Integrative single-cell analysis of epigenomic and transcriptomic states in patients with systemic sclerosis-associated interstitial lung disease

Shifang Li, Meijiao Gong

Laboratory of Immunology and Vaccinology, FARAH, ULiège, Liège 4000, Belgium.

SF and MJ share senior authorship.

Correspondence:

Shifang Li, fruceslee@gmail.com

Laboratory of Immunology and Vaccinology, FARAH, ULiège, Liège 4000, Belgium

Abstract

Interstitial lung disease is the primary cause of death in individuals who have systemic sclerosis, one of the autoimmune connective tissue diseases. Understanding the pathophysiology of the disease is crucial to developing treatment options. Here, we performed a single-cell multi-omic analysis on lung tissue samples from patients with systemic sclerosis-associated interstitial lung disease (SSC-ILD), profiling chromatin accessibility and gene expression in the same samples and discovering significant cellular heterogeneity. Systemic-venous endothelial cells (ECs) have been shown to be pro-inflammatory and highly active. In addition, it was shown that the transcription factor FOSL2 targets the genes involved in response to unfolded proteins in systemic-venous ECs. Furthermore, we prioritized functional risk variants for systemic sclerosis using a genome-wide association study. Ligand-receptor analysis revealed that ECs significantly increased interaction with B cells via CXCL10-CXCR3 in patients with SSC-ILD. Overall, our analysis emphasizes epigenetic and transcriptional patterns in systemic-venous ECs, which might be beneficial in understanding the pathogenesis of SSC-ILD.

Introduction

Systemic sclerosis (SSC) is a multi-organ disease with clinical symptoms such as skin and internal organ fibrosis and vascular abnormalities. Because of limited therapy choices, SSC causes significant morbidity and mortality. One of the primary causes of SSC-related death is interstitial lung disease (ILD) [1]. However, the

processes underlying fibrosis in SSC-associated interstitial lung disease (SSC-ILD) remain largely unknown. According to research, epithelial cell damage, activation of innate and adaptive immunity, and recruitment and activation of fibroblasts may all contribute to the excessive synthesis of extracellular matrix and scar formation in SSC-ILD, similar to the mechanisms in other fibrotic lung diseases [2]. Furthermore, in SSC-ILD, myofibroblasts have been demonstrated to be essential effector cells for extracellular matrix remodeling [2-3]. Recent research has given more insight into the role of SPP1 macrophages in pro-fibrosis in SSC-ILD [4]. However, the role of the diversity of other cellular subpopulations found in healthy and SSC-ILD lungs in shaping SSC-ILD pathophysiology is unclear.

Single-cell sequencing, which includes single-cell RNA sequencing (scRNA-seq) and the assay for transposase-accessible chromatin sequencing (scATAC-seq), is a valuable method for investigating the molecular and cellular heterogeneity of cell types. EC, one of the innermost cell types lined along vessels in all organs and tissues through hosts, is also in the bloodstream, where it can modulate inflammation by regulating immune cell transport, activation status, and function [5-7]. Furthermore, EC in the lungs and liver not only promotes immune homeostasis but also mediates the balance between tolerance and inflammation [6]. However, the role of endothelial cells (ECs) in lung fibrosis is not well-defined. A previous study showed that ECs to mesenchymal phenotypic change was observed in the lung tissues of SSC-ILD patients [8]. A recent study found that ECs in the lungs regulate pulmonary fibrosis via FOXF1/R-Ras signaling [9]. However, our understanding of EC heterogeneity in SSC-ILD at the single-cell level remains limited, particularly in the context of transcriptome and chromatin accessibility. Here, scRNA-seq and scATAC-seq methods were employed to investigate the role of ECs involved in SSC-ILD in order to better understand cellular heterogeneity at the single-cell multiomics level. Furthermore, we used the scATAC-seq with SSC genome-wide association studies (GWAS) loci to elucidate mechanisms for functional GWAS loci.

Methods

Data download

scRNA-seq for 11 individuals and matched scATAC-seq for 5 individuals were conducted in the recent study [4]. Therefore, all these raw fastq files were downloaded from the Gene Expression Omnibus (accession number: GSE212109)

scATAC-seq processing

scATAC-seq data from the scATAC-seq experiment were aligned to the GRCh19 (hg19) reference genome and quantified using the Cellranger count tool (10× Genomics, v.2.0). ArchR v1.0.2 [10] was used to construct arrow files by reading in the available read fragments for each sample and applying the default augments. We selected cells with less than 1000 unique fragments and enrichment at TSSs below 4 to ensure that each cell had a strong signal and was well-sequenced. Meanwhile, doublets were inferred and filtered by ArchR. The method scOpen (v.1.0.0) [11] was used to construct a low-dimensional matrix of the cells for dimensionality reduction. To correct the batch effects (sample and chemistry) and integrate the data, the Harmony (v.0.1.1) algorithm [12] was used, and cells were clustered using the Leiden algorithm with a resolution of 1, respectively. A gene activity score matrix was built using the method "addGeneScoreMatrix" to annotate the clusters and marker genes were discovered for each cluster using the function "getMarkerFeatures". We also utilized the MAGIC [13] imputed weight approach on the obtained gene activity scores to reduce noise in the data scarcity in scATAC-seq. The clusters were annotated using the same markers from the scRNA-seq data. Clustering with "FindClusters" with a default resolution of 0.5 during the reclustering process to better find small clusters. In two-dimensional space, all data were represented using uniform manifold approximation and projection (UMAP). MACS2 was utilized to accomplish peak calling [14], and the peak matrix was then generated using "addReproduciblePeakSet" and these pseudobulk duplicates. Non-overlapping pseudobulk replicates were constructed from groups of cells using the "addGroupCoverages" function with distinct inputs for differential comparisons of cell types and clinical stages. The pseudobulk peak set was subjected to motif enrichment and motif deviation studies. The chromVAR [15] deviation scores for these motifs were also estimated using the ArchR implementation. Using the "getMarkerFeatures" function, the pseudobulked

91 peak set was employed for differential analysis between different cell types and
92 disease states. To calculate the *p*-value and false discovery rate (FDR) between any
93 pair of samples, the Wilcoxon test and the Benjamini-Hochberg (BH) multiple tests
94 were utilized. FDR<=0.01 and log2-fold change (FC)>=1 were used to determine
95 differentially accessible distal peaks. Meanwhile, we used the ChIPseeker package's
96 "annotatePeak" function [16] to annotate the closest genes in the peak region using
97 default inputs. Tn5 insertions in genome-wide motifs were measured and normalized
98 by eliminating the Tn5 bias from the footprinting signal. We normalized these
99 footprints using mean values 200-250 from the motif center. To identify peak-to-gene
100 links, we used the ArchR "addPeak2GeneLinks" function and set the parameters to
101 "corCutOff=0.45" and "resolution=10,000", with "reducedDims" as the
102 dimensionality reduction results after batch correction.

103 **Overlap of SSC GWAS SNPs with cell-type peaks**

104 We downloaded summary statistics from López-Isac E *et al* [17] who performed
105 GWAS meta-analysis in 26,679 European populations involving 9,095 cases and
106 17,584 controls for comparison with GWASs of SSC via GWAS catalog
107 (https://www.ebi.ac.uk/gwas) (accession codes GCST009131). The same strategy as
108 described by Turner *et al* was used to detect overlap between disease-related SNPs
109 and cell-type peaks [18]. In brief, the lead variants were utilized as input to PLINK
110 (version 1.9) [19] first, and then kept variants with r^2>0.8 utilizing the European
111 population. To accommodate for SSC variants located close to a peak, a 100-bp
112 window with the SNP in the center was evaluated by extending the SNP location 50
113 bp in each direction using BEDOPS (v.2.4.37) [20]. Finally, bedtools (v.2.26.0) [21]
114 was used to intersect these SNP with each cell type peak.

115 **scRNA-seq processing**

116 For each sample, the raw fastq files were processed and aligned to the GRCh38
117 reference using CellRanger software with the default parameters
118 (https://support.10×genomics.com, version 7.1.0). Following that, the Seurat R
119 package v4.2.0 [22] was used for data quality control, preprocessing, and dimensional
120 reduction analysis, as previously described [23]. After generating gene-cell data

matrices from combined datasets, matrices were merged, and poor-quality cells with 200 or more expressed genes and mitochondrial gene percentages greater than 50 were excluded, leaving 48,622 cells for further analysis. Doublets were detected using the Scrublet package (version 2.0.2) using the "scrub.scrub_doublets()" function [24] and removed by Seurat. After normalizing the data using log transformation and scaling, the Harmony R package v1.0 was used to correct batch effects (sample and chemistry). For downstream UMAP visualization and clustering, the top 30 dimensions of Harmony embeddings were used. Cells were clustered using the Louvain algorithm with a resolution of 0.2, yielding 16 unique cell clusters. We utilized the Seurat package's "FindAllMarkers" function with the default parameters for scRNA-seq differential expression analysis. A BH-adjusted *p*-value of less than 0.05 was regarded as statistically significant. CellTypist [25] was used to determine the cell type annotations for each cluster based on the expression patterns of known marker genes.

Multiomics data processing

An anchoring strategy as implemented in the Archr packages was employed to carry out an integrated analysis of scATAC-seq and scRNA-seq data. Briefly, a collection of "anchors" between datasets presumed to be in a comparable biological state was first found (the "FindTransferAnchors" function) using canonical-correlation analysis (CCA). These "anchors" are then utilized as a reference to integrate the datasets, and the cluster labels from the scRNA-seq data can be projected onto the scATAC-seq data using the "TransferData" function. To assess the agreement between annotated and predicted cell labels in scATAC-seq data, an adjusted rand index was computed.

Pseudotime and RNA velocity analysis

In the scATAC-seq and integrated datasets, we created a differentiation trajectory and aligned single cells along it. Using batch-corrected Harmony embeddings, a cellular trajectory was built in a low-dimension space based on the trajectory backbone. The cells were assigned based on pseudotime value estimates, and differential feature z-scores were used to generate a heatmap. For pseudotime

151 trajectory analysis in scRNA-seq, R packages Monocle3 (version 1.3.1) [26] were
152 utilized with their default parameters. The root point was determined as "EC venous
153 systemic". For RNA velocity analysis, CellRanger produced BAM files in conjunction
154 with velocyto were used to generate a loom file containing the quantification of
155 spliced and unspliced RNA [27]. Following that, we concatenated samples and
156 computed RNA velocity using the packages scvelo with the default settings, as
157 previously described [23].

158 **Receptor-ligand pair analysis**

159 The Python software CellPhoneDB [28] with database version 4 was used to
160 infer cell-cell communication networks from single-cell transcriptome data to
161 measure cellular crosstalk across various cell types. Only receptors and ligands
162 expressed in more than 5% of the cells in a given cell type were taken into account.
163 We retained only statistically significant averages ($p<0.05$) of interaction partners
164 from the CellPhoneDB output, which served as a proxy for mean expression of both
165 the ligand and the receptor of a given anticipated relationship.

166 **Identifying TF target genes**

167 To identify significantly shared TFs and their directly regulated target genes in
168 SSC-ILD disease, we used the method described by Kuppe *et al* [29], with minor
169 modifications. In brief, following pseudotime analysis, the transcription factors and
170 RNA expression along the trajectory were sorted and a pseudotime label for each
171 transcription factor/gene was assigned. To associate the selected TF with targets, we
172 explored the correlation of peak accessibility and gene expression to identify
173 peak-to-gene links and further subset these links by those containing the TF motif.
174 Significantly correlated links (FDR<0.0001) with a positive correlation were selected.
175 To build a quantitative transcription factor-gene-regulatory network, we estimated the
176 correlation of the transcription factor binding activity between scATAC-seq and target
177 gene expression from scRNA-seq data, and only those interactions with a Pearson
178 correlation >0.8 were considered.

179 **Results and Discussion**

180 **Single-cell landscape in systemic sclerosis-associated interstitial lung disease**

181 A total of 48,622 cells were investigated using scRNA-seq data (**Supplemental**
182 **Figure 1A**), including 18,591 cells from 6 healthy individuals and 30,031 cells from 5
183 SSC-ILD patients, while 2 samples from SSC-ILD and the remaining 3 samples were
184 included in scATAC-seq (**Supplemental Figure 1B**). Following that, batch
185 adjustment for samples and chemistry was conducted using the R package "Harmony"
186 on both the scATAC-seq and scRNA-seq datasets. Based on the expression of
187 lineage-specific markers, 15 main cell types in the lung were identified in scRNA-seq
188 (**Figure 1A-1C**). In detail, macrophages, monocytes, T cells, monocyte-derived Mph,
189 multiciliated, fibroblasts, ECs, basal resting, NK cells, alveolar type 2 epithelial cells
190 (AT2), mast cells, smooth muscle, AT1, B cells, lymphatic EC mature, and an
191 unidentified cell population were reported (**Figure 1C**). We obtained a total of 14,717
192 high-quality cells after filtering for scATAC-seq and identified 13 clusters in ArchR
193 (**Figure 1D-1E**). 12 cell types were annotated based on gene activity scores provided
194 to each cluster (**Supplemental Figure 1C**). Label transfer was accomplished by
195 generating a gene-activity matrix from scATAC-seq data, which is a measurement of
196 chromatin accessibility inside the gene body and promoter of protein-coding genes.
197 Transfer anchors have been identified between the "reference" scRNA-seq dataset and
198 the "query" gene activity matrix, and anticipated cell types were assigned. When
199 comparing the annotated label using the scRNA-seq dataset, the 12 cell types
200 identified for scATAC-seq showed high consistency, demonstrating that scATAC-seq
201 is equivalent to scRNA-seq in the detection and assignment of cell identities (**Figure**
202 **1F**). Following that, the MACS2 was used to evaluate the differences in chromatin
203 accessibility (DAR) between cell types. A master set of 50,229 peaks was generated
204 by using a log2-fold-change threshold of 1 and adjusted *p*-values to measure DAR
205 significance (FDR<=0.01). Furthermore, 26% of DAR were found at a promoter
206 region less than 3 kb from the nearest transcriptional start point (**Figure 1G**). Intronic
207 was the second most common site, and DAR distribution was relatively consistent
208 across cell types. Meanwhile, motif enrichment analysis for these marker peaks was
209 undertaken to identify the TFs likely driving the regulatory profiles in each cell type
210 (**Figure 1H**). Top enriched motifs in T cells (ERG, ELK4, and ERF) were discovered.

211 Furthermore, SOX13 motifs were shown to be enriched in ECs, as well as ZNF238
212 and ZBTB42 motifs in fibroblasts. Overall, the integrated dataset allows us to detect
213 cell type heterogeneity at the transcriptional and chromatin accessibility levels
214 (**Figure 1I**).

215 **Heterogeneity of endothelial cell subset**

216 Several investigations have demonstrated that ECs have a role in the
217 pathophysiology of SSC [30-31]. The chromatin accessibility of ECs from control and
218 SSC-ILD lungs, however, remains unclear. To investigate the heterogeneity of ECs
219 subsets in SSC-ILD, we first examined the differences in ECs subsets between HC
220 and SSC-ILD patients in scRNA-seq including 11 individuals due to the higher
221 statistical power. Following the re-clustered ECs (4,492 cells), 5 cell subpopulations
222 were identified based on particular gene expression (**Figure 2A** and **Supplemental**
223 **Figure 2A-2C**). Interestingly, our study identified two cell subpopulations of ECs
224 with unequal frequencies in SSC-ILD patients and healthy individuals ($p<0.01$)
225 (**Figure 2B** and **Supplemental Figure 2D**). The frequency of SSC-ILD was found to
226 be higher in EC venous systemic, which substantially expressed RBP7; however, the
227 frequency of SSC-ILD was significantly lower in EC aerocyte capillary, which highly
228 expressed TBX2. Our findings from RNA velocity and pseudotime analysis utilizing
229 scATAC-seq revealed the differentiation of EC venous systemic to EC aerocyte
230 capillary, indicating that EC venous systemic could serve as a precursor for EC
231 aerocyte capillary (**Supplemental Figure 2E-2F**).

232 The integrated analysis of scRNA-seq and scATAC-seq was performed to better
233 understand the chromatin structure of ECs subpopulations (**Figure 2C-2D**). Label
234 transfer from scRNA-seq to scATAC-seq revealed the three cell types, with a
235 consistent lower frequency for EC venous systemic and a greater frequency for EC
236 aerocyte capillary, as expected (**Figure 2C-2D** and **Supplemental Figure 2G-2H**). A
237 pseudotime analysis was undertaken to uncover the dynamics and modulation of cell
238 fate decisions in EC subpopulations (**Figure 2E**). A lineage route from EC venous
239 systemic to EC aerocyte capillary was discovered by pseudotime analysis utilizing
240 scATAC-seq. The examination of *cis*-elements near certain genes and the accessibility

241 of TF motifs revealed distinct regulatory patterns of accessibility over the pseudotime
242 (**Figure 2F-2H**). Interestingly, *cis*-element accessibility in some loci involved in
243 inflammatory response, such as IL1R1 and SELE, as well as response to unfolded
244 protein, such as SERPINH1 and DNAJB1, gradually decreased from EC venous
245 systemic to EC aerocyte capillary, demonstrating that the expression of inflammatory
246 response-related genes was accurately regulated by a distinct *cis*-element network in
247 the different states (**Figure 2F-2G**). The distinct TF activities across the pseudotime
248 of EC venous systemic to EC aerocyte capillary paths were revealed by integrating
249 chromVAR TF-activity and TF RNA expression studies (**Figure 2H**). For example, by
250 assigning a path of accessibility from differentiated EC venous systemic to EC
251 aerocyte capillary, we identified enriched MEOX1 motifs at the start of the trajectory,
252 followed by enrichment of FOS, FOSL2, and JUNB motifs in EC aerocyte capillary,
253 and finally PBX1, KLF11, and KLF4 motifs. According to the trajectory analysis, the
254 EC venous systemic showed significantly higher occupancy in FOSL2 in the footprint
255 analysis than the EC aerocyte capillary, while the EC aerocyte capillary showed a
256 more pronounced DNA occupancy of KLF4, implying a role for FOSL2 and KLF4 as
257 distinct transcription factors in shaping EC maturation (**Figure 2I**). We used a
258 recently described approach to identify probable TF target genes based on the merged
259 scATAC-seq and scRNA-seq datasets to demonstrate the TFs regulatory program
260 network in the ECs [28]. Only Pearson correlations greater than 0.8 between
261 transcription factor binding activity from scATAC-seq and target gene expression
262 from scRNA-seq data were evaluated. Using this method, we discovered 456 genes
263 controlled by KLF4 and 179 genes regulated by FOSL2 (**Figure 2J**). For example,
264 four genes implicated in the response to unfolded protein, namely HSPA8,
265 HSP90AB1, DNAJA1, and HSPB1, were shown to be regulated by FOSL2, which
266 likewise showed decreased activity along the pseudotime axis (**Figure 2K**).
267 Furthermore, our analysis found that KLF4 regulates the expression of genes involved
268 in immune response regulation, such as BTNL9, ITGA4, and ICAM2, in EC aerocyte
269 capillaries. In the pseudotime analysis, we found increased KLF4 binding activity and
270 SITGA4 expression, supporting the significance of KLF4 as a possible activator of

271 immune response activation (**Figure 2K**). The chromVAR motif variations in the EC
272 venous systemic and EC aerocyte capillary between HC and SSC-ILD patients were
273 also compared (**Figure 2L**). AP-1 factors, including FOS and JUNB, were shown to
274 be abundant in both EC venous systemic and EC aerocyte capillary of SSC-ILD
275 patients. Furthermore, transcription factors including as GATA2 and GATA3 that are
276 involved in the negative regulation of ECs apoptosis were shown to be abundant in
277 the EC venous systemic of healthy individuals. Overall, our analysis indicated
278 epigenetic reprogramming in the formation of ECs, with activation characteristics and
279 compositional and epigenomic abnormalities in SSC-ILD patients' EC aerocyte
280 capillaries.

281 **Annotation of target cell types at SSC GWAS loci**

282 Indeed, the paired scATAC-seq/scRNA-seq datasets enabled us to deconstruct
283 the mechanisms underlying previously reported causative risk variations from
284 GWASs and to uncover disease-relevant cell types associated with these loci. Several
285 studies have found that noncoding GWAS variants are concentrated in CREs and
286 frequently function in cell-type-specific ways [32-34]. In this study, a multi-tiered
287 strategy was employed to prioritize putative functional SSC GWAS variants, as
288 described by Turner *et al* [18]. In brief, after identifying CRE variants, we compared
289 SSC lead variants (and variants in high LD (r^2>0.8; EUR)) from the most recent SSC
290 GWAS meta-analyses in a European cohort of 26,679 individuals with scATAC peaks
291 (**Figure 3A**). This identified a subset of variants (50 bp) that overlapped GWASs loci
292 and marker peaks, with the majority of SSC SNPs located within macrophage peaks,
293 followed by monocytes, T cells, AT2/basal resting cells, and B cells peaks (**Figure**
294 **3B**). We next identified target cell types for SSC regulatory variants based on these
295 overlaps. Several top potential SSC variants, for example, map to cell-type-specific
296 peaks, such as rs3792783 at the TNIP1 locus within an AT2 and basal resting cells
297 peak and rs1005714 at the NUP85 locus within macrophages (**Figure 3C**). Other SSC
298 variants, such as rs2305743 at IL12RB1 and rs1378942 at MIR4513, map to peaks
299 shared by T cells, monocytes, and macrophages. We reclustered AT2 using a
300 scRNA-seq dataset of 11 individuals to characterize its role in SSC-ILD, and four

301 clusters were identified (**Supplemental Figure 3A-3C**). Notably, the frequencies of
302 cluster 3, which highly expressed SERPINF1, were considerably greater in patients
303 with SSC-ILD (**Supplemental Figure 3D**). RNA velocity analysis suggests a C1 to
304 C3 lineage trajectory (**Supplemental Figure 3E**). Due to the small number of cells
305 observed in scRNA-seq for AT2 cluster 3, which may represent a rare cell population,
306 scATAC-seq did not disclose this subpopulation (data not shown). Furthermore,
307 because of the small number of samples included in the study, no substantial overlap
308 with GWAS loci was discovered for the cell-type peaks of fibroblasts, one of the main
309 effector cells in the extracellular matrix remodelling of the disease [3]. A larger
310 dataset should be used in the future to explore these subpopulations in depth.

311 Because noncoding variants cannot always regulate the adjacent gene, we used
312 co-accessibility and scRNA-seq integration to link candidate variants to target gene
313 expression (**Supplemental Figure 4A**). When all cell types were aggregated, this
314 orthogonal method yielded a total of 31,432 Peak2Gene linkages (**Supplemental**
315 **Figure 4B**). It is worth noting that the macrophages peak-containing variant,
316 rs1005714, has a co-accessibility and gene expression relationship with MRPS7, and
317 is also an MRPS7 lung-specific expression quantitative trait locus (eQTL) in the
318 Genotype-Tissue Expression project (GTEx) (**Figure 3E**). MRSPS7, but not NUP85,
319 is differently expressed between healthy individuals and SSC-ILD patients
320 (uncorrected p=2.9e-05) (**Supplemental Figure 4C**). These results suggested that
321 MRSPS7 could be a candidate gene associated with SSC, a large dataset as well as
322 experiments should be performed to validate these results.

323 **Cell-cell communication network**

324 To identify the reciprocal communication between different cell types in the lung,
325 we investigated the accumulated ligand/receptor interaction database CellPhoneDB to
326 detect reciprocal communication. The number of cell-cell interactions was increased
327 in SSC-ILD patients than in healthy individuals, as expected (**Figure 4A**). Fibroblasts,
328 in particular, demonstrated the greatest capacity for cell-cell interactions and shared
329 an increased number of projected connections with lymphatic EC mature, ECs, and
330 fibroblasts in patients with SSC-ILD. Furthermore, with the exception of

331 monocyte-derived Mph, ECs increased cell-cell interactions with practically all other
332 cell types in SSC-ILD patients. These findings further suggest that ECs may play an
333 essential role in disease progression.

334 In SSC-ILD patients, ECs had enhanced CXCL10-DPP4 and CXCL10-CXCR3
335 interactions with lymphatic EC mature and B cells (**Figure 4B**). According to gene
336 ontology enrichment analysis, CXCL10 was substantially enriched in the RIG-I-like
337 receptor signaling pathway and ECs activation (p.adj<0.01), whereas CXCL10 and
338 CXCR3 were significantly involved in the inflammatory response (p.adj=0.00081).
339 CXCL10 and its receptors CXCR3 were shown to be more accessible in SSC-ILD
340 patients, with increased RNA expression of CXCL10 and its receptor CXCR3 (**Figure**
341 **4C-4D**). Interestingly, TNFSF10-RIPK1 and TNF-FFAR2 signaling was anticipated
342 to be active during AT2 and monocyte interactions with other cell types in SSC-ILD
343 patients. Furthermore, SOSTDC1-LRP6 interactions were expected to be reduced in
344 the communication between EC cells with basal resting cells and AT2 in SSC-ILD
345 patients (**Figure 4B**). In fact, this has been identified as a critical regulatory
346 interaction in the negative control of the canonical Wnt signaling pathway [35].
347 Overall, our findings point to a complicated interaction network between healthy
348 individuals and patients. More research is needed to determine whether ECs are
349 involved in the regulation that leads to the onset of the inflammatory response in B
350 cells.

351 In conclusion, by integrating single-cell multiomics data, our research
352 demonstrated that systemic-venous ECs may play a significant pro-inflammatory role
353 in SSC patients, providing insights into the pathophysiology of SSC and therapeutic
354 alternatives.

355 **Competing interests**

356 None.

357 **Acknowledgments**

358 The authors would like to thank Anna and her colleagues for their nice work and the
359 generation of the single-cell multiomics dataset.

360

361 **Figure Legend**

362 **Figure 1 An overview of the epigenetic and transcriptional landscape of cell types**
363 **in the lungs of SSC-ILD patients and healthy individuals.** (**A**) UMAP projection of
364 48,622 scRNA-seq profiles of cell types in the lung from 11 individuals. Individual
365 cells are represented by dots, and different cell types are indicated by different colors.
366 (**B**) Dot plots of gene expression of the marker genes in the scRNA-seq dataset. The
367 size of the dots represents the percentage of cells in each cluster that have the gene of
368 interest. Color intensity denoted the standardized gene expression level. (**C**) Barplots
369 depicting the cell-type proportions of scRNA-seq data across all samples. The various
370 colors represent the major cell types. (**D**) UMAP projection of 14,717 scATAC-seq
371 profiles of lung cell types from 5 individuals. Individual cells are represented by dots,
372 while different cell types are represented by different colors. (**E**) Barplots of
373 scATAC-seq cell-type proportions across all samples. The colors represent the several
374 major cell types. (**F**) Validation of the scATAC-seq cell type annotation using paired 5
375 scRNA-seq data. Heatmap displaying evaluation findings based on the adjusted rand
376 index (ARI). (**G**) Annotated DAR location bar plot for each cell type. (**H**) Heatmap of
377 TF motifs found in peak sequences of cell-type markers. The color indicates the
378 normalized motif enrichment score in ArchR obtained with HOMER and the
379 hypergeometric test. (**I**) Multi-omics integration for scATAC-seq and scRNA-seq data
380 processing.

381

382 **Figure 2 Epigenomic and transcriptional signatures of EC subsets.** (**A**) UMAP
383 projection of 4,492 EC subpopulation scATAC-seq profiles. Individual cells are
384 represented by dots, while different cell types are represented by different colors. (**B**)
385 Differences in chosen population proportions across healthy individuals (n=6) and
386 SSC-ILD groups (n=5). The Wilcoxon test was used to get the *p*-values. (**C**) UMAP
387 projection of 1,760 scATAC-seq profiles of lung cell types from 5 individuals.
388 Individual cells are represented by dots, while different cell types are represented by
389 different colors. (**D**) Barplots of scRNA-seq data cell-type proportions across all
390 samples. The colors represent the indicated cell types. (**E**) UMAP depicting the ECs

391 lineage trajectory. Pseudotime values were superimposed on the UMAP embedding;
392 the smoothed line and arrow indicate the representation of the spline fit trajectory
393 route. (**F**) Heatmaps depicting the 323 ordered gene-score and matched RNA
394 expression trajectory from EC venous systemic to EC aerocyte capillary across
395 pseudotime. (**G**) Accessibility tracks surrounding the IL1R (up) and SELE (down)
396 loci. (**H**) Heatmaps of 10 positive TF regulators whose TF motif activity (left) and
397 corresponding gene expression (right) are positively associated across the ECs
398 differentiation pseudotime trajectory in (E). (**I**) Comparison of aggregate TF
399 footprints for FOSL2 and KLF4 in EC venous systemic and EC aerocyte capillary,
400 respectively. (**J**) A TF regulatory network depicting the TFs (FOSL2 and KLF4) and
401 probable target genes. (**K**) chromVAR bias-corrected deviation scores, RNA
402 expression, and selected target gene expression across ECs pseudotime for the
403 specified TFs. In an individual pseudotime-ordered scATAC-seq profile, each dot
404 indicates the deviation score or RNA expression. The smoothed fit spanning
405 pseudotime and chromVAR deviation scores or RNA expression is represented by the
406 line. (**L**) Volcano plots exhibiting the differential TF motif accessibility in the
407 chromVAR TF bias-corrected deviation between health and SSC-ILD individuals in
408 the identified ECs subpopulations using the mean TF motif accessibility. The *p*-values
409 were obtained using a Wilcoxon test, and the FDR was corrected using the
410 Benjamini-Hochberg method.
411
412 **Figure 3 scATAC-seq identifies mechanisms for functional SSC GWAS variants.**
413 (**A**) Prioritizing potential SSC-associated GWAS SNPs using a multi-tiered method,
414 as described by Turner *et al*. (**B**) LD-expanded (r^2>0.8; EUR) SSC GWAS variants
415 (50 bp) overlap with cell-type peaks. (**C**) Heatmap displaying the number of peaks per
416 cell type that intersect SSC GWAS loci. (**D**) Normalized chromatin accessibility
417 landscape for cell-type-specific pseudobulk tracks in the vicinity of the TNIP1 (left)
418 and NUP85 (right) loci. Below the ATAC-seq tracks are the positions of the
419 ATAC-seq peaks, the GWAS lead SNP, the SNP candidates in LD with the lead SNP,
420 and the candidate functional SNP. (**E**) Lung eQTL in GTEx for rs3792783 and

421 rs1005714. The G allele for rs3792783 is strongly related to lower TNIP1 mRNA
422 levels in GTEx lung (left two panels). The G allele for rs1005714 is strongly related
423 to elevated MRPS7 and NUP85 mRNA levels (right two panels). The eQTL *p*-value is
424 displayed as a result of the GTEx pipeline's linear regression between genotype and
425 normalized gene expression levels. The boxplot within the violin plot includes the
426 median and interquartile range (IQR) from 25% to 75%.
427
428 **Figure 4 The cellular-cellular communications of cell types in lung.** (**A**) A dotplot
429 depicting the differences in ligand-receptor interaction events between cell types in
430 the healthy and SSC-ILD groups. The dotplot (left) and heatmap (right) show the
431 number of interactions and fold changes (SSC-ILD vs healthy individuals). (**B**) Circos
432 plots depicting genes implicated in ligand-receptor interactions between cell types.
433 Color scale is proportional to interaction strength as assessed by CellPhoneDB
434 interaction means of statistically significant interactions. (**C**) Dot plots demonstrating
435 an increase in gene expression for the specified genes. The size of the dots represents
436 the percentage of cells that express the marker. The color scale shows the average
437 gene expression levels. The arrows represent ligand-receptor pairings generated from
438 CellPhoneDB. The X represents signaling that has been disabled. (**D**) Single-cell
439 chromatin accessibility in the CXCL10 and CXCR3 loci as seen by genome browser
440 tracks.
441
442 **Figure S1 shows an overview of the single-cell RNA and ATAC datasets.** (**A**)
443 UMAP embedding of scRNA-seq data from all samples for cell types, samples, and
444 disease states. (**B**) UMAP embedding of scATAC-seq data from all samples for cell
445 types, samples, and disease states. CCL18 (macrophages), S100A8 (monocytes),
446 CD3D (T cells), FCN3 (endothelial cells), TMEM190 (multiciliated), DCN
447 (fibroblasts), BPIFB1 (basal resting), SFTPA1 (AT2), TPSAB1 (mast cells), ACTA2
448 (smooth muscle), CD79A (B cells), and CCL21 (lymphatic EC mature).
449
450 **Figure S2 Sub-clustering of ECs and integration of scRNA and scATAC datasets.**

451 (**A**) UMAP embedding of scRNA-seq data from 11 samples in the SSC group and
452 health individuals. (**B**) Dot plots showing marker gene expression in the scRNA-seq
453 dataset. The dot size represents the percentage of cells in each cluster that have the
454 gene of interest. The color intensity reflected the standardized gene expression level.
455 (**C**) Barplots of scRNA-seq data cell-type proportions across all samples. The colors
456 represent the several primary cell kinds. (**D**) Proportions of EC subpopulations differ
457 between healthy individuals (n=6) and SSC-ILD groups (n=5). The Wilcoxon test was
458 used to compute the *p*-values. (**E**) Monocle 3 inferred scRNA-seq EC venous
459 systemic differentiation route towards EC aerocyte capillary. Pseudotime colors the
460 cells. (**F**) RNA velocity cell trajectory analysis of ECs. (**G**) UMAP plot of
461 scATACA-seq dataset by gene score for cell-type marker genes, including RPS29 (EC
462 venous systemic), DKK2 (EC arterial), and APLN (EC aerocyte capillaries). (**H**)
463 Validation of scATAC-seq cell type annotation using scRNA-seq data. Heatmap
464 showing evaluation results by adjusted rand index (ARI).
465
466 **Figure S3 Sub-clustering of AT2 using scRNA datasets.** (**A**) UMAP embedding of
467 scRNA-seq data from 11 SSC and healthy individuals. (**B**) Dot plots showing marker
468 gene expression in the scRNA-seq dataset. The dot size represents the percentage of
469 cells in each cluster that have the gene of interest. The color intensity reflected the
470 standardized gene expression level. (**C**) Barplots of scRNA-seq data cell-type
471 proportions across all samples. The colors represent the several primary cell kinds. (**D**)
472 Differences in the proportions of four AT2 clusters among healthy individuals (n=6)
473 and SSC-ILD groups (n=5). The Wilcoxon test was used to compute the *p*-values. (**E**)
474 AT2 RNA velocity cell trajectory analysis. AT2, alveolar type 2 epithelial cell.
475
476 **Figure S4 Co-accessibility of scATAC-seq and integration with scRNA-seq link**
477 **putative regulatory elements to target gene.** (**A**) Heatmap summary of ArchR
478 Peak2Gene connections (n=31,432) at a resolution of 10 kb, where chromatin
479 accessibility is substantially linked with target gene expression. The Z-scores for
480 scATAC-seq peak accessibility are shown on the left, while the Z-scores for RNA

481 expression are shown on the right. (**B**) *cis*-regulatory architecture and lung cell types
482 at the MRPS7 GWAS loci. (**C**) Violin plot of gene expression in the indicated illness
483 stages. Each dot represents the expression of a single gene in a single cell.

484 **Supplementary Table 1. The statistics generated in the study.**

485

486 **Reference**

487 1 Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis,
488 1972-2002. *Ann Rheum Dis* 2007;66:940-4.

489 2 Tamby MC, Chanseaud Y, Guillevin L, et al. New insights into the pathogenesis
490 of systemic sclerosis. *Autoimmun Rev* 2003;2:152-7.

491 3 Valenzi E, Bulik M, Tabib T, et al. Single-cell analysis reveals fibroblast
492 heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung
493 disease. *Ann Rheum Dis* 2019;78:1379-87.

494 4 Papazoglou A, Huang M, Bulik M, et al. Epigenetic Regulation of Profibrotic
495 Macrophages in Systemic Sclerosis-Associated Interstitial Lung Disease. *Arthritis*
496 *Rheumatol* 2022;74:2003-14.

497 5 Shao Y, Saredy J, Yang WY, et al. Vascular Endothelial Cells and Innate
498 Immunity. *Arterioscler Thromb Vasc Biol* 2020;40:E138-52.

499 6 Amersfoort J, Eelen G, Carmeliet P. Immunomodulation by endothelial cells-
500 partnering up with the immune system? *Nat Rev Immunol* 2022;22:576-88.

501 7 Lu Y, Sun Y, Xu K, et al. Editorial: Endothelial cells as innate immune cells.
502 *Front Immunol* 2022;13:1-5.

503 8 Mendoza FA, Piera-Velazquez S, Farber JL, et al. Endothelial Cells Expressing
504 Endothelial and Mesenchymal Cell Gene Products in Lung Tissue from Patients with
505 Systemic Sclerosis-Associated Interstitial Lung Disease. *Arthritis Rheumatol*
506 2016;68:210-7.

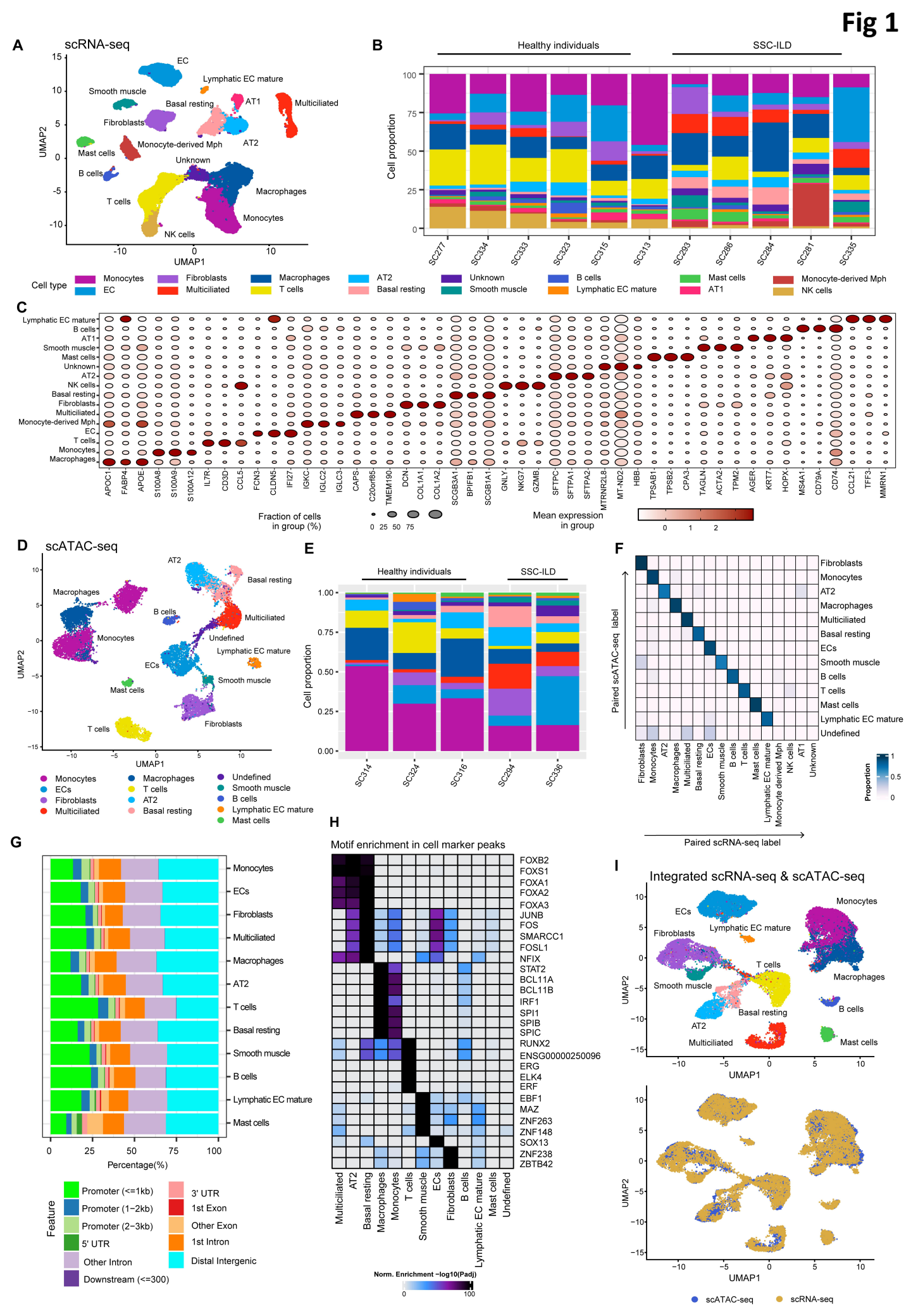
507 9 Bian F, Lan YW, Zhao S, et al. Lung endothelial cells regulate pulmonary fibrosis
508 through FOXF1/R-Ras signaling. *Nat Commun* 2023;14.

509 10 Granja, JM, Corces MR, Pierce SE. et al. ArchR is a scalable software package

510 for integrative single-cell chromatin accessibility analysis. *Nat Genet*
511 2021;53:403-411.
512 11 Li Z, Kuppe C, Ziegler S, et al. Chromatin-accessibility estimation from
513 single-cell ATAC-seq data with scOpen. *Nat Commun* 2021;12.
514 12 Korsunsky I, Millard N, Fan J, *et al*. Fast, sensitive and accurate integration of
515 single-cell data with harmony. *Nat Methods* 2019;16:1289-96.
516 13 van Dijk D, Sharma R, Nainys J, et al. Recovering Gene Interactions from
517 Single-Cell Data Using Data Diffusion. *Cell* 2018;174:716-729.e27.
518 14 Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS).
519 Genome Biol 2008;9.
520 15 Schep A, Wu B, Buenrostro J. et al. chromVAR: inferring
521 transcription-factor-associated accessibility from single-cell epigenomic data. *Nat*
522 *Methods* 2017;14:975-978.
523 16 Yu G, Wang LG, He QY. ChIP seeker: An R/Bioconductor package for ChIP peak
524 annotation, comparison and visualization. *Bioinformatics* 2015;31:2382-3.
525 17 López-Isac E, Acosta-Herrera M, Kerick M, et al. GWAS for systemic sclerosis
526 identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat*
527 *Commun* 2019;10:1-14.
528 18 Turner AW, Hu SS, Mosquera JV, et al. Single-nucleus chromatin accessibility
529 profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat Genet*
530 2022;54:804-16.
531 19 Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome
532 association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75.
533 20 Neph S, Kuehn MS, Reynolds AP, et al. BEDOPS: High-performance genomic
534 feature operations. *Bioinformatics* 2012;28:1919-20.
535 21 Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing
536 genomic features. *Bioinformatics* 2010;26:841-2.
537 22 Butler A, Hoffman P, Smibert P, et al. Integrating single-cell transcriptomic data
538 across different conditions, technologies, and species. *Nat Biotechnol* 2018;36:411-20.
539 23 Li SF & Gong MJ. Single-cell RNA sequencing reveals the heterogeneity of

540 endothelial cells in human dupuytren's disease. Preprint in *Researchsquare.*
541 https://doi.org/10.21203/rs.3.rs-3005025/v2
542 24 Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell
543 Doublets in Single-Cell Transcriptomic Data. *Cell Syst* 2019;8:281-291.e9.
544 25 Domínguez Conde C, Xu C, Jarvis LB, et al. Cross-tissue immune cell analysis
545 reveals tissue-specific features in humans. *Science* (80-) 2022;376.
546 26 Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The
547 dynamics and regulators of cell fate decisions are revealed by pseudotemporal
548 ordering of single cells. *Nat Biotechnol* 2014. 32, 381-386.
549 27 La Manno G, Soldatov R, Zeisel A, et al. RNA velocity of single cells. *Nature*
550 2018;560:494-8.
551 28 Garcia-Alonso L, Lorenzi V, Mazzeo CI, et al. Single-cell roadmap of human
552 gonadal development. *Nature* 2022;607:540-7.
553 29 Kuppe C, Ramirez Flores RO, Li Z, et al. Spatial multi-omic map of human
554 myocardial infarction. *Nature* 2022;608:766-77.
555 30 Perelas A, Silver RM, Arrossi A V., et al. Systemic sclerosis-associated interstitial
556 lung disease. *Lancet Respir Med* 2020;8:304-20.
557 31 Di Martino ML, Frau A, Losa F, et al. Role of circulating endothelial cells in
558 assessing the severity of systemic sclerosis and predicting its clinical worsening. *Sci*
559 *Rep* 2021;11:1-9.
560 32 Corces MR, Shcherbina A, Kundu S, et al. Single-cell epigenomic analyses
561 implicate candidate causal variants at inherited risk loci for Alzheimer's and
562 Parkinson's diseases. *Nat Genet* 2020;52:1158-68.
563 33 Alexi N, Inge RH, Nicole GC, et al. Brain cell type-specific enhancer-promoter
564 interactome maps and disease-risk association. *Science* (80-) 2019;1139:1134-9.
565 34 Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common
566 disease-associated variation in regulatory DNA. *Science* (80-) 2012;337:1190-5.
567 35 Ahn Y, Sanderson BW, Klein OD, et al. Inhibition of Wnt signaling by wise
568 (Sostdc1) and negative feedback from Shh controls tooth number and patterning.
569 *Development* 2010;137:3221-31.

Fig 1



A
scRNA-seq
UMAP1
UMAP2
EC
Lymphatic EC mature
Smooth muscle
Basal resting
AT1
Multiciliated
Fibroblasts
Monocyte-derived Mph
AT2
Mast cells
Unknown
B cells
Macrophages
T cells
Monocytes
NK cells
Cell type
Monocytes
EC
Fibroblasts
Multiciliated
Macrophages
T cells
AT2
Basal resting
Unknown
Smooth muscle
B cells
Lymphatic EC mature
Mast cells
AT1
Monocyte-derived Mph
NK cells
B
Healthy individuals
SSC-ILD
Cell proportion
SC277
SC334
SC333
SC323
SC315
SC313
SC293
SC286
SC284
SC281
SC335
C
Lymphatic EC mature
B cells
AT1
Smooth muscle
Mast cells
Unknown
AT2
NK cells
Basal resting
Fibroblasts
Multiciliated
Monocyte-derived Mph
EC
T cells
Monocytes
Macrophages
APOC1
FABP4
APOE
S100A8
S100A9
S100A12
IL7R
CD3D
CCL5
FCN3
CLDN5
IFI27
IGKC
IGLC2
IGLC3
CAPS
C20orf85
TMEM190
DCN
COL1A1
COL1A2
SCGB3A1
BPIFB1
SCGB1A1
GNLY
NKG7
GZMB
SFTPC
SFTPA1
SFTPA2
MTRNR2L8
MT-ND2
HBB
TPSAB1
TPSB2
CPA3
TAGLN
ACTA2
TPM2
AGER
KRT7
HOPX
MS4A1
CD79A
CD74
CCL21
TFF3
MMRN1
Fraction of cells in group (%)
0 25 50 75
Mean expression in group
0 1 2
D
scATAC-seq
UMAP1
UMAP2
AT2
Basal resting
Macrophages
B cells
Multiciliated
Monocytes
Undefined
Lymphatic EC mature
ECs
Smooth muscle
Mast cells
Fibroblasts
T cells
Monocytes
ECs
Fibroblasts
Multiciliated
Macrophages
T cells
AT2
Basal resting
Undefined
Smooth muscle
B cells
Lymphatic EC mature
Mast cells
E
Healthy individuals
SSC-ILD
Cell proportion
SC314
SC324
SC316
SC294
SC336
F
Paired scATAC-seq label
Paired scRNA-seq label
Fibroblasts
Monocytes
AT2
Macrophages
Multiciliated
Basal resting
ECs
Smooth muscle
B cells
T cells
Mast cells
Lymphatic EC mature
Undefined
Monocyte derived Mph
NK cells
AT1
Unknown
Proportion
G
Monocytes
ECs
Fibroblasts
Multiciliated
Macrophages
AT2
T cells
Basal resting
Smooth muscle
B cells
Lymphatic EC mature
Mast cells
Percentage(%)
Feature
Promoter (<=1kb)
Promoter (1−2kb)
Promoter (2−3kb)
5' UTR
Other Intron
Downstream (<=300)
3' UTR
1st Exon
Other Exon
1st Intron
Distal Intergenic
H
Motif enrichment in cell marker peaks
FOXB2
FOXS1
FOXA1
FOXA2
FOXA3
JUNB
FOS
SMARCC1
FOSL1
NFIX
STAT2
BCL11A
BCL11B
IRF1
SPI1
SPIB
SPIC
RUNX2
ENSG00000250096
ERG
ELK4
ERF
EBF1
MAZ
ZNF263
ZNF148
SOX13
ZNF238
ZBTB42
Multiciliated
AT2
Basal resting
Macrophages
Monocytes
T cells
Smooth muscle
ECs
Fibroblasts
B cells
Lymphatic EC mature
Mast cells
Undefined
Norm. Enrichment −log10(Padj)
0 100
I
Integrated scRNA-seq & scATAC-seq
UMAP1
UMAP2
ECs
Monocytes
Fibroblasts
Lymphatic EC mature
T cells
Smooth muscle
Macrophages
Basal resting
B cells
AT2
Multiciliated
Mast cells
scATAC-seq
scRNA-seq

Fig 2

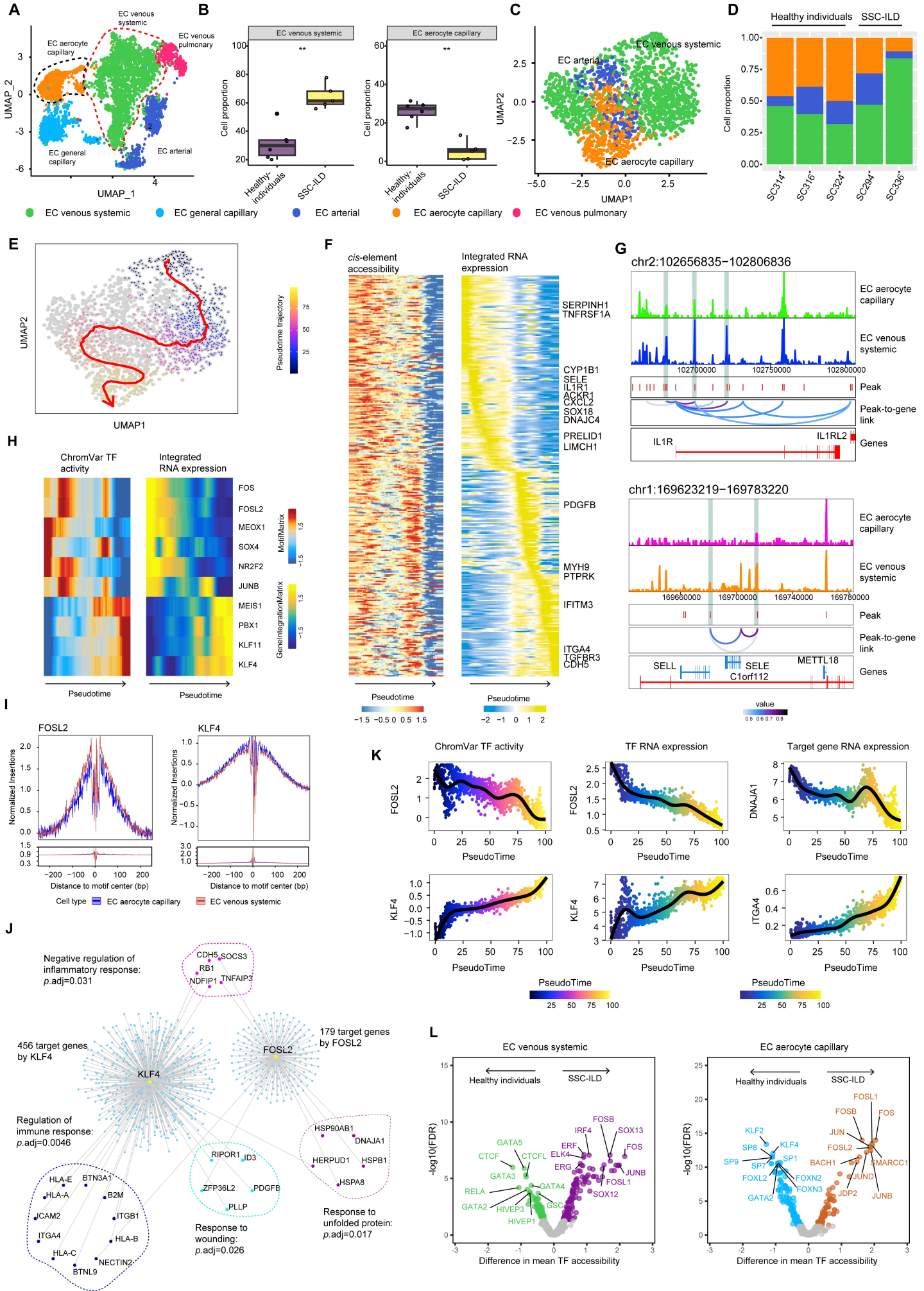

A
EC venous systemic
EC venous pulmonary
EC aerocyte capillary
EC general capillary
EC arterial
UMAP_2
UMAP_1
B
EC venous systemic
EC aerocyte capillary
Cell proportion
Healthy-individuals
SSC-ILD
C
EC venous systemic
EC arterial
EC aerocyte capillary
UMAP2
UMAP1
D
Healthy individuals
SSC-ILD
Cell proportion
SC314
SC316
SC324
SC294
SC336
EC venous systemic
EC general capillary
EC arterial
EC aerocyte capillary
EC venous pulmonary
E
UMAP2
UMAP1
Pseudotime trajectory
F
cis-element accessibility
Integrated RNA expression
SERPINH1
TNFRSF1A
CYP1B1
SELE
IL1R1
ACKR1
CXCL2
SOX18
DNAJC4
PRELID1
LIMCH1
PDGFB
MYH9
PTPRK
IFITM3
ITGA4
TGFBR3
CDH5
Pseudotime
G
chr2:102656835−102806836
EC aerocyte capillary
EC venous systemic
Peak
Peak-to-gene link
Genes
IL1R
IL1RL2
chr1:169623219−169783220
SELL
SELE
METTL18
C1orf112
value
H
ChromVar TF activity
Integrated RNA expression
FOS
FOSL2
MEOX1
SOX4
NR2F2
JUNB
MEIS1
PBX1
KLF11
KLF4
MotifMatrix
GeneIntegrationMatrix
Pseudotime
I
FOSL2
KLF4
Normalized insertions
Distance to motif center (bp)
Cell type
EC aerocyte capillary
EC venous systemic
J
Negative regulation of inflammatory response: p.adj=0.031
CDH5
SOCS3
RB1
NDFIP1
TNFAIP3
456 target genes by KLF4
KLF4
FOSL2
179 target genes by FOSL2
Regulation of immune response: p.adj=0.0046
HSP90AB1
DNAJA1
HERPUD1
HSPB1
HSPA8
RIPOR1
ID3
ZFP36L2
PDGFB
PLLP
HLA-E
BTN3A1
HLA-A
B2M
ICAM2
ITGB1
ITGA4
HLA-B
HLA-C
NECTIN2
BTNL9
Response to wounding: p.adj=0.026
Response to unfolded protein: p.adj=0.017
K
ChromVar TF activity
TF RNA expression
Target gene RNA expression
FOSL2
DNAJA1
KLF4
ITGA4
PseudoTime
L
EC venous systemic
EC aerocyte capillary
Healthy individuals
SSC-ILD
-log10(FDR)
Difference in mean TF accessibility
FOSB
IRF4
SOX13
GATA5
ERF
FOS
CTCF
CTCFL
ELK4
ERG
JUNB
GATA3
FOSL1
GATA4
RELA
SOX12
GATA2
HIVEP3
GSC
HIVEP1
FOSL1
FOSB
FOS
KLF2
JUN
SP8
KLF4
FOSL2
SP9
SP1
BACH1
SMARCC1
SP7
FOXL2
FOXN2
JUND
FOXN3
JDP2
GATA2
JUNB

Fig 3

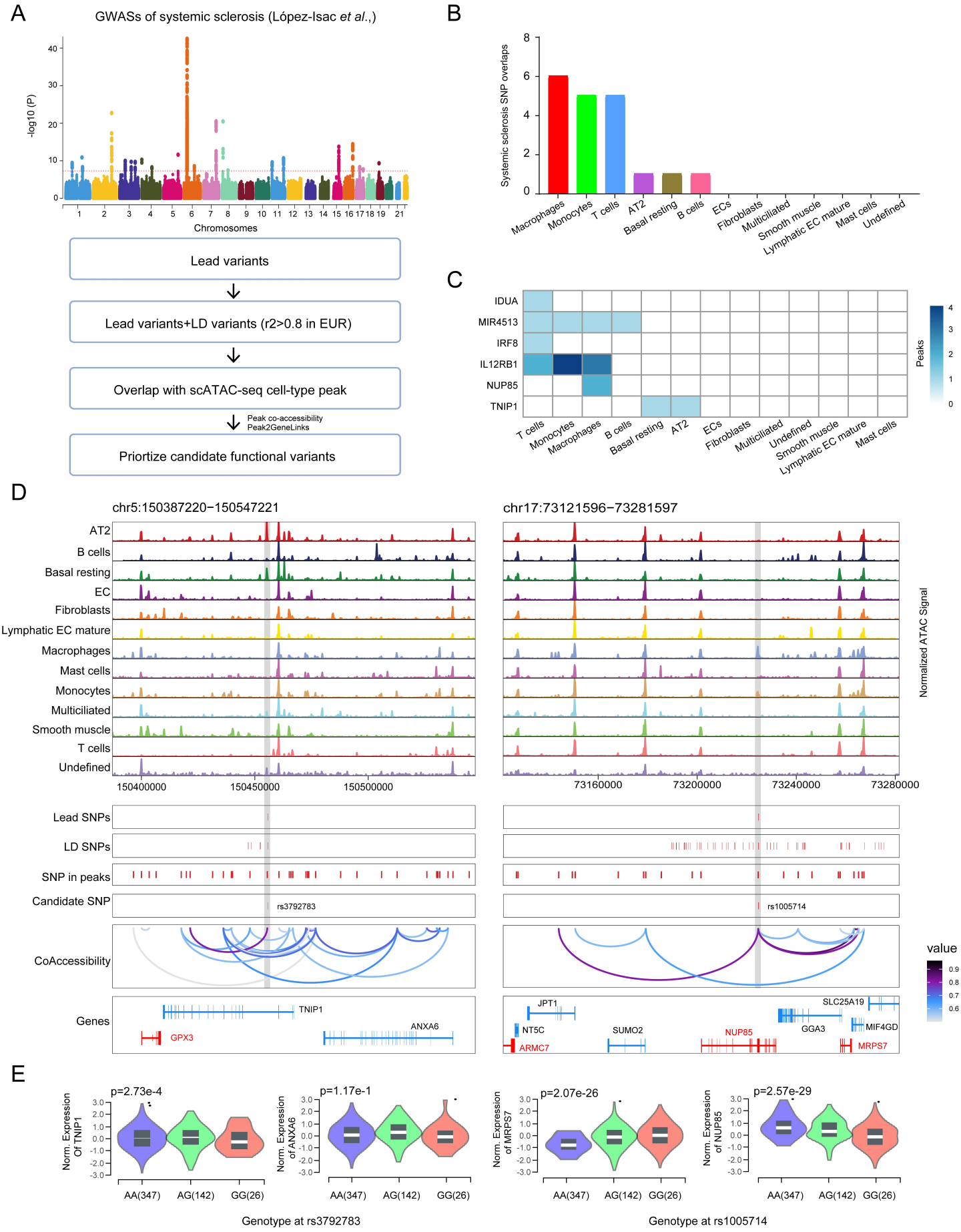

A
GWASs of systemic sclerosis (López-Isac et al.,)
-log10 (P)
Chromosomes
Lead variants
Lead variants+LD variants (r2>0.8 in EUR)
Overlap with scATAC-seq cell-type peak
Peak co-accessibility
Peak2GeneLinks
Priortize candidate functional variants
B
Systemic sclerosis SNP overlaps
Macrophages
Monocytes
T cells
AT2
Basal resting
B cells
ECs
Fibroblasts
Multiciliated
Smooth muscle
Lymphatic EC mature
Mast cells
Undefined
C
IDUA
MIR4513
IRF8
IL12RB1
NUP85
TNIP1
Peaks
D
chr5:150387220−150547221
chr17:73121596−73281597
Normalized ATAC Signal
Lead SNPs
LD SNPs
SNP in peaks
Candidate SNP
rs3792783
rs1005714
CoAccessibility
value
Genes
TNIP1
ANXA6
GPX3
JPT1
SLC25A19
NT5C
SUMO2
NUP85
GGA3
MIF4GD
ARMC7
MRPS7
E
p=2.73e-4
p=1.17e-1
p=2.07e-26
p=2.57e-29
Norm. Expression Of TNIP1
Norm. Expression of ANXA6
Norm. Expression of MRPS7
Norm. Expression of NUP85
AA(347)
AG(142)
GG(26)
Genotype at rs3792783
Genotype at rs1005714

Fig 4

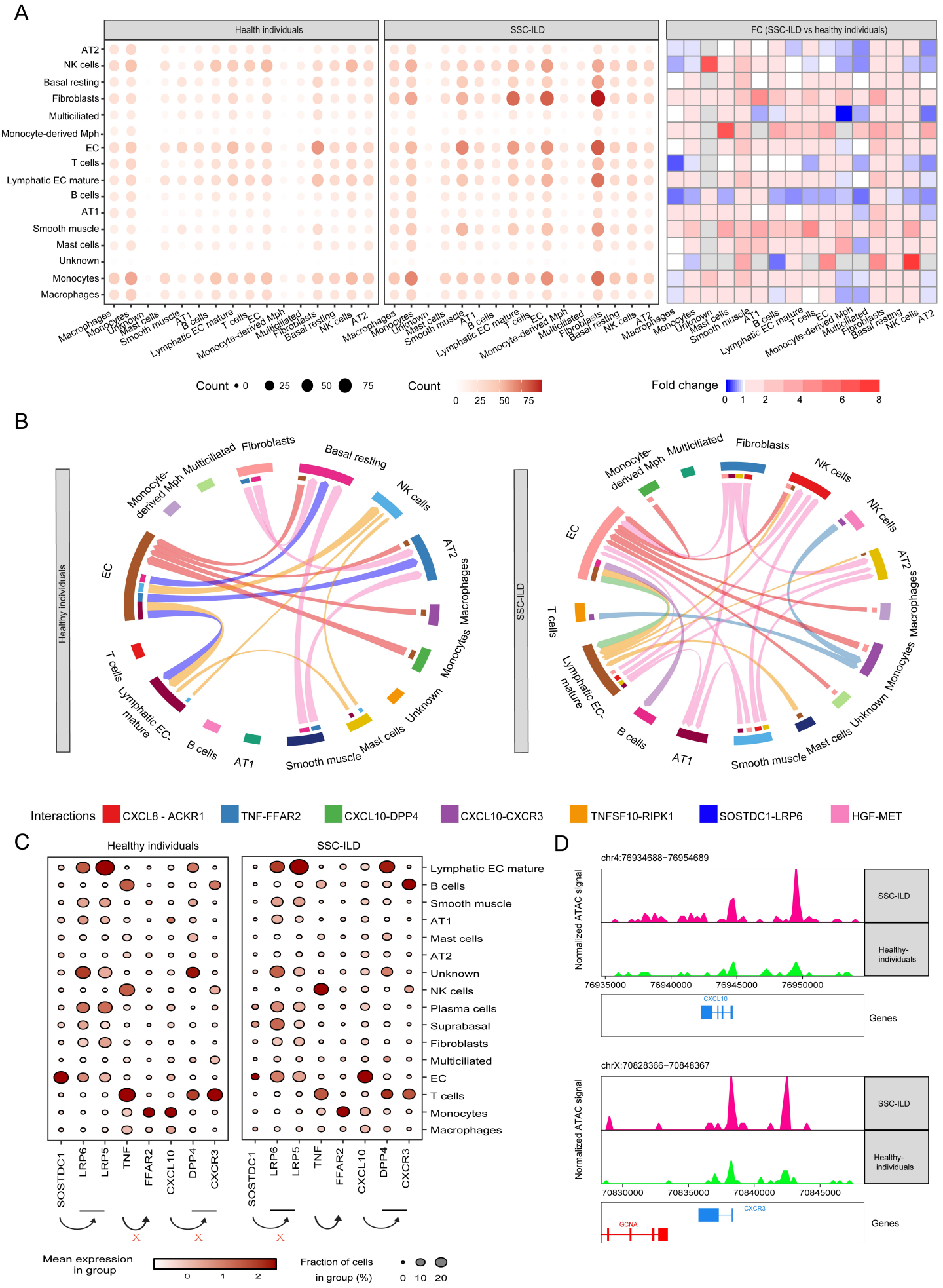

A
Health individuals
SSC-ILD
FC (SSC-ILD vs healthy individuals)
AT2
NK cells
Basal resting
Fibroblasts
Multiciliated
Monocyte-derived Mph
EC
T cells
Lymphatic EC mature
B cells
AT1
Smooth muscle
Mast cells
Unknown
Monocytes
Macrophages
Count 0 25 50 75
Count 0 25 50 75
Fold change 0 1 2 4 6 8
B
Healthy individuals
SSC-ILD
Interactions
CXCL8 - ACKR1
TNF-FFAR2
CXCL10-DPP4
CXCL10-CXCR3
TNFSF10-RIPK1
SOSTDC1-LRP6
HGF-MET
C
Healthy individuals
SSC-ILD
Lymphatic EC mature
B cells
Smooth muscle
AT1
Mast cells
AT2
Unknown
NK cells
Plasma cells
Suprabasal
Fibroblasts
Multiciliated
EC
T cells
Monocytes
Macrophages
SOSTDC1
LRP6
LRP5
TNF
FFAR2
CXCL10
DPP4
CXCR3
Mean expression in group 0 1 2
Fraction of cells in group (%) 0 10 20
D
chr4:76934688−76954689
Normalized ATAC signal
SSC-ILD
Healthy-individuals
76935000
76940000
76945000
76950000
CXCL10
Genes
chrX:70828366−70848367
Normalized ATAC signal
SSC-ILD
Healthy-individuals
70830000
70835000
70840000
70845000
CXCR3
GCNA
Genes

Fig S1

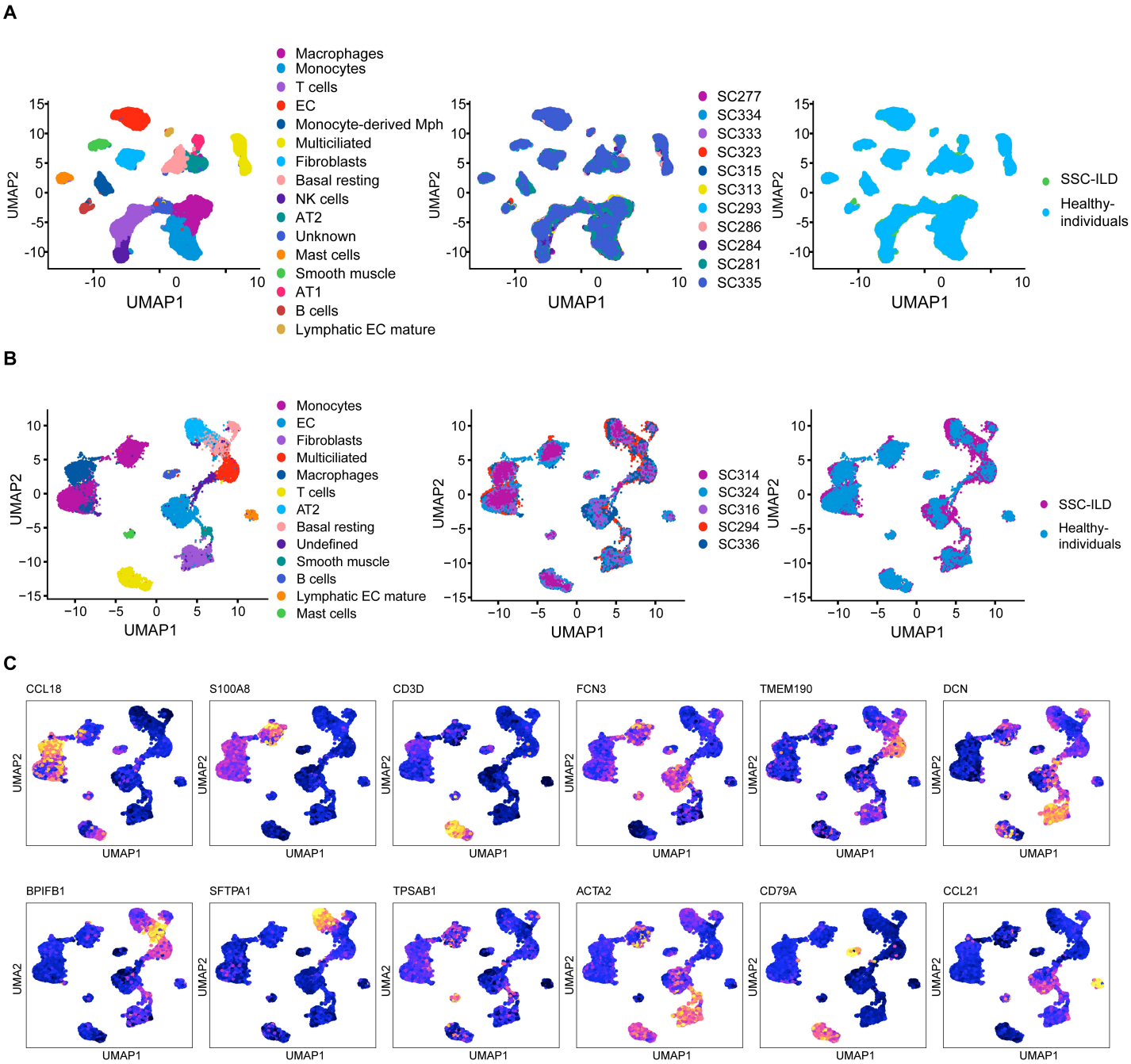

A
Macrophages
Monocytes
T cells
EC
Monocyte-derived Mph
Multiciliated
Fibroblasts
Basal resting
NK cells
AT2
Unknown
Mast cells
Smooth muscle
AT1
B cells
Lymphatic EC mature
SC277
SC334
SC333
SC323
SC315
SC313
SC293
SC286
SC284
SC281
SC335
SSC-ILD
Healthy-individuals
UMAP1
UMAP2
B
Monocytes
EC
Fibroblasts
Multiciliated
Macrophages
T cells
AT2
Basal resting
Undefined
Smooth muscle
B cells
Lymphatic EC mature
Mast cells
SC314
SC324
SC316
SC294
SC336
SSC-ILD
Healthy-individuals
C
CCL18
S100A8
CD3D
FCN3
TMEM190
DCN
BPIFB1
SFTPA1
TPSAB1
ACTA2
CD79A
CCL21

Fig S2

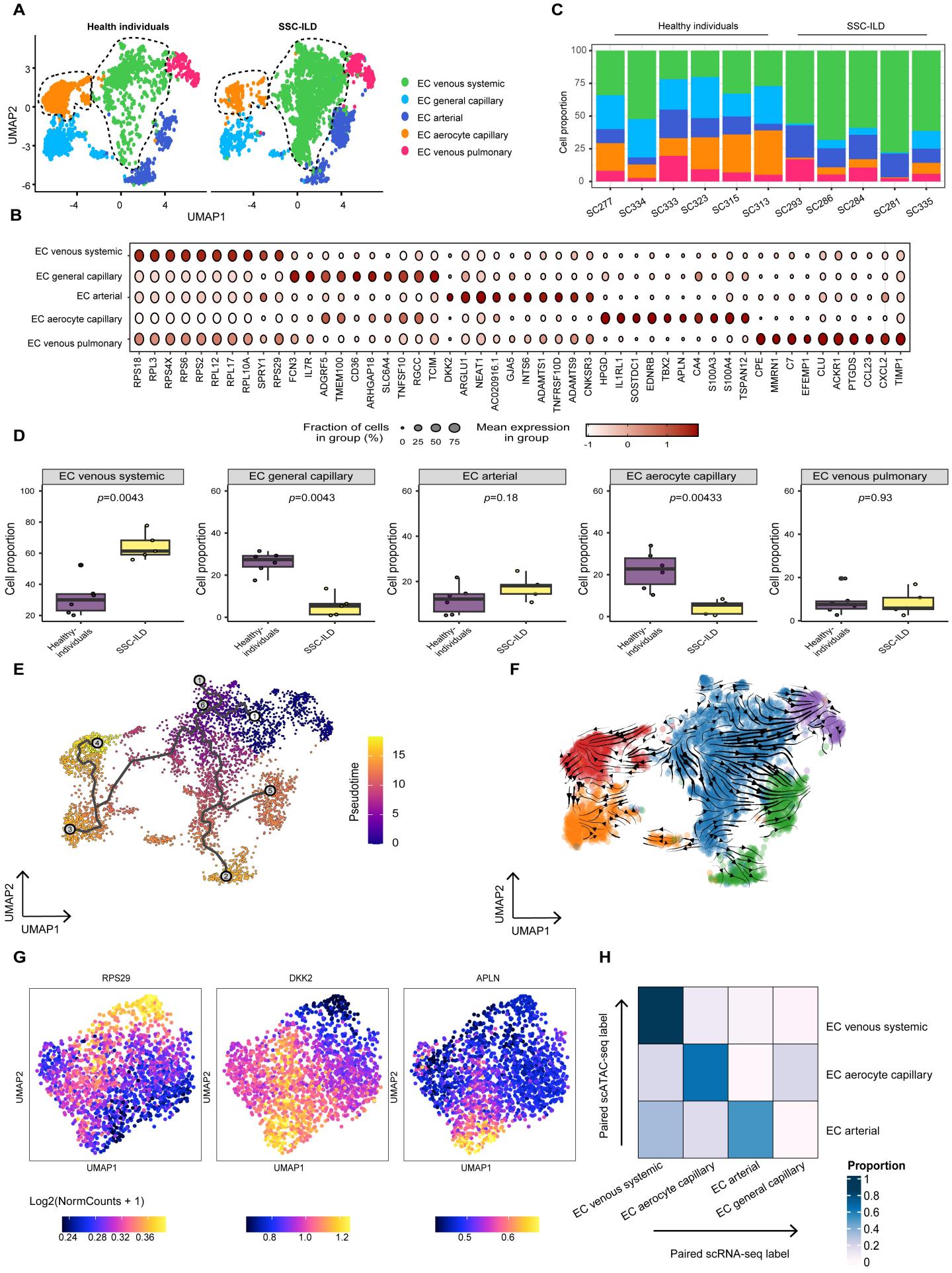

A
Health individuals
SSC-ILD
UMAP2
UMAP1
EC venous systemic
EC general capillary
EC arterial
EC aerocyte capillary
EC venous pulmonary
C
Healthy individuals
SSC-ILD
Cell proportion
SC277
SC334
SC333
SC323
SC315
SC313
SC293
SC286
SC284
SC281
SC335
B
RPS18 RPL3 RPS4X RPS6 RPS2 RPL12 RPL17 RPL10A SPRY1 RPS29 FCN3 IL7R ADGRF5 TMEM100 CD36 ARHGAP18 SLC6A4 TNFSF10 RGCC TCIM DKK2 ARGLU1 NEAT1 AC020916.1 GJA5 INTS6 ADAMTS1 TNFRSF10D ADAMTS9 CNKSR3 HPGD IL1RL1 SOSTDC1 EDNRB TBX2 APLN CA4 S100A3 S100A4 TSPAN12 CPE MMRN1 C7 EFEMP1 CLU ACKR1 PTGDS CCL23 CXCL2 TIMP1
Fraction of cells in group (%)
Mean expression in group
D
EC venous systemic
p=0.0043
EC general capillary
p=0.0043
EC arterial
p=0.18
EC aerocyte capillary
p=0.00433
EC venous pulmonary
p=0.93
Healthy individuals
SSC-ILD
E
Pseudotime
F
G
RPS29
DKK2
APLN
Log2(NormCounts + 1)
H
Paired scATAC-seq label
Paired scRNA-seq label
EC venous systemic
EC aerocyte capillary
EC arterial
EC general capillary
Proportion

Fig S3

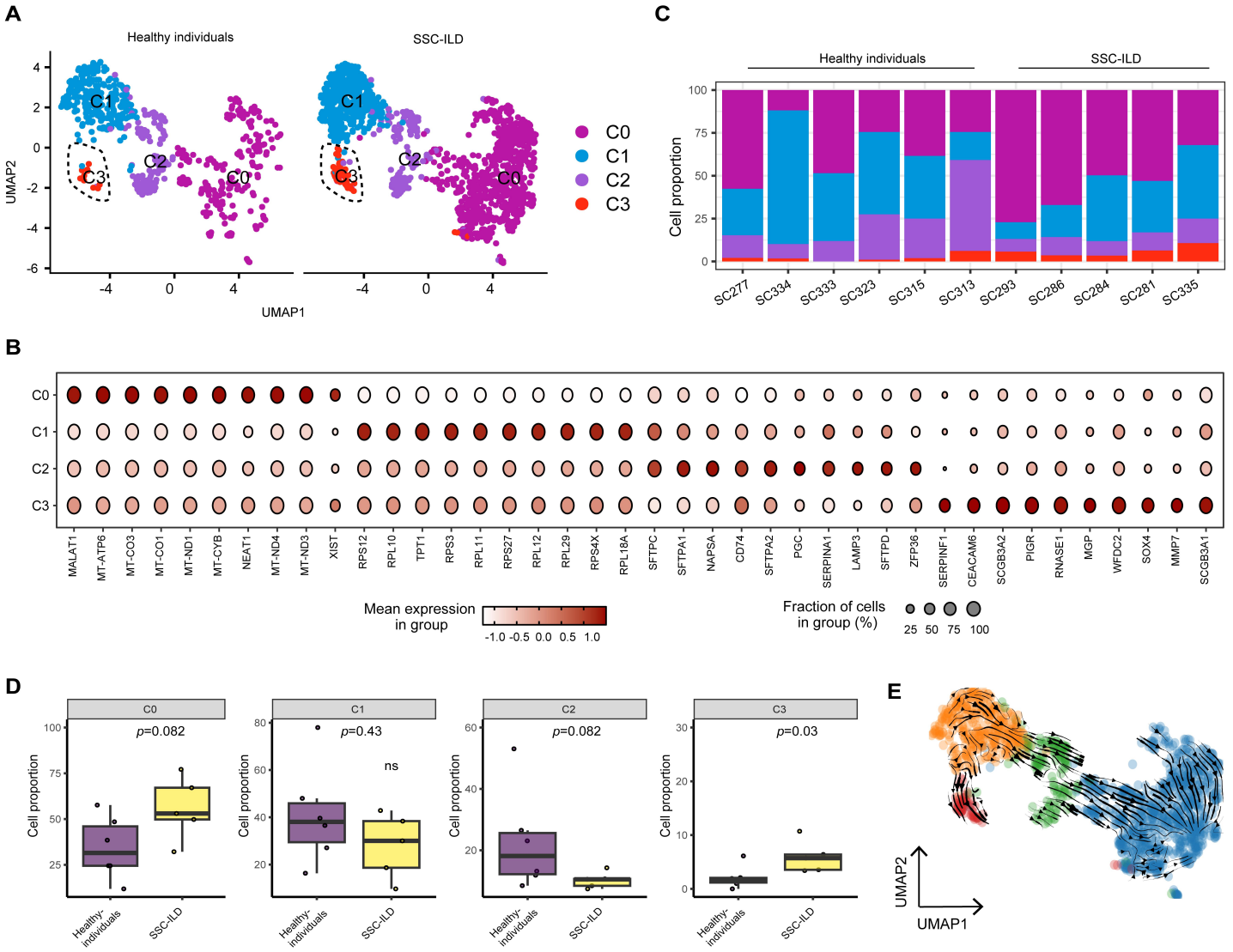
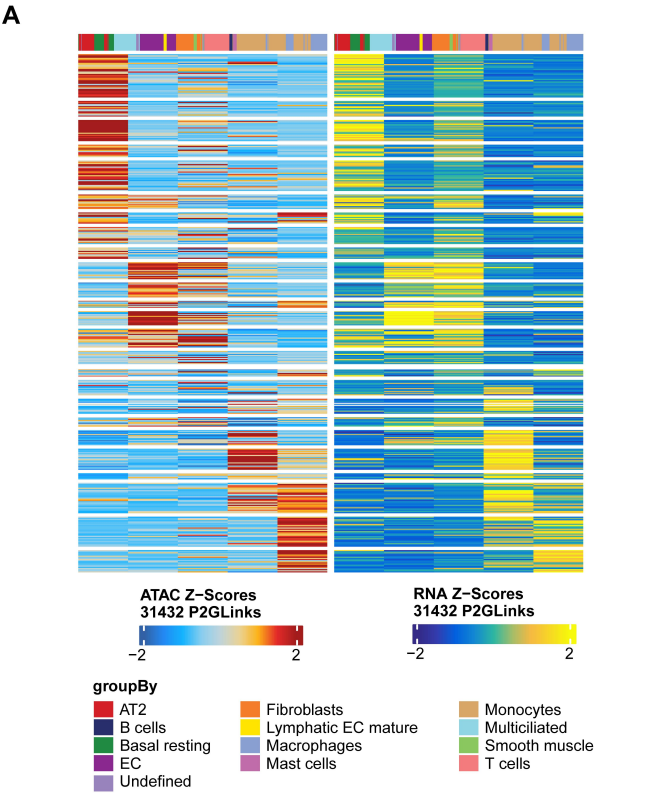
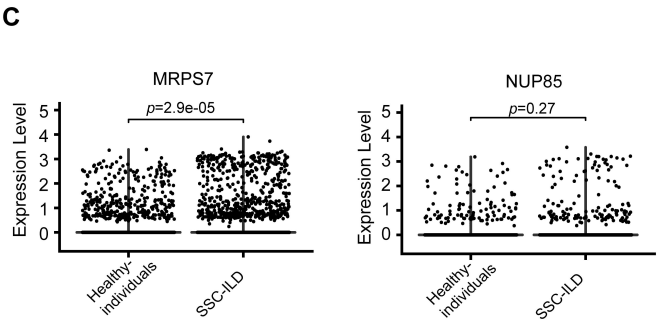
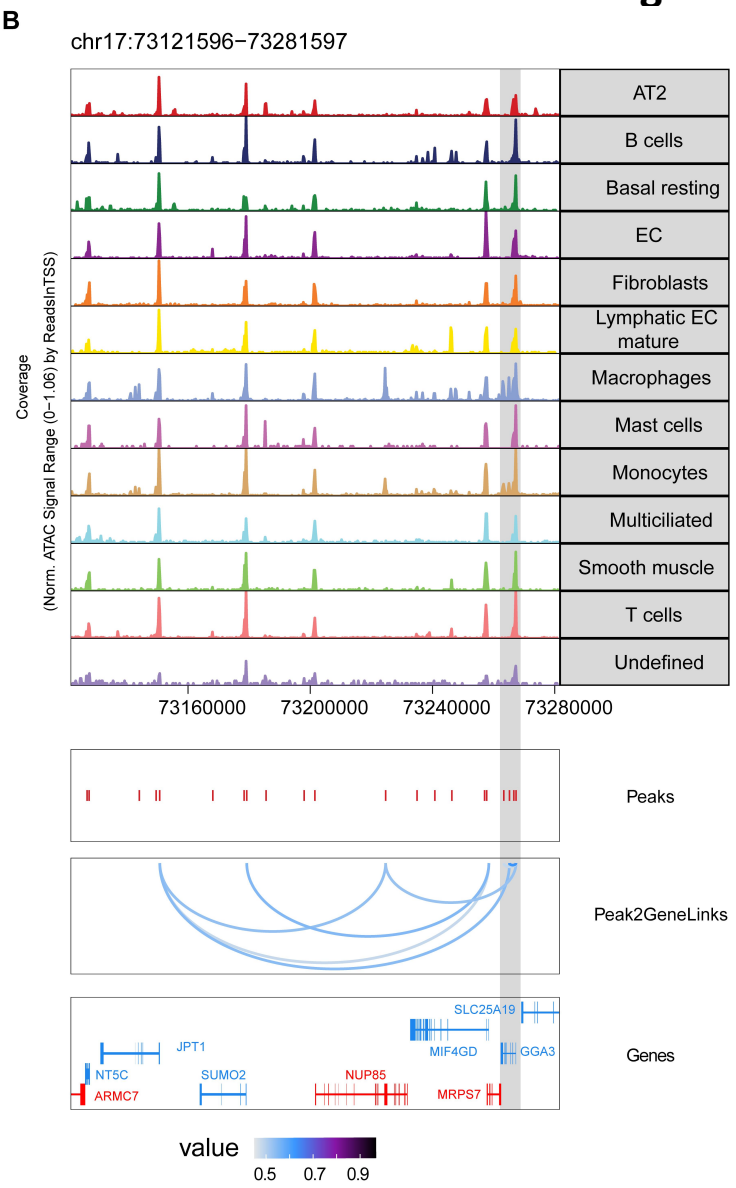

A
Healthy individuals
SSC-ILD
UMAP2
UMAP1
C0
C1
C2
C3
C
Healthy individuals
SSC-ILD
Cell proportion
SC277
SC334
SC333
SC323
SC315
SC313
SC293
SC286
SC284
SC281
SC335
B
MALAT1
MT-ATP6
MT-CO3
MT-CO1
MT-ND1
MT-CYB
NEAT1
MT-ND4
MT-ND3
XIST
RPS12
RPL10
TPT1
RPS3
RPL11
RPS27
RPL12
RPL29
RPS4X
RPL18A
SFTPC
SFTPA1
NAPSA
CD74
SFTPA2
PGC
SERPINA1
LAMP3
SFTPD
ZFP36
SERPINF1
CEACAM6
SCGB3A2
PIGR
RNASE1
MGP
WFDC2
SOX4
MMP7
SCGB3A1
Mean expression in group
-1.0 -0.5 0.0 0.5 1.0
Fraction of cells in group (%)
25 50 75 100
D
C0
C1
C2
C3
p=0.082
p=0.43
ns
p=0.082
p=0.03
Cell proportion
Healthy-individuals
SSC-ILD
E
UMAP2
UMAP1

Fig S4


A
ATAC Z−Scores
31432 P2GLinks
−2
2
RNA Z−Scores
31432 P2GLinks
−2
2
groupBy
AT2
B cells
Basal resting
EC
Undefined
Fibroblasts
Lymphatic EC mature
Macrophages
Mast cells
Monocytes
Multiciliated
Smooth muscle
T cells


C
MRPS7
p=2.9e-05
Expression Level
Healthy-individuals
SSC-ILD
NUP85
p=0.27
Expression Level
Healthy-individuals
SSC-ILD


B
chr17:73121596−73281597
Coverage
(Norm. ATAC Signal Range (0−1.06) by ReadsInTSS)
AT2
B cells
Basal resting
EC
Fibroblasts
Lymphatic EC mature
Macrophages
Mast cells
Monocytes
Multiciliated
Smooth muscle
T cells
Undefined
73160000
73200000
73240000
73280000
Peaks
Peak2GeneLinks
Genes
SLC25A19
JPT1
MIF4GD
GGA3
NT5C
SUMO2
NUP85
ARMC7
MRPS7
value
0.5
0.7
0.9