types whose average
1076 proportions were higher among samples in the CTAP compared to their average across all
1077 samples (**Supplementary Table 4**). To assess the stability of CTAPs, 1) We first bootstrapped
1078 the patient samples and clustered the resampled dataset, 2) For every original CTAP subgroup,
1079 we found the most similar cluster (based on Jaccard similarity) in each resampled clustering and
1080 recorded that value, giving us the maximum Jaccard similarity coefficient for each CTAP. The
1081 Jaccard similarity coefficient can be a value between 0 and 1, where 1 indicates complete
1082 overlap and 0 indicates no overlap between two sets of the clustering results, 3) We repeated
1083 the above two steps 1e4 times and calculated the mean Jaccard similarity coefficient. We
1084 performed this process on different possible numbers of patient subgroups ranging from 2 to 10,

1085 and evaluated the statistical stability retaining in-group similarity. We selected six clusters as
1086 CTAPs because they gave us relatively high stability (mean Jaccard similarity coefficient=0.727)
1087 and also high granularity of biologically meaningful interpretations.

1088

1089 **Covarying neighborhood analysis (CNA) to identify cell populations associated with** 1090 **patient CTAP membership**

1091 We evaluated whether the global RA CTAPs are associated with changes in the relative
1092 abundances of cell states within each of our six major cell types, which would indicate that these
1093 CTAP groupings reflect both coarse (relative abundance of major cell types) and fine-scale
1094 heterogeneity in synovial tissue composition.

1095

1096 For each major cell type, we used CNA⁷⁷ to associate sample-level attributes to the abundances
1097 of cell states within that cell type. CNA defines many small cell neighborhoods in the batch-
1098 corrected low-dimensional space and stores that fractional abundance of cells from each
1099 sample in each neighborhood in a neighborhood abundance matrix (NAM). By decomposing the
1100 NAM with principal component analysis, CNA defines NAM-PCs within each cell type that
1101 capture axes of heterogeneity defined by groups of neighborhoods whose abundances vary in a
1102 coordinated manner. Here, we use CNA to test for associations between sample-level clinical
1103 characteristics and the abundance of covarying neighborhood groups. For associations with
1104 histologic metrics such as histology density and aggregate scores, we only used samples that
1105 passed histology-level QC grades (Grade A and B). We also use CNA to identify neighborhoods
1106 that are associated with one CTAP compared to other CTAPs.

1107

1108 To perform CNA, we used the `tl.association()` function in the `cna` Python package with default
1109 parameters and top four NAM-PCs as inputs, while controlling for the “age”, “sex”, and “number
1110 of cells per sample” as covariates. As CNA utilizes a permutation test, we determined a

1111 significant association based on a global permutation $p < 0.05$. For visualization of local
1112 associations, which indicate the particular neighborhoods driving a found global association, we
1113 used the 5% FDR threshold from CNA to determine which neighborhoods featured a locally
1114 significant correlation. In violin plots, we plotted this threshold as dotted lines. In UMAP plots,
1115 we colored neighborhoods that pass local significance based on the intensity of their correlation,
1116 with red indicating a higher positive correlation, while we colored neighborhoods that did not
1117 attain local significance as grey. We used a modified version of CNA, available on Github, which
1118 included the following features: 1) scaling the variance per neighborhood within the NAM
1119 inversely to the sample size of the source sample for that neighborhood's anchor cell such that
1120 total variance across all neighborhoods anchored on cells from the same sample sums to 1, and
1121 2) the addition of a pseudo-count, a small number that was added to each entry in the NAM.
1122 Using CNA, we tested associations of cell neighborhoods that are associated with histology,
1123 ultrasound, clinical metrics, and also each CTAP group. The statistics are in **Supplementary**
1124 **Table 9**.

1125

1126 **Modeling histologic, clinical, and demographic characteristics using CTAPs**

1127 We used linear mixed modeling to model each histologic parameter and clinical demographic
1128 variable using single-cell CTAPs. Only samples that passed histology-level QC (Grade A and B)
1129 were included to seek an association between molecular-level categories and histologic metrics.
1130 Taking histologic density Y as an example, we fitted a mixed-effect model for each CTAP with
1131 the number of cells per sample as a cell-level fixed effect, age and sex as demographical level
1132 fixed effects, and clinical collection site as a random effect covariate:

1133 Full model: $Y_i = 0 + \sum_j \beta_j X_{i,j} + \beta_{age} X_{i,age} + \beta_{sex} X_{i,sex} + \beta_{tech} X_{i,tech} + (1|site)$,

1134 Null model: $Y_i = 0 + \beta_{age} X_{i,age} + \beta_{sex} X_{i,sex} + \beta_{tech} X_{i,tech} + (1|site)$

1135

1136 where β_j is the effect size for each CTAP j for sample i , β_{age} is a vector of age values and β_{sex}
1137 is a vector of sex values, β_{tech} is a vector containing a technical covariate that captures the
1138 number of cells for each single sample, X_i is the one-hot encoded variable for sample i in CTAP
1139 j as appropriate, and $(1|site)$ is the random effect for clinical collection sites. Thus, we used the
1140 full model to calculate the corrected values of CTAPs accounting for these technical, cell-level,
1141 and donor-level covariates. For modeling age and disease duration, we used a similar model
1142 but we removed the age fixed effect from both the full and null model. We obtained percent of
1143 variance explained by the CTAPs only by subtracting the variance explained by the null model
1144 from the variance explained by the full model. ANOVA p-value was also calculated. The R
1145 package lme4 was used for the mixed effect modeling¹⁰⁴.

1146

1147 **Classifying flow cytometry samples into RA CTAPs**

1148 We provided a proof-of-concept framework to assign RA samples processed by other data
1149 modalities (e.g., flow cytometry) to the RA CTAPs generated from single-cell technology.
1150 Specifically for **Figure 6E**, 1) we quantified the major cell type abundances in a sample using
1151 flow cytometry based on cell type markers derived from the single-cell technology, then 2) we
1152 mapped each flow sample to the principal component space generated from the CTAP single-
1153 cell cell type abundance based on the same features. Here, the features are T, B, Myeloid,
1154 stromal, endothelial, and NK cell canonical markers. Now that each flow sample has a loading in
1155 the original single-cell abundance space, 3) we built a Mahalanobis-distance-based nearest-
1156 neighbor classifier to measure the distance of a flow sample to each of the CTAP centroids. We
1157 use Mahalanobis distance to handle the covariance, because our CTAP clusters in PC space
1158 are elliptical shaped covariances rather than circular shapes. 4) For each flow sample, we
1159 assigned a CTAP label based on which CTAP centroid had the smallest Mahalanobis distance.

1160 We calculated the accuracy of our classifications based on a subset (n=15) of RA synovial
1161 tissues sent to both single-cell CITE-seq and flow cytometry.

1162

1163

1164 **Acknowledgements**

1165 This work was supported by the Accelerating Medicines Partnership (AMP) in Rheumatoid
1166 Arthritis and Lupus Network. AMP is a public-private partnership (AbbVie Inc., Arthritis
1167 Foundation, Bristol-Myers Squibb Company, Foundation for the National Institutes of Health,
1168 GlaxoSmithKline, Janssen Research and Development, LLC, Lupus Foundation of America,
1169 Lupus Research Alliance, Merck Sharp & Dohme Corp., National Institute of Allergy and
1170 Infectious Diseases, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Pfizer
1171 Inc., Rheumatology Research Foundation, Sanofi and Takeda Pharmaceuticals International,
1172 Inc.) created to develop new ways of identifying and validating promising biological targets for
1173 diagnostics and drug development. Funding was provided through grants from the National
1174 Institutes of Health (UH2-AR067676, UH2-AR067677, UH2-AR067679, UH2-AR067681, UH2-
1175 AR067685, UH2- AR067688, UH2-AR067689, UH2-AR067690, UH2-AR067691, UH2-
1176 AR067694, and UM2- AR067678). Accelerating Medicines Partnership and AMP are registered
1177 service marks of the U.S. Department of Health and Human Services. This work was also
1178 supported by a Rheumatology Research Foundation Investigator Award and Arthritis National
1179 Research Foundation award (to A.H.J.); NIH NHGRI T32HG002295 and NIAMS T32AR007530
1180 (to A. Nathan); NIH NIAMS K08AR077037, Rheumatology Research Foundation Innovative
1181 Research Award, and Burroughs Wellcome Fund Career Award in Medical Sciences (to K.
1182 Wei); NIH NIGMS T32GM007753 (to J.B.K.); NIH NIAID T32AR007258 (to K.S.); Research into
1183 Inflammatory Arthritis Centre Versus Arthritis (22072), IMI-RTCure (777357) and the NIHR
1184 Birmingham Biomedical Research Centre (BRC-1215-20009) (to A.F. and D.S.T.); NIH NIAMS

1185 K08AR072791 and Burroughs Wellcome Fund Career Award in Medical Sciences (to D.A.R.);
1186 NIH NIAID R01AI148435 (to L.T.D.); NIH NIAMS R21AR071670 and P30 AR069655 (to J.H.A.);
1187 NIH NIAMS R01AR073833 and R01AR073290 (to M.B.B.); NIH NHGRI U01HG009379 and
1188 NIAMS R01AR063759 (to S.R.). We especially acknowledge people in the AMP RA/SLE
1189 Network: Arnon Arazi, Celine Berthier, Jill Buyon, Maria Dall'Era, Anne Davidson, Betty
1190 Diamond, Andrea Fava, Jennifer Grossman, Nir Hacohen, David Hildeman, Jeffrey Hodgins,
1191 Tiffany Hwang, Mariko Ishimori, Ken Kalunian, Diane Kamen, Matthias Kretzler, Holden
1192 Maecker, Rong Mao, Maureen McMahon, Fernanda Payan-Schober, Michelle Petri, Chaim
1193 Putterman, Daimon Simmons, Thomas Tuschl, David Wofsy, Steve Woodle, and Aaron Wyse.

1194

1195

1196 **Author contributions**

1197 L.G-P., K.D.D., D.T., A.C., G.S.F., M.M., I.S., A.B-A., A.M.M., A. Nerviani, F.R., C.P., L.B.H.,
1198 and D.H., recruited patients and obtained synovial tissues. L.W.M., S.M.G., H.P., V.M.H., A.F.,
1199 V.P.B., and J.H.A. contributed to the procurement and processing of samples and design of the
1200 AMP study. E.D., E.M.G., and B.F.B., performed histological assessment of tissues. D.W.,
1201 K.P.L., A.F., and V.P.B. curated and analyzed histologic and clinical data. W.A. provided project
1202 management and curated histologic and clinical data. K. Wei, A.H.J, G.F.M.W., A. Nathan, and
1203 M.B.B. designed and implemented the tissue disaggregation, cell sorting, and single-cell
1204 sequencing pipeline. A.H.J., K. Wei, and G.F.M.W supervised and executed the tissue
1205 disaggregation pipeline. F.Z., A. Nathan, N.M., Q.X., M.G-A., J.B.K, K. Weinand, J.M., L.R., and
1206 S.R. conducted computational and statistical analysis. A.H.J., K. Wei, M.B.B., J.H.A., L.T.D.,
1207 D.A.R., F.Z., A. Nathan, S.R., D.E.O., J.R-M., and A.F. provided input on cellular analysis and
1208 interpretation. D.E.O., J.R-M., A.F., and J.H.A. provided input on histologic analyses. N.M. and
1209 K.S. implemented the website. S.R., M.B.B., J.H.A., L.T.D., and D.A.R. supervised the research.

1210 F.Z., A.H.J., A. Nathan, N.M., Q.X., and S.R. wrote the initial draft. F.Z., A.H.J., A. Nathan, K.
1211 Wei, N.M., D.A.R, L.T.D. J.H.A, M.B.B., and S.R. edited the draft. AMP: RA/SLE Network
1212 members contributed to this work by managing patient recruitment, curating clinical data,
1213 obtaining and processing synovial tissue samples, managing biorepositories, conducting
1214 histological or computational analysis, providing software code, providing website support,
1215 and/or providing input on data analysis and interpretation. All authors participated in editing the
1216 final manuscript.

1217

1218

1219 **Competing interests**

1220 A.H.J. reports research support from Amgen, outside the submitted work. K.W. is a consultant
1221 for Mestag Therapeutics and Gilead Sciences and reports grant support from Gilead Sciences.
1222 S.M.G. reports research support from Novartis and is a consultant for UCB, outside the
1223 submitted work. V.M.H. is a co-founder of Q32 Bio and has previously received sponsored
1224 research from Janssen and been a consultant for Celgene and BMS, outside the submitted
1225 work. A.F. reports personal fees from Abbvie, Roche, and Janssen and grant support from
1226 Roche, UCB, Nascient, Mestag, GlaxoSmithKline, and Janssen, outside the submitted work.
1227 D.A.R. reports personal fees from Pfizer, Janssen, Merck, Scipher Medicine, GlaxoSmithKline,
1228 and Bristol-Myers Squibb and grant support from Janssen and Bristol-Myers Squibb, outside the
1229 submitted work. In addition, D.A.R. is a co-inventor on a patent submitted on T peripheral helper
1230 cells. M.B.B. is a founder for Mestag Therapeutics and a consultant for GlaxoSmithKline, 4FO
1231 Ventures, and Scailyte AG. S.R. is a founder for Mestag Therapeutics, a scientific advisor for
1232 Janssen and Pfizer, and a consultant for Gilead and Rheos Medicines.

1233

1234

1235 **Data availability**

1236 All raw and processed data will be available upon acceptance. A cell browser website will be
1237 available to visualize our data and results.

1238

1239

1240 **Code availability**

1241 All source code will be available on Github upon acceptance. Supplementary Information is
1242 available for this paper.

1243

1244

1245

1246

1247

1248

1249

1250

1251

1252

1253

1254

1255

1256

1257

1258

1259

1260 **References**

- 1261 1. Alamanos, Y., Voulgari, P. V. & Drosos, A. A. Incidence and prevalence of rheumatoid
1262 arthritis, based on the 1987 American College of Rheumatology criteria: a systematic
1263 review. *Semin. Arthritis Rheum.* **36**, 182–188 (2006).
- 1264 2. Smolen, J. S. *et al.* Rheumatoid arthritis. *Nature Reviews Disease Primers* **4**, 1–23 (2018).
- 1265 3. Koduri, G. *et al.* Interstitial lung disease has a poor prognosis in rheumatoid arthritis: results
1266 from an inception cohort. *Rheumatology* **49**, 1483–1489 (2010).
- 1267 4. McInnes, I. B. & Schett, G. The pathogenesis of rheumatoid arthritis. *N. Engl. J. Med.* **365**,
1268 2205–2219 (2011).
- 1269 5. Orr, C. *et al.* Synovial tissue research: a state-of-the-art review. *Nat. Rev. Rheumatol.* **13**,
1270 463–475 (2017).
- 1271 6. Kerrigan, S. A. & McInnes, I. B. Reflections on ‘older’ drugs: learning new lessons in
1272 rheumatology. *Nature Reviews Rheumatology* vol. 16 179–183 (2020).
- 1273 7. Nagy, G. & van Vollenhoven, R. F. Sustained biologic-free and drug-free remission in
1274 rheumatoid arthritis, where are we now? *Arthritis Res. Ther.* **17**, 181 (2015).
- 1275 8. Smolen, J. S. & Aletaha, D. Rheumatoid arthritis therapy reappraisal: strategies,
1276 opportunities and challenges. *Nat. Rev. Rheumatol.* **11**, 276–289 (2015).
- 1277 9. Alivernini, S., Laria, A., Gremese, E., Zoli, A. & Ferraccioli, G. ACR70-disease activity score
1278 remission achievement from switches between all the available biological agents in
1279 rheumatoid arthritis: a systematic review of the literature. *Arthritis Res. Ther.* **11**, R163
1280 (2009).
- 1281 10. Viatte, S. & Barton, A. Genetics of rheumatoid arthritis susceptibility, severity, and
1282 treatment response. *Semin. Immunopathol.* **39**, 395–408 (2017).
- 1283 11. Amariuta, T., Luo, Y., Knevel, R., Okada, Y. & Raychaudhuri, S. Advances in genetics
1284 toward identifying pathogenic cell states of rheumatoid arthritis. *Immunol. Rev.* **294**, 188–

- 1285 204 (2020).
- 1286 12. Terao, C. *et al.* Distinct HLA Associations with Rheumatoid Arthritis Subsets Defined by
1287 Serological Subphenotype. *Am. J. Hum. Genet.* **105**, 880 (2019).
- 1288 13. Pitzalis, C., Choy, E. H. S. & Buch, M. H. Transforming clinical trials in rheumatology:
1289 towards patient-centric precision medicine. *Nat. Rev. Rheumatol.* **16**, 590–599 (2020).
- 1290 14. Han, B. *et al.* Fine mapping seronegative and seropositive rheumatoid arthritis to shared
1291 and distinct HLA alleles by adjusting for the effects of heterogeneity. *Am. J. Hum. Genet.*
1292 **94**, 522–532 (2014).
- 1293 15. Oliver, J., Plant, D., Webster, A. P. & Barton, A. Genetic and genomic markers of anti-TNF
1294 treatment response in rheumatoid arthritis. *Biomark. Med.* **9**, 499–512 (2015).
- 1295 16. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis
1296 with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann.*
1297 *Rheum. Dis.* **79**, 685–699 (2020).
- 1298 17. Fraenkel, L. *et al.* 2021 American College of Rheumatology Guideline for the Treatment of
1299 Rheumatoid Arthritis. *Arthritis Care Res.* **73**, 924–939 (2021).
- 1300 18. Aletaha, D. & Smolen, J. S. Diagnosis and Management of Rheumatoid Arthritis: A Review.
1301 *JAMA* **320**, 1360–1372 (2018).
- 1302 19. Lewis, M. J. *et al.* Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and
1303 Treatment Response Phenotypes. *Cell Rep.* **28**, 2455–2470.e5 (2019).
- 1304 20. Humby, F. *et al.* Rituximab versus tocilizumab in anti-TNF inadequate responder patients
1305 with rheumatoid arthritis (R4RA): 16-week outcomes of a stratified, biopsy-driven,
1306 multicentre, open-label, phase 4 randomised controlled trial. *Lancet* **397**, 305–317 (2021).
- 1307 21. Zhang, F. *et al.* Defining inflammatory cell states in rheumatoid arthritis joint synovial
1308 tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* (2019)
1309 doi:10.1038/s41590-019-0378-1.
- 1310 22. Kuo, D. *et al.* HBEGF+ macrophages in rheumatoid arthritis induce fibroblast invasiveness.

- 1311 *Sci. Transl. Med.* **11**, (2019).
- 1312 23. Zhang, F. *et al.* IFN- γ and TNF- α drive a CXCL10+ CCL2+ macrophage phenotype
1313 expanded in severe COVID-19 lungs and inflammatory diseases with tissue inflammation.
1314 *Genome Med.* **13**, 64 (2021).
- 1315 24. Rao, D. A. *et al.* Pathologically expanded peripheral T helper cell subset drives B cells in
1316 rheumatoid arthritis. *Nature* **542**, 110–114 (2017).
- 1317 25. Alivernini, S. *et al.* Distinct synovial tissue macrophage subsets regulate inflammation and
1318 remission in rheumatoid arthritis. *Nat. Med.* **26**, 1295–1306 (2020).
- 1319 26. Wang, Y. *et al.* Rheumatoid arthritis patients display B-cell dysregulation already in the
1320 naïve repertoire consistent with defects in B-cell tolerance. *Sci. Rep.* **9**, 1–13 (2019).
- 1321 27. Wei, K. *et al.* Notch signalling drives synovial fibroblast identity and arthritis pathology.
1322 *Nature* **582**, 259–264 (2020).
- 1323 28. Bocharnikov, A. V. *et al.* PD-1hiCXCR5- T peripheral helper cells promote B cell responses
1324 in lupus via MAF and IL-21. *JCI Insight* **4**, (2019).
- 1325 29. Christophersen, A. *et al.* Distinct phenotype of CD4+ T cells driving celiac disease identified
1326 in multiple autoimmune conditions. *Nat. Med.* **25**, 734–737 (2019).
- 1327 30. Ekman, I. *et al.* Circulating CXCR5-PD-1hi peripheral T helper cells are associated with
1328 progression to type 1 diabetes. *Diabetologia* **62**, 1681–1688 (2019).
- 1329 31. Martin, J. C. *et al.* Single-Cell Analysis of Crohn’s Disease Lesions Identifies a Pathogenic
1330 Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493–1508.e20
1331 (2019).
- 1332 32. Na, Y. R., Stakenborg, M., Seok, S. H. & Matteoli, G. Macrophages in intestinal
1333 inflammation and resolution: a potential therapeutic target in IBD. *Nat. Rev. Gastroenterol.*
1334 *Hepatol.* **16**, 531–543 (2019).
- 1335 33. Kochi, Y. Genetics of autoimmune diseases: perspectives from genome-wide association
1336 studies. *Int. Immunol.* **28**, 155–161 (2016).

- 1337 34. Matzaraki, V., Kumar, V., Wijmenga, C. & Zernakova, A. The MHC locus and genetic
1338 susceptibility to autoimmune and infectious diseases. *Genome Biol.* **18**, 76 (2017).
- 1339 35. Krenn, V. *et al.* Grading of chronic synovitis--a histopathological grading system for
1340 molecular and diagnostic pathology. *Pathol. Res. Pract.* **198**, 317–325 (2002).
- 1341 36. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
1342 Harmony. *Nat. Methods* (2019) doi:10.1038/s41592-019-0619-0.
- 1343 37. Kang, J. B. *et al.* Efficient and precise single-cell reference atlas mapping with Symphony.
1344 *Nat. Commun.* **12**, 1–21 (2021).
- 1345 38. Ehrenstein, M. R. *et al.* Compromised function of regulatory T cells in rheumatoid arthritis
1346 and reversal by anti-TNFalpha therapy. *J. Exp. Med.* **200**, 277–285 (2004).
- 1347 39. Dominguez-Villar, M., Baecher-Allan, C. M. & Hafler, D. A. Identification of T helper type 1-
1348 like, Foxp3+ regulatory T cells in human autoimmune disease. *Nat. Med.* **17**, 673–675
1349 (2011).
- 1350 40. MacDonald, K. G. *et al.* Regulatory T cells produce profibrotic cytokines in the skin of
1351 patients with systemic sclerosis. *J. Allergy Clin. Immunol.* **135**, 946–955.e9 (2015).
- 1352 41. McClymont, S. A. *et al.* Plasticity of human regulatory T cells in healthy subjects and
1353 patients with type 1 diabetes. *J. Immunol.* **186**, 3918–3926 (2011).
- 1354 42. Johnson, J. L. *et al.* The Transcription Factor T-bet Resolves Memory B Cell Subsets with
1355 Distinct Tissue Distributions and Antibody Specificities in Mice and Humans. *Immunity* **52**,
1356 842–855.e6 (2020).
- 1357 43. Wang, S. *et al.* IL-21 drives expansion and plasma cell differentiation of autoreactive
1358 CD11c^{hi} T-bet⁺ B cells in SLE. *Nat. Commun.* **9**, 1758 (2018).
- 1359 44. Jenks, S. A. *et al.* Distinct Effector B Cells Induced by Unregulated Toll-like Receptor 7
1360 Contribute to Pathogenic Responses in Systemic Lupus Erythematosus. *Immunity* **49**, 725–
1361 739.e6 (2018).
- 1362 45. Yuseff, M.-I. *et al.* Polarized secretion of lysosomes at the B cell synapse couples antigen

- 1363 extraction to processing and presentation. *Immunity* **35**, 361–374 (2011).
- 1364 46. Rubtsov, A. V. *et al.* CD11c-Expressing B Cells Are Located at the T Cell/B Cell Border in
1365 Spleen and Are Potent APCs. *J. Immunol.* **195**, 71–79 (2015).
- 1366 47. Radtke, D. & Bannard, O. Expression of the Plasma Cell Transcriptional Regulator Blimp-1
1367 by Dark Zone Germinal Center B Cells During Periods of Proliferation. *Front. Immunol.* **9**,
1368 3106 (2018).
- 1369 48. Palm, A.-K. E. & Kleinau, S. Marginal zone B cells: From housekeeping function to
1370 autoimmunity? *J. Autoimmun.* **119**, 102627 (2021).
- 1371 49. Weller, S. *et al.* Human blood IgM ‘memory’ B cells are circulating splenic marginal zone B
1372 cells harboring a prediversified immunoglobulin repertoire. *Blood* **104**, 3647–3654 (2004).
- 1373 50. Tull, T. J. *et al.* Human marginal zone B cell development from early T2 progenitors. *J. Exp.*
1374 *Med.* **218**, (2021).
- 1375 51. Descatoire, M. *et al.* Identification of a human splenic marginal zone B cell precursor with
1376 NOTCH2-dependent differentiation properties. *J. Exp. Med.* **211**, 987–1000 (2014).
- 1377 52. Weyand, C. M. & Goronzy, J. J. Ectopic germinal center formation in rheumatoid synovitis.
1378 *Ann. N. Y. Acad. Sci.* **987**, 140–149 (2003).
- 1379 53. Schröder, A. E., Greiner, A., Seyfert, C. & Berek, C. Differentiation of B cells in the
1380 nonlymphoid tissue of the synovial membrane of patients with rheumatoid arthritis. *Proc.*
1381 *Natl. Acad. Sci. U. S. A.* **93**, 221–225 (1996).
- 1382 54. Dogra, P. *et al.* Tissue Determinants of Human NK Cell Development, Function, and
1383 Residence. *Cell* **180**, 749–763.e13 (2020).
- 1384 55. Spits, H. *et al.* Innate lymphoid cells--a proposal for uniform nomenclature. *Nat. Rev.*
1385 *Immunol.* **13**, 145–149 (2013).
- 1386 56. Ebbo, M., Crinier, A., Vély, F. & Vivier, E. Innate lymphoid cells: major players in
1387 inflammatory diseases. *Nat. Rev. Immunol.* **17**, 665–678 (2017).
- 1388 57. Cella, M. *et al.* Subsets of ILC3-ILC1-like cells generate a diversity spectrum of innate

- 1389 lymphoid cells in human mucosal tissues. *Nat. Immunol.* **20**, 980–991 (2019).
- 1390 58. Gordon, S. Phagocytosis: An Immunobiologic Process. *Immunity* **44**, 463–475 (2016).
- 1391 59. Jakubzick, C. V., Randolph, G. J. & Henson, P. M. Monocyte differentiation and antigen-
1392 presenting functions. *Nat. Rev. Immunol.* **17**, 349–362 (2017).
- 1393 60. Lim, H. Y. *et al.* Hyaluronan Receptor LYVE-1-Expressing Macrophages Maintain Arterial
1394 Tone through Hyaluronan-Mediated Regulation of Smooth Muscle Cell Collagen. *Immunity*
1395 **49**, 326–341.e7 (2018).
- 1396 61. Mulder, K. *et al.* Cross-tissue single-cell landscape of human monocytes and macrophages
1397 in health and disease. *Immunity* **54**, 1883–1900.e5 (2021).
- 1398 62. Liao, M. *et al.* Single-cell landscape of bronchoalveolar immune cells in patients with
1399 COVID-19. *Nat. Med.* (2020) doi:10.1038/s41591-020-0901-9.
- 1400 63. Schuch, K. *et al.* Osteopontin affects macrophage polarization promoting endocytic but not
1401 inflammatory properties. *Obesity* **24**, 1489–1498 (2016).
- 1402 64. Remmerie, A. *et al.* Osteopontin Expression Identifies a Subset of Recruited Macrophages
1403 Distinct from Kupffer Cells in the Fatty Liver. *Immunity* **53**, 641–657.e14 (2020).
- 1404 65. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells,
1405 monocytes, and progenitors. *Science* **356**, (2017).
- 1406 66. Zhang, Q. *et al.* Landscape and Dynamics of Single Immune Cells in Hepatocellular
1407 Carcinoma. *Cell* **179**, 829–845.e20 (2019).
- 1408 67. Korsunsky, I. *et al.* Cross-tissue, single-cell stromal atlas identifies shared pathological
1409 fibroblast phenotypes in four chronic inflammatory diseases. *bioRxiv* 2021.01.11.426253
1410 (2021) doi:10.1101/2021.01.11.426253.
- 1411 68. Mizoguchi, F. *et al.* Functionally distinct disease-associated fibroblast subsets in
1412 rheumatoid arthritis. *Nat. Commun.* **9**, 789 (2018).
- 1413 69. Buechler, M. B. *et al.* Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–
1414 579 (2021).

- 1415 70. Yazici, Y. *et al.* Efficacy of tocilizumab in patients with moderate to severe active
1416 rheumatoid arthritis and a previous inadequate response to disease-modifying
1417 antirheumatic drugs: the ROSE study. *Ann. Rheum. Dis.* **71**, 198–205 (2012).
- 1418 71. Genovese, M. C. *et al.* Interleukin-6 receptor inhibition with tocilizumab reduces disease
1419 activity in rheumatoid arthritis with inadequate response to disease-modifying antirheumatic
1420 drugs: the tocilizumab in combination with traditional disease-modifying antirheumatic drug
1421 therapy study. *Arthritis Rheum.* **58**, 2968–2980 (2008).
- 1422 72. Genovese, M. C. *et al.* Sarilumab Plus Methotrexate in Patients With Active Rheumatoid
1423 Arthritis and Inadequate Response to Methotrexate: Results of a Phase III Study. *Arthritis*
1424 *Rheumatol* **67**, 1424–1437 (2015).
- 1425 73. Wei, K., Nguyen, H. N. & Brenner, M. B. Fibroblast pathology in inflammatory diseases. *J.*
1426 *Clin. Invest.* **131**, (2021).
- 1427 74. Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**,
1428 764–779.e20 (2020).
- 1429 75. Wigle, J. T. *et al.* An essential role for Prox1 in the induction of the lymphatic endothelial
1430 cell phenotype. *EMBO J.* **21**, 1505–1513 (2002).
- 1431 76. Fujimoto, N. *et al.* Single-cell mapping reveals new markers and functions of lymphatic
1432 endothelial cells in lymph nodes. *PLoS Biol.* **18**, e3000704 (2020).
- 1433 77. Reshef, Y. A. *et al.* Co-varying neighborhood analysis identifies cell populations associated
1434 with phenotypes of interest from single-cell transcriptomics. *Nat. Biotechnol.* 1–9 (2021).
- 1435 78. Crotty, S. T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. *Immunity*
1436 **50**, 1132–1148 (2019).
- 1437 79. Dennis, G., Jr *et al.* Synovial phenotypes in rheumatoid arthritis correlate with response to
1438 biologic therapeutics. *Arthritis Res. Ther.* **16**, R90 (2014).
- 1439 80. Humby, F. *et al.* Synovial cellular and molecular signatures stratify clinical response to
1440 csDMARD therapy and predict radiographic progression in early rheumatoid arthritis

- 1441 patients. *Ann. Rheum. Dis.* **78**, 761–772 (2019).
- 1442 81. Pitzalis, C., Kelly, S. & Humby, F. New learnings on the pathophysiology of RA from
1443 synovial biopsies. *Curr. Opin. Rheumatol.* **25**, 334–344 (2013).
- 1444 82. Rao, D. A. T Cells That Help B Cells in Chronically Inflamed Tissues. *Front. Immunol.* **9**,
1445 1924 (2018).
- 1446 83. Seth, A. & Craft, J. Spatial and functional heterogeneity of follicular helper T cells in
1447 autoimmunity. *Curr. Opin. Immunol.* **61**, 1–9 (2019).
- 1448 84. Kroot, E. J. *et al.* The prognostic value of anti-cyclic citrullinated peptide antibody in patients
1449 with recent-onset rheumatoid arthritis. *Arthritis Rheum.* **43**, 1831–1835 (2000).
- 1450 85. Jansen, L. M. A. *et al.* The predictive value of anti-cyclic citrullinated peptide antibodies in
1451 early arthritis. *J. Rheumatol.* **30**, 1691–1695 (2003).
- 1452 86. Folkman, J. & D'Amore, P. A. Blood vessel formation: what is its molecular basis? *Cell* **87**,
1453 1153–1155 (1996).
- 1454 87. Holmes, A. B. *et al.* Single-cell analysis of germinal-center B cells informs on lymphoma cell
1455 of origin and outcome. *J. Exp. Med.* **217**, (2020).
- 1456 88. Milpied, P. *et al.* Human germinal center transcriptional programs are de-synchronized in B
1457 cell lymphoma. *Nat. Immunol.* **19**, 1013–1024 (2018).
- 1458 89. Glass, D. R. *et al.* An Integrated Multi-omic Single-Cell Atlas of Human B Cell Identity.
1459 *Immunity* **53**, 217–232.e5 (2020).
- 1460 90. Nocturne, G. & Mariette, X. B cells in the pathogenesis of primary Sjögren syndrome. *Nat.*
1461 *Rev. Rheumatol.* **14**, 133–145 (2018).
- 1462 91. Prevoo, M. L. *et al.* Modified disease activity scores that include twenty-eight-joint counts.
1463 Development and validation in a prospective longitudinal study of patients with rheumatoid
1464 arthritis. *Arthritis Rheum.* **38**, 44–48 (1995).
- 1465 92. Wells, G. *et al.* Validation of the 28-joint Disease Activity Score (DAS28) and European
1466 League Against Rheumatism response criteria based on C-reactive protein against disease

- 1467 progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on
1468 erythrocyte sedimentation rate. *Ann. Rheum. Dis.* **68**, 954–960 (2009).
- 1469 93. Choi, I. Y. *et al.* Stromal cell markers are differentially expressed in the synovial tissue of
1470 patients with early arthritis. *PLoS One* **12**, e0182751 (2017).
- 1471 94. Krenn, V. *et al.* Synovitis score: discrimination between chronic low-grade and high-grade
1472 synovitis. *Histopathology* **49**, 358–364 (2006).
- 1473 95. Donlin, L. T. *et al.* Methods for high-dimensional analysis of cells dissociated from
1474 cryopreserved synovial tissue. *Arthritis Res. Ther.* **20**, 139 (2018).
- 1475 96. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell
1476 Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291.e9 (2019).
- 1477 97. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21
1478 (2019).
- 1479 98. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells.
1480 *Nat. Methods* **14**, 865–868 (2017).
- 1481 99. Baglama, J. & Reichel, L. Augmented implicitly restarted lanczos bidiagonalization
1482 methods. *SIAM J. Sci. Comput.* **27**, 19–42 (2005).
- 1483 100. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection
1484 for Dimension Reduction. *arXiv [stat.ML]* (2018).
- 1485 101. Nathan, A. *et al.* Multimodally profiling memory T cells from a tuberculosis cohort identifies
1486 cell state associations with demographics, environment and disease. *Nat. Immunol.* **22**,
1487 781–793 (2021).
- 1488 102. González, I., Déjean, S., Martin, P. & Baccini, A. CCA: An R Package to Extend Canonical
1489 Correlation Analysis. *Journal of Statistical Software, Articles* **23**, 1–14 (2008).
- 1490 103. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities
1491 in large networks. *arXiv [physics.soc-ph]* (2008).
- 1492 104. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using

- 1493 lme4. *Journal of Statistical Software, Articles* **67**, 1–48 (2015).
- 1494 105.Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference
1495 resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–62 (2016).
- 1496 106.Fonseka, C. Y. *et al.* Mixed-effects association of single cells identifies an expanded
1497 effector CD4+ T cell subset in rheumatoid arthritis. *Sci. Transl. Med.* **10**, (2018).