Journal of Autoimmunity xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

Integrated bulk and single-cell RNA-sequencing identified disease-relevant monocytes and a gene network module underlying systemic sclerosis

Satomi Kobayashi [a], Yasuo Nagafuchi [a,b,*], Mai Okubo [a], Yusuke Sugimori [a], Harumi Shirai [a], Hiroaki Hatano [a], Maeda Junko [a], Haruyuki Yanaoka [a,c], Yusuke Takeshima [a,b], Mineto Ota [a,b], Yukiko Iwasaki [a], Shuji Sumitomo [a], Tomohisa Okamura [a,b], Kazuhiko Yamamoto [a,d], Hirofumi Shoda [a], Keishi Fujio [a,**]

[a] Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan
[b] Department of Functional Genomics and Immunological Diseases, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan
[c] Immuno-Rheumatology Center, St. Luke's International Hospital, St. Luke's International University, 9-1 Akashi-cho, Chuo-ku, Tokyo, 104-8560, Japan
[d] Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan

ARTICLE INFO

Keywords:
Systemic sclerosis
Immune cells
Bulk RNA-seq
Single cell RNA-seq
Monocytes

ABSTRACT

Objective: Immunological disturbances have been reported in systemic sclerosis (SSc). This study assessed the transcriptome disturbances in immune cell subsets in SSc and characterized a disease-related gene network module and immune cell cluster at single cell resolution.

Methods: Twenty-one Japanese SSc patients were enrolled and compared with 13 age- and sex-matched healthy controls (HC). Nineteen peripheral blood immune cell subsets were sorted by flow cytometry and bulk RNA-seq analysis was performed for each. Differential expression and pathway analyses were conducted. Iterative weighted gene correlation network analysis (iWGCNA) of each subset revealed clustered co-expressed gene network modules. Random forest analysis prioritized a disease-related gene module. Single cell RNA-seq analysis of 878 monocytes was integrated with bulk RNA-seq analysis and with a public database for single cell RNA-seq analysis of SSc patients.

Results: Inflammatory pathway genes were differentially expressed in widespread immune cell subsets of SSc. An inflammatory gene module from $CD16^{+}$ monocytes, which included *KLF10, PLAUR, JUNB* and *JUND,* showed the greatest discrimination between SSc and HC. One of the clusters of SSc monocytes identified by single-cell RNA-seq analysis characteristically expressed these inflammatory co-expressed genes and was similar to lung infiltrating $FCN1^{hi}$ monocytes expressing *IL1B*.

Conclusions: Our integrated analysis of bulk and single cell RNA-seq analysis identified an inflammatory gene module and a cluster of monocytes that are relevant to SSc pathophysiology. They could serve as candidate novel therapeutic targets in SSc.

1. Introduction

Systemic sclerosis (SSc) is an autoimmune disease characterized by microvasculopathy and organ fibrosis, including skin and lung [1,2]. The progressive organ fibrosis causes life-threating disorders, functional impairment and a reduced quality of life [1]. In addition, the rate of

* Corresponding author. Department of Allergy and Rheumatology and Department of Functional Genomics and Immunological Diseases, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan.
** Corresponding author.
E-mail addresses: satokobayashi-tky@umin.org (S. Kobayashi), nagafuchi-tky@umin.ac.jp (Y. Nagafuchi), mookubo-tky@umin.ac.jp (M. Okubo), ysugimori-tky@umin.ac.jp (Y. Sugimori), hare80415@hotmail.com (H. Shirai), hhatano-tky@umin.ac.jp (H. Hatano), jmaeda0512@gmail.com (M. Junko), mich@luke.ac.jp (H. Yanaoka), ytakeshima-tky@umin.ac.jp (Y. Takeshima), miota-tky@umin.ac.jp (M. Ota), yunyan-todai@umin.ac.jp (Y. Iwasaki), sumitos-tky@umin.ac.jp (S. Sumitomo), tomohisa-tky@umin.ac.jp (T. Okamura), kazuhiko.yamamoto@riken.jp (K. Yamamoto), shoda-tky@umin.ac.jp (H. Shoda), kfujio-tky@umin.ac.jp (K. Fujio).

https://doi.org/10.1016/j.jaut.2020.102547
Received 12 April 2020; Received in revised form 15 September 2020; Accepted 25 September 2020

disease-related mortality is higher in SSc patients than in patients with other autoimmune diseases because effective therapy has not yet been established [3]. Immunosuppressive therapies, including the use of steroids, cyclophosphamide or mycophenolate mofetil [4] have been used for SSc, however, the adverse reactions are intolerable for many patients and serious infection sometimes occurs. Recently, clinical trials have begun to examine the efficacy of biologic reagents and anti-fibrotic compounds. In particular, an IL-6 inhibitor, tocilizumab, improved dermal sclerosis effectively in a phase 2 trial [5]. These results support the notion that immunological dysregulation and inflammation play pivotal roles in the pathogenesis of fibrosis in SSc. Nonetheless, remission or cure of SSc are still unachievable even with these therapies, and there is a significant need for improved understanding of the underlying pathogenesis of SSc and developing new therapeutic strategies.

The pathogenesis of SSc is complex. The infiltration of both innate and adaptive immune cells into organs and the cells' activation have been reported in SSc [6,7]. Hematopoietic stem cell transplantation has been used for the treatment of severe SSc [8,9]. Thus, dysregulation of the immune system and bone marrow-derived immune cells play a key role in the pathogenesis of SSc, even after disease progression. Genetic studies suggested that T cells and B cells play pivotal roles [10,11]. Recently, monocyte/macrophages were also reported as key players in fibrosis [12–14]. Therefore, several immune cell subsets are implicated as crucial participants in SSc pathogenesis. However, the disease-specific characteristics of these subsets have not been revealed.

The development of single cell RNA sequencing (scRNA-seq) technologies provided us a way to focus on the characterization of individual cells in human diseases [15,16]. For example, scRNA-seq analysis revealed the subpopulations of invading cells in the synovium of rheumatoid arthritis patients and in the lung of SSc [17,18]. The aim of our present study was to identify the characteristic gene expression patterns and the disease-relevant subpopulations in SSc. Here, we sorted peripheral blood immune cells into 19 subsets and performed RNA-sequencing (RNA-seq) analysis of each cell subset. In addition, we performed scRNA-seq analysis of the identified subsets. By integrating the bulk and single cell transcriptome data, we determined an overall picture of the dysregulated immune system in SSc.

2. Methods and materials

2.1. Patients and healthy volunteers

SSc patients were enrolled in the University of Tokyo Hospital. All of the SSc patients met the American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) 2013 classification criteria [19]. Those cases with malignancy, acute infection, or under treatment with more than 20 mg prednisolone or biologics were excluded. Age- and sex-matched healthy controls (HCs) were also enrolled. All of the patients and HCs were Japanese. We obtained written informed consent from all of the participants. This study was approved by local ethical committees (G10095 and 11828) and performed in accordance with the latest version of the Helsinki declaration.

2.2. Peripheral blood mononuclear cell collection for flow cytometry and cell sorting

Peripheral blood mononuclear cells (PBMCs) were isolated from the peripheral blood of SSc patients and HCs by density gradient centrifugation with Ficoll-Paque Plus (GE Healthcare). Ammonium chloride/potassium (ACK) lysing buffer (pH 7.2, 150 mM NH_4Cl + 10 mM $KHCO_3$ + 0.1 mM Na_2EDTA) was used to lyse red blood cells (RBCs). PBMCs were blocked with human Fc receptor binding inhibitor (eBioscience) and stained with fluorescently labeled antibodies listed in Supplementary Table 1.

Nineteen immune cell subsets were separated with a cell sorter (MoFlo XDP, Beckman Colter). The definition of each immune cell is listed in Supplementary Table 2. Five thousand cells for each subset were sorted for bulk samples and lysed in RLT Buffer (QIAGEN) containing 1% 2-mercaptoethanol.

2.3. RNA-sequencing (RNA-seq) for bulk samples

Full-length complementary DNA (cDNA) and sequencing libraries were prepared using SMART-seq v4 Ultra Low Input RNA kit for Sequencing (TaKaRa) in accordance with the manufacturer's protocol. Prepared libraries were sequenced on HiSeq2500 (Illumina) to generate 100 base paired-end reads. To perform QC of the sequencing data, the adaptor sequences and 3′ low quality bases (Phred quality score < 20) were trimmed with Cutadapt. Short reads (<50bp) and reads containing many low quality bases (Phred quality score <20 in >20% of the bases) were removed. Reads were mapped to hg38 (UCSC human genome 38) with STAR [20] and count expression of all transcripts for each gene with HTSeq [21]. If the uniquely mapped rate was less than 80%, or the number of uniquely mapped reads was 5.00×10^6 reads, the sample was removed before further analysis. The average of the uniquely mapped rates was 95.9%, and the average number of uniquely mapped reads was 1.01×10^7 read pairs.

For quality control of RNA-seq analysis data of bulk samples, we calculated the correlation coefficient of the expression data between two samples belonging to the same cell subset and calculated the average of the correlation coefficient (Di). Samples for which Di was less than 0.9 were removed. We analyzed 627 samples after quality control.

2.4. Transcriptome analysis of bulk samples

We mapped t-distributed stochastic neighbor embedding (t-SNE) plots using R package RtSNE (https://CRAN.R-project.org/package=Rtsne). Differential expression analysis was performed with edgeR [22] after TMM normalization and removing batch effects with RUVSeq [23]. Differential expression genes (DEGs) were detected with false discovery rate (FDR) < 0.05. Upstream regulator analysis was performed with Ingenuity Pathways Analysis (IPA, Ingenuity) to identify the upstream factors of the gene sets.

Iterative weighted gene correlation network analysis (iWGCNA) was performed with WGCNA [24] and iWGCNA (https://github.com/cstoeckert/iterativeWGCNA). Using iWGCNA, we found modules of strongly correlated genes and calculated principal component 1, the representative expression value of each module for so-called "eigengenes". We calculated the correlation coefficient between each eigengene and clinical parameters. When one gene expression value had a higher correlation with the eigengene, the gene was defined as a hub gene regulating the module gene's expression.

To select the important gene modules for distinguishing between SSc and HC, random forest, a machine learning algorithm, was used. We calculated the average of decreased Gini index 100 times. We visualized the module correlation network with igraph R package (https://igraph.org/).

2.5. Single cell (sc) RNA-seq analysis of monocyte subsets with ICELL8 single-CELL system

As mentioned above, peripheral $CD14^+CD3^-CD19^-$ monocytes were sorted regardless of CD16 positivity and suspended in STEM-CELLBANKER GMP grade (ZENOAQ) for storage at −80 °C. After the samples thawed, they were resuspended in GT-T551 (TaKaRa). scRNA-seq analysis was performed with an ICELL8 Single-CELL System in accordance with the manufacturer's protocol [25]. Briefly, the samples were dispensed into each well on an ICELL8 Chip and a single living cell sample was selected with live cell staining. cDNA libraries were prepared using 3′-seq and sequencing was performed by Hiseq2500. The sequencing primer, poly-A, and poly-T were trimmed by Cutadapt and low-quality ends (phred quality score < 25) were also trimmed. After

trimming, the read (the length of which was <20 bases) was also removed. Mapping was performed with STAR [26] with the Ensembl GRCh38 reference sequence. Single cell data were analyzed with R package SEURAT (https://satijalab.org/seurat/). The expression data were normalized as: ln((one gene count/all genes counts) x 10^4) + 1). Clustering was based on the k-nearest neighbor method and data visualization with t-SNE was performed. We defined each cluster signature as the average of the Z-score of cluster marker genes that were expressed in more than 25% of cells in one cluster and the expression was higher in one cluster than in others (p-values < 0.01). Public lung scRNA-seq data from 4 SSc-associated interstitial lung disease patients (8 samples) were downloaded from GEO (GSE128169, http://www.ncbi.nlm.nih.gov/geo/). Integrative analysis of our peripheral blood monocyte data and the public lung data was performed with Seurat's IntegrateData function to minimize batch effects between data [27]. Nineteen clusters of integrated single cells were identified, and 2 monocyte clusters (L4, L11) and 4 macrophage clusters (L0, L2, L14, L17) were selected for visualization based on canonical markers, including A1F1 for macrophages [18].

2.6. Statistics

Categorical data were tested with Fisher's exact test. Differences between two groups of normally and non-normally distributed continuous data were tested for significance with Student's T tests and nonparametric Mann–Whitney U tests. Correlations were evaluated by Pearson's product moment correlation coefficient or nonparametric Spearman's rank correlation coefficients. P values less than 0.05 were considered significant. Statistical analyses were performed with R version 3.4.1 or 3.5.1 (R Foundation for Statistical Computing).

3. Results

3.1. Demographical and clinical information of study participants

Using 19 sorted immune cell subsets from 21 SSc patients and 13 age-sex-matched HCs, we performed RNA-seq analysis. The demographic information and clinical manifestations of the participants are listed in Table 1. The median ages were 68 (range; 37–82) and 63 (42–80) years for SSc cases and HCs, respectively. The proportions of female were 95.2% in SSc and 76.9% in HCs. Detailed clinical characteristics are summarized in Supplementary Table 3.

For scRNA-seq analysis, 3 SSc patients were included, and their demographical and clinical information are described in Supplementary Table 4. The median age was 69 (range; 42–79), and all were female.

Table 1
Clinical characteristics.

	SSc n = 21	HC n = 13	p-value
Demographics			
age, median (range), years	68 (37–82)	63 (42–80)	0.32
female:male (female %)	20:1 (95.2%)	10:3 (76.9%)	0.27
Clinical parameters			
disease duration (years)	13.0 (0.3–36)		
modified Rodnan total skin thickness score	9.0 (0–26)		
interstitial lung disease (%)	12/21 (57.1%)		
Laboratory tests			
Anti-nuclear antibody positive	21/21 (100%)		
Anti-topoisomerase I antibody positive	9/21 (42.8%)		
Anti-centromere antibody positive	11/21 (52.3%)		
CRP median (range), mg/dl	0.09 (0.02–2.16)		

Mann-Whitney *U* test, or Fisher's exact probability test.
SSc; systemic sclerosis, HC; healthy control.

3.2. The gene module of CD16$^+$ monocytes distinguished SSc from HCs

First, the overview of the peripheral blood immune cell transcriptome was summarized with tSNE (Supplementary Fig. 1). Clustering of each immune cell subset and high expression of key marker genes confirmed the purity and validity of our peripheral blood RNA-seq analysis data. We performed DEGs analysis between SSc and HC in each of the cell subsets. Variable numbers of DEGs were detected in each subset, with CD16$^-$ "classical" monocytes having the most (Fig. 1A and B). Of note, CD16$^-$ "classical" monocytes from SSc patients differentially expressed 3 Egr transcription factors that play important roles in monocyte to macrophage differentiation (Fig. 1B) [28]. Then, upstream regulator analysis of DEGs in each subset was performed with IPA (Fig. 1C). Pro-inflammatory cytokines, such as IL-1β, TNF, the interferons and growth factors PDGF-BB and EGF were identified as candidates with the most activating upstream regulators in most of the immune cell subsets in SSc.

To identify the characteristic gene expression network module that discriminated SSc from HCs, iWGCNA analysis was performed. Subsequently, random forest, a machine learning algorithm, was utilized to select modules to discriminate SSc from HCs. Those modules were ranked according to the mean Gini decrease to prioritize which co-expressed genes were important in the discrimination (Fig. 2A). As a result, CD16$^+$ monocyte module 5 (CD16p_Mono_M5) was ranked first. The top 30 hub genes of CD16p_Mono_M5 included *CXCL8, FOS*, *JUN*, and *NFKBIA* (Supplementary Table 5). The eigengenes, the representative expression value of the module CD16p_Mono_M5, was significantly higher in monocytes from SSc than those from HCs (Fig. 2B). By upstream regulator analysis, pro-inflammatory cytokines and growth factors were identified as the upstream regulators of this module (Fig. 2C). The correlation of clinical parameters and the expression of eigengenes of the CD16$^+$ monocyte modules were also analyzed (Supplementary Fig. 2). CD16p_Mono_M5 correlated with the existence of SSc (correlation coefficient 0.78, $p = 4 \times 10^{-8}$). Therefore, we defined CD16p_Mono_M5 as an "SSc-related module". Importantly, by correlation analysis, the top modules for distinguishing SSc from HCs with random forest were strongly correlated to each other (correlation coefficient > 0.80, Fig. 2D, Supplementary Fig. 3) and upstream regulator analysis revealed almost identical regulators (data not shown). Therefore, the SSc-related module in the CD16$^+$ monocyte subset and co-regulated modules in other immune cells could discriminate SSc from HCs.

In addition, we characterized iWGCNA modules that were correlated with higher age and the presence of anti-topoisomerase I antibody in SSc patients (Supplementary Fig. 4, Supplementary Tables 6 and 7). Higher age was especially associated with gene modules of the naïve CD8 T cell subset. Interferons and T cell receptor (TCR) signaling were identified as upstream regulators of these modules. We also noted a correlation between the presence of anti-topoisomerase I antibody and naïve and DN B cell modules regulated by cytokines, such as interferon alpha or TNF. While statistically limited by the relatively small number of stratified SSc populations, these cell subsets and highly correlated genes reflect the immunological heterogeneity within SSc patients.

3.3. Single cell RNA-sequencing revealed clusters of monocytes in SSc

The results of RNA-seq analysis of bulk sorted samples implied that the gene co-expression network module of monocytes had characteristic pro-inflammatory features in SSc. However, the results obtained with bulk sorted samples could represent a mixture of a variety of monocyte subpopulations. To investigate more specific details of monocyte subsets in SSc, we performed scRNA-seq analysis of SSc monocytes. Towards that end, 878 cell samples were analyzed, and we assessed clustering based on the k-nearest neighbor method. A T-SNE plot of scRNA-seq analysis data indicated that monocyte samples from SSc were divided into 7 clusters (Fig. 3A). Cluster PM0 and PM4 showed marker gene

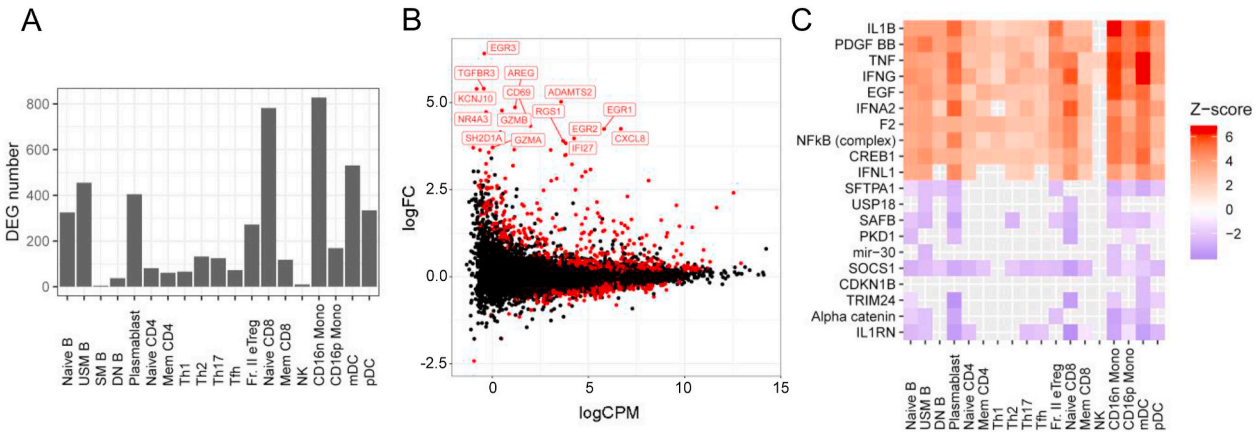

Fig. 1. Pro-inflammatory differential expression of SSc immune cell subsets. (A) Differentially expressed gene (DEG) numbers in SSc were estimated for each cell population (FDR < 0.05). (B) Representative plot of the CD16$^-$ classical monocyte subset (CD16n_Mono). Positive log FC means increased expression in SSc. DEGs are colored red and the top 15 most significant genes are labeled. (C) Knowledge-based estimation of upstream regulator analysis were performed with IPA. Z-score >2 or <2 means significant activation or inhibition. Regulators were sorted in the order of averaged Z-scores, and the top 10 activation and inhibition regulators are shown. Log CPM: $\log_2$ counts per million. Log FC: $\log_2$ fold change. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

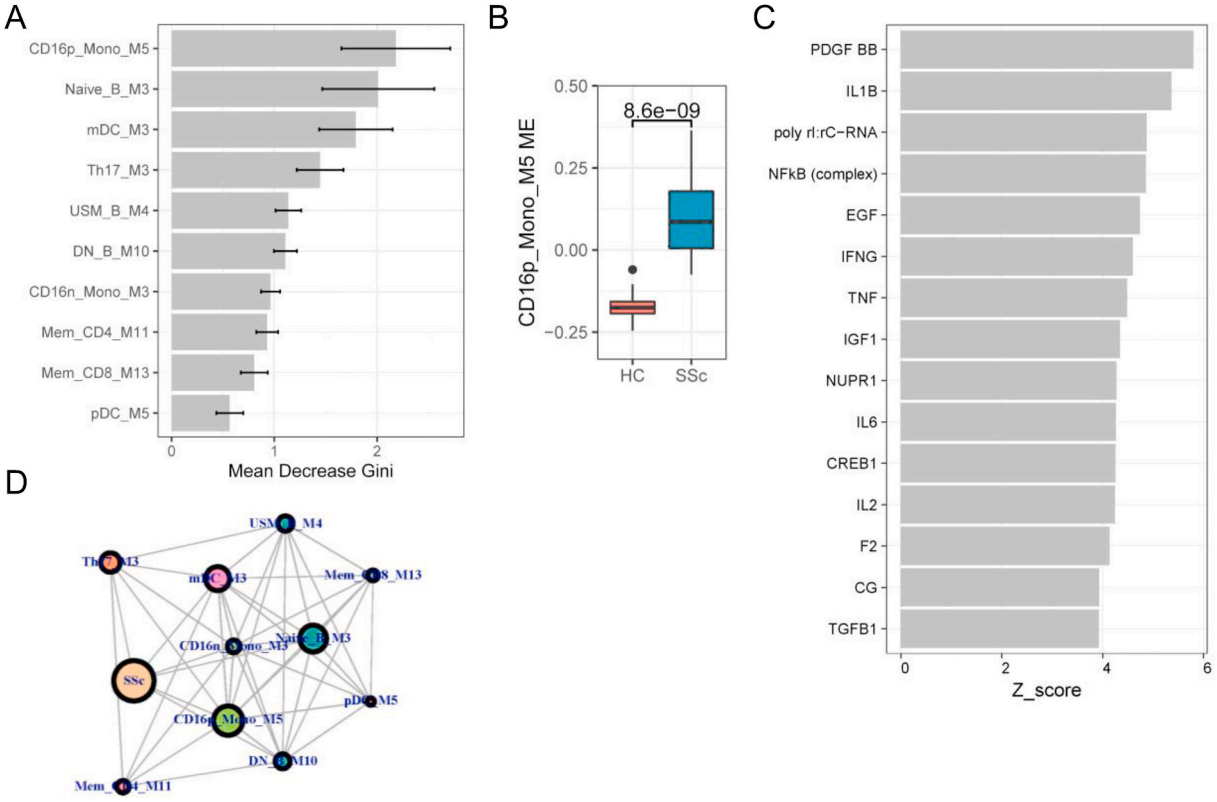

Fig. 2. Inflammatory gene module in CD16$^+$ monocytes is important for the discrimination of SSc. (A) Top 10 modules that discriminate SSc from HC. iWGCNA analysis identified 204 gene modules in 19 immune cell subsets and ranked them by random forest mean decreased Gini index. (B) CD16$^+$ monocyte (CD16p_Mono) M5 module eigengene (ME) expression in HC and SSc. Error bars represent standard deviations. (C) IPA upstream regulator analysis of CD16p_Mono_M5 module genes. Top 15 activation upstream regulators are shown. (D) Correlation network of Top 10 modules (A). Two module nodes with Spearman's correlation coefficients >0.8 are connected with a line. Module size is proportional to the discriminant importance in random forest analysis (A). Overall network is shown in Supplementary Fig. 3.

expressions that suggested that they were distinct monocyte populations with characteristic highly expressed genes (Fig. 3B and C). Other clusters likely reflected intermediate differentiating conditions. By calculating cluster signature scores in bulk-sorted RNA-seq analysis data, we found that the cluster PM0 signature score was significantly higher in SSc CD16$^+$ monocytes than in HCs (Fig. 3D). Intriguingly, SSc-related module genes were highly expressed in cluster PM0 (odds ratio 5.35, $p < 1 \times 10^{-6}$) (Fig. 3E). Also, the DEGs in bulk-sorted CD16$^+$ monocytes significantly overlapped with single cell RNA-seq analysis of cluster PM0 marker genes (Odds ratio 2.94, $p = 0.0036$, Fig. 3E). Nine genes, *KLF10*, *PLAUR*, *JUNB*, *JUND*, *JUN*, *ZFP36*, *IER2*, *PUF60*, and *HLA-DRB5*, were commonly detected in DEGs, SSc-related module genes, and cluster PM0

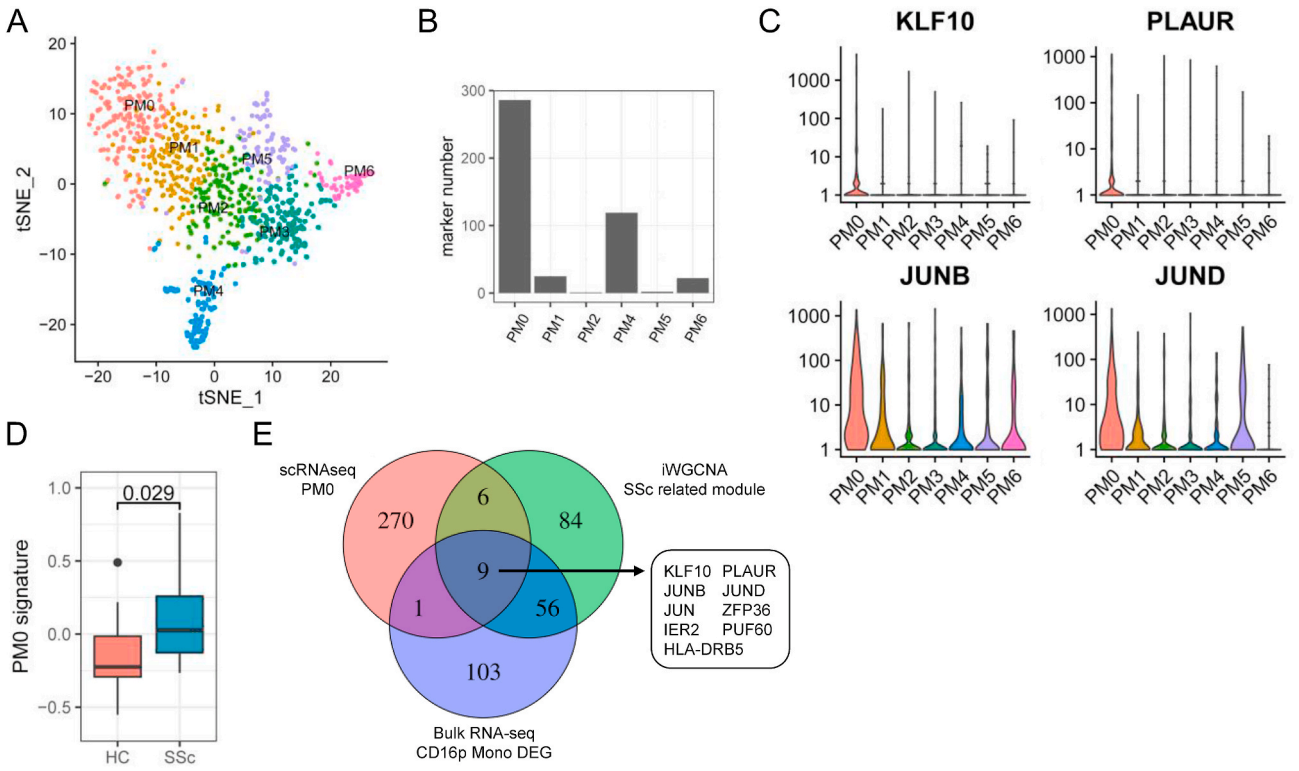

Fig. 3. Single cell RNA-seq analysis of SSc monocytes and its overlap with bulk RNA-seq analysis. (A) T-SNE of 878 single monocyte cells RNA-seq (scRNA-seq) analysis from SSc patients. Cells were clustered by k-nearest neighbor method and labeled accordingly (PM0-6). (B) Marker gene number of each of the 7 clusters of SSc monocytes. (C) Violin plots of selected cluster PM0 marker genes (differential expression p-values < 0.01). (D) Cluster PM0 signature score of bulk-sorted $CD16^{+}$ monocytes. (E) Venn diagram of the scRNA-seq cluster PM0 marker genes, iWGCNA SSc-related module, and $CD16^{+}$ monocyte (CD16p_Mono) differentially expressed genes (DEG).

maker genes (Supplementary Table 8).

3.4. *Integrative analysis of peripheral blood and lung single cell RNA-seq data from SSc patients*

To gain insight into the contributions of the identified peripheral blood monocyte single cell PM0 cluster in the affected organs, we performed integrative analysis with public lung scRNA-seq data from SSc-associated interstitial lung disease patients [18]. As reported by Valenzi et al., infiltrating lung macrophages and monocytes were divided into 3 primary subgroups: $SPP1^{hi}$ and $FABP4^{hi}$ macrophages and $FCN1^{hi}$ monocytes (Fig. 4A). For the peripheral blood monocyte cluster PM0, 57% clustered together with lung $FCN1^{hi}$ monocytes and was most similar to this population in the 7 peripheral blood monocyte clusters (Figs. 3A, 4B and C). The lung $FCN1^{hi}$ monocytes characteristically expressed *IL1B* and *IL1R2* (Supplementary Fig. 5).

4. Discussion

In this study of SSc, RNA-seq analysis of 19 peripheral blood immune cell subsets and scRNA-seq analysis of monocytes were performed. Previously, some transcriptome analysis had been reported in SSc.

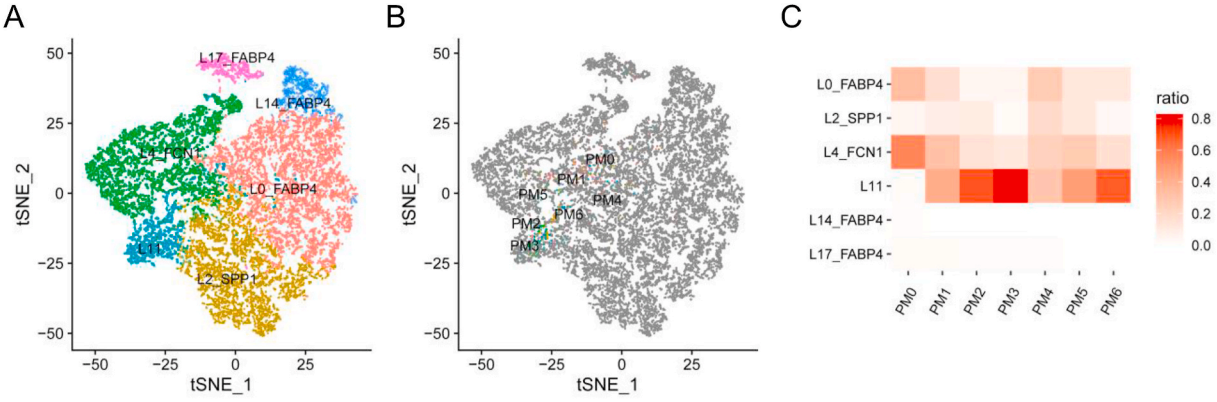

Fig. 4. Integrative analysis of peripheral blood and lung scRNA-seq data. (A, B) T-SNE of 17,254 monocytes/macrophages single cell RNA-seq analyses (scRNA-seq) from SSc patients. Peripheral blood monocyte data were merged with public SSc lung scRNA-seq data and 6 monocyte/macrophage clusters out of 19 clusters were selected for visualization. Cells were labeled with integrated monocyte ($L4\_FCN1^{hi}$, L11) and macrophage ($L0\_FABP4^{hi}$, $L2\_SPP1^{hi}$, $L14\_FABP4^{hi}$, $L17\_FABP4^{hi}$) clusters (A) and peripheral blood monocyte clusters as in Fig. 3 (B). (C) Overlap between peripheral blood monocyte clusters (PM0-6) and integrated monocyte/macrophage clusters.

Assassi et al. performed transcriptome analysis of skin biopsies and reported that the macrophage signature was the most prominent inflammatory signature in SSc skin [29]. They reported that the majority of samples with significantly higher fibroblast scores had significantly higher macrophage scores. Taroni et al. compared gene expression profiles of biopsies from 4 SSc-affected tissues, skin, lung, esophagus and peripheral blood. They found profibrotic-macrophage signatures in multiple tissues [30]. However, these analyses were based on heterogeneous biopsy samples, and therefore greatly influenced by the proportion of cell populations. Our sorted immune cell transcriptome analysis is a different (but complimentary) approach to directly assessing qualitative differences among immune cell subsets, and it showed the importance of inflammatory gene alterations in monocytes in the pathophysiology of SSc.

The gene module in $CD16^+$ monocytes was the most prominent characteristic in SSc in our analysis. $CD16^+$ monocytes are mainly "non-classical" monocytes, and they are considered to be pro-inflammatory due to their production of pro-inflammatory cytokines [31–33]. Lescoat et al. reported that the number of circulating $CD16^+$ monocytes was increased in SSc compared to that in HCs and that they were associated with fibrotic manifestations [34]. Scott et al. reported increased monocyte counts as a cellular biomarker for poor outcomes in fibrotic diseases, including SSc [13]. These reports show that the quantitative increase of monocytes, especially $CD16^+$ pro-inflammatory monocytes is associated with the fibrotic phenotype of SSc. Also, Moreno-Moral and Bagnati et al. reported the enrichment of SSc susceptibility genes in differentially regulated genes in monocyte-derived macrophages of SSc [12]. This result suggested that these qualitative alterations of gene expressions in SSc monocytes/macrophages have a genetic basis, and are causally associated with the pathogenesis of SSc.

We identified a characteristic gene network module and a cluster of monocytes in SSc. As monocytes are key producers of pro-inflammatory cytokines, such as IL-1, IL-6 and TNF-α, they could contribute to SSc pathophysiology [35–37]. Another possibility is that a subset of monocytes migrates to the periphery and contributes to fibrosis directly. In mice, a fraction of the monocytes or macrophages are involved in tissue fibrosis [14,38,39]. Integrative analysis using lung scRNA-seq data revealed transcriptomic similarities between peripheral blood scRNA-seq cluster PM0 and lung *IL1B*$^+$ FCN1^{hi} monocytes. These data suggest a link between circulating PM0 monocytes and lung *IL1B*$^+$ FCN1^{hi} monocytes. *IL1B*$^+$ FCN1^{hi} monocytes could contribute to local fibrosis not only in the lung but also in the other fibrotic organs such as skin. Of note, *IL1B* is a marker of a pro-fibrotic intermediate between monocytes and alveolar macrophages [38]. Moreover, IL-1Ra-KO mice showed enhanced IL-1 signaling and more severe skin and lung fibrosis in a bleomycin-induced SSc model [36]. Therefore, IL-1β itself could be one of the key pro-fibrotic mediators. Thus far, a small phase 1/2 biomarker trial of IL1-trap, rilonacept, has not reported any treatment-related effect on SSc biomarkers [40] and a phase 2 trial of the IL-1α inhibitor Bermekimab is ongoing (ClinicalTrials.gov Identifier: NCT04045743).

Among the overlapping genes identified in the DEGs, SSc-related module genes and scRNA-seq cluster PM0 signature genes (Fig. 3E and Supplementary Table 8), *KLF10* is reported to mediate TGF-β signaling and modulate fibrosis [41,42]. *PLAUR* (UPAR) gene polymorphism is associated with SSc and its vascular phenotype and SSc-like histopathologic features are reported in uPAR-deficient mice [43]. Also, *JUNB* and *JUND* contribute to fibroblast activation in SSc [44,45]. In addition, the *HLA-DRB5* transcript was reported to be the only gene commonly upregulated in SSc patients with interstitial lung disease [46]. Therefore, we suggest that these overlapping SSc-related monocyte gene lists offer candidate therapeutic targets for SSc. Further functional studies will be required to validate them.

We also observed elevated expression of interferon-induced gene expression in immune cells from SSc patients (Fig. 1C). The interferon signature has been implicated in the pathophysiology of systemic sclerosis and other autoimmune diseases [47,48]. In a mouse model of SSc, disease was prevented by the depletion of plasmacytoid dendritic cells that produce type I interferons [49]. A randomized, double-blind, placebo-controlled trial showed that recombinant IFNα therapy might be deleterious in SSc because those who received the IFNα treatment showed less improvement in skin scores and greater deterioration of lung function than the placebo group [50]. Therefore, the interferons and their regulatory mechanisms are attractive candidate therapeutic targets for SSc.

The limitations of this study include the use of an established SSc patient cohort with relatively long disease duration that might have hindered the identification of early phase immunological disturbances. However, with the intractable and life-long nature of SSc, identification of pro-inflammatory immune disturbances in established SSc has potentially therapeutic meanings. Our scRNA-seq analysis was limited by a relatively small sample size and cell number. Additionally, functional elucidation of the identified monocyte cluster and gene module is needed to clarify their roles in SSc pathophysiology.

5. Conclusions

In summary, our integrated analysis of bulk and single cell RNA-seq identified an inflammatory gene module and a cluster of monocytes that are associated with SSc. We suggest that they could serve as novel therapeutic targets in SSc.

Author contributions

S.K., Y.N., H.Shoda, and K.F. conceived, designed and analyzed the experiments and contributed to writing the manuscript. M.Okubo, Y.S., H.Shirai., H.H., J.M., H.Y., Y.T., M.Ota, Y.I., and S.S. helped with obtaining RNA sequencing data. H.H., H.Y., Y.T., and M.Ota contributed to RNA-seq data analysis. T.O., K.Y., H.Shoda, and K.F supervised the study. All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published.

Author statement

S.K. and Y.N. Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft. H.Shoda, and K.F. Supervision, Conceptualization, Methodology, Writing - Review & Editing. S.K. M. Okubo, Y.S., H.Shirai., H.H., J.M., H.Y., Y.T., M.Ota, Y.I., and S.S. Resources, Data Curation. H.H., H.Y., Y.T., and M.Ota Software. T.O., K.Y., Supervision.

Funding

This research was performed in collaboration with Chugai Pharmaceutical Co., Ltd. It was supported by a Grant-in-Aid for Young Scientists (B) (17K16204) and a Grant-in-Aid for Scientific Research (B) (19H03697) from the Japan Society for the Promotion of Science, Ministry of Health, Labor and Welfare and the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Data availability statement

The peripheral blood cell subsets RNA-seq data and monocyte single cell RNA-seq data are deposited in National Bioscience Database Center (NBDC) Human Database (http://humandbs.biosciencedbc.jp/) with the accession number of hum0214, JGAD00000000309 and hum0218, E-GEAD-344. The data are available upon reasonable request.

Declaration of competing interest

Y. Nagafuchi has received financial support or fees from BMS,

Chugai, Kissei, Mitsubishi Tanabe, and Pfizer. M. Ota has received speaking fees from BMS, MSD and Chugai. S. Sumitomo received financial support or fees from AbbVie, Eisai, Chugai, UCB, BMS, Takeda, AstraZeneca, Pfizer, MitsubishiTanabe, Daiichi-Sankyo, Novartis, and Asahi Kasei. T. Okamura received financial support or fees from Chugai, Novartis Pharma. K. Yamamoto received financial support or fees from AbbVie, Astellas, BMS, Daiichi-Sankyo, MitsubishiTanabe, Pfizer, Sanofi, Santen, Takeda, Teijin, Boehringer Ingelheim, Chugai, Eisai, Ono, Taisho Toyama, UCB, ImmunoFuture, Asahi Kasei, and Janssen. H. Shoda has received speaking fees from Takeda, BMS, Asahi Kasei, Sanofi, Eli Lilly, Janssen, Astellas, Pfizer, Chugai, UCB, and Daiichi-Sankyo. K. Fujio has received grants, consulting fees, speaking fees, and/or honoraria from Takeda, BMS, Mitsubishi Tanabe, Asahi Kasei, Sanofi, Eli Lilly, Daiichi-Sankyo, Ono, Janssen, AbbVie, Astellas, Eisai, Pfizer, Chugai, Novartis, UCB, Tsumura, Taisho Toyama, Nihon Kayaku and Ayumi. All the other authors declared no conflicts of interest. Department of Functional Genomics and Immunological Diseases, Graduate School of Medicine, The University of Tokyo was supported by Chugai Pharmaceutical Co., Ltd.

Acknowledgements

We thank all the study participants and all the members of the Department of Allergy and Rheumatology for the collection of clinical data. The supercomputing resource SHIROKANE was provided by the Human Genome Center at The University of Tokyo.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jaut.2020.102547.

References

[1] C.P. Denton, D. Khanna, Systemic sclerosis, Lancet 390 (2017) 1685–1699.
[2] J. Varga, D. Abraham, Systemic sclerosis: a prototypic multisystem fibrotic disorder, J. Clin. Invest. 117 (2007) 557–567.
[3] A.J. Tyndall, B. Bannert, M. Vonk, P. Airo, F. Cozzi, P.E. Carreira, et al., Causes and risk factors for death in systemic sclerosis: a study from the EULAR Scleroderma Trials and Research (EUSTAR) database, Ann. Rheum. Dis. 69 (2010) 1809–1815.
[4] E.R. Volkmann, D.P. Tashkin, N. Li, M.D. Roth, D. Khanna, A.M. Hoffmann-Vold, et al., Mycophenolate mofetil versus placebo for systemic sclerosis-related interstitial lung disease: an analysis of scleroderma lung studies I and II, Arthritis Rheum. 69 (2017) 1451–1460.
[5] D. Khanna, C.P. Denton, C.J.F. Lin, J.M. van Laar, T.M. Frech, M.E. Anderson, et al., Safety and efficacy of subcutaneous tocilizumab in systemic sclerosis: results from the open-label period of a phase II randomised controlled trial (faSScinate), Ann. Rheum. Dis. 77 (2018) 212–220.
[6] L.I. Sakkas, I.C. Chikanza, C.D. Platsoucas, Mechanisms of Disease: the role of immune cells in the pathogenesis of systemic sclerosis, Nat. Clin. Pract. Rheumatol. 2 (2006) 679–685.
[7] G. Kania, M. Rudnik, O. Distler, Involvement of the myeloid cell compartment in fibrogenesis and systemic sclerosis, Nat. Rev. Rheumatol. 15 (2019) 288–302.
[8] R.K. Burt, S.J. Shah, K. Dill, T. Grant, M. Gheorghiade, J. Schroeder, et al., Autologous non-myeloablative haemopoietic stem-cell transplantation compared with pulse cyclophosphamide once per month for systemic sclerosis (ASSIST): an open-label, randomised phase 2 trial, Lancet 378 (2011) 498–506.
[9] J.M. van Laar, D. Farge, J.K. Sont, K. Naraghi, Z. Marjanovic, J. Larghero, et al., Autologous hematopoietic stem cell transplantation vs intravenous pulse cyclophosphamide in diffuse cutaneous systemic sclerosis: a randomized clinical trial, J. Am. Med. Assoc. 311 (2014) 2490–2498.
[10] C. Terao, T. Kawaguchi, P. Dieude, J. Varga, M. Kuwana, M. Hudson, et al., Transethnic meta-analysis identifies GSDMA and PRDM1 as susceptibility genes to systemic sclerosis, Ann. Rheum. Dis. 76 (2017) 1150–1158.
[11] G. Murdaca, M. Contatore, R. Gulli, P. Mandich, F. Puppo, Genetic factors and systemic sclerosis, Autoimmun. Rev. 15 (2016) 427–432.
[12] A. Moreno-Moral, M. Bagnati, S. Koturan, J.H. Ko, C. Fonseca, N. Harmston, et al., Changes in macrophage transcriptome associate with systemic sclerosis and mediate GSDMA contribution to disease risk, Ann. Rheum. Dis. 77 (2018) 596–601.
[13] M.K.D. Scott, K. Quinn, Q. Li, R. Carroll, H. Warsinske, F. Vallania, et al., Increased monocyte count as a cellular biomarker for poor outcomes in fibrotic diseases: a retrospective, multicentre cohort study, Lancet Respir. Med. 7 (2019) 497–508.
[14] T. Satoh, K. Nakagawa, F. Sugihara, R. Kuwahara, M. Ashihara, F. Yamane, et al., Identification of an atypical monocyte and committed progenitor involved in fibrosis, Nature 541 (2017) 96–101.
[15] A.A. Kolodziejczyk, J.K. Kim, V. Svensson, J.C. Marioni, S.A. Teichmann, The technology and biology of single-cell RNA sequencing, Mol. Cell 58 (2015) 610–620.
[16] C. Ziegenhain, B. Vieth, S. Parekh, B. Reinius, A. Guillaumet-Adkins, M. Smets, et al., Comparative analysis of single-cell RNA sequencing methods, Mol. Cell 65 (2017), 631-43.e4.
[17] F. Zhang, K. Wei, K. Slowikowski, C.Y. Fonseka, D.A. Rao, S. Kelly, et al., Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry, Nat. Immunol. (2019).
[18] E. Valenzi, M. Bulik, T. Tabib, C. Morse, J. Sembrat, H. Trejo Bittar, et al., Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease, Ann. Rheum. Dis. 78 (2019) 1379–1387.
[19] F. van den Hoogen, D. Khanna, J. Fransen, S.R. Johnson, M. Baron, A. Tyndall, et al., Classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative, Arthritis Rheum. 65 (2013) 2737–2747, 2013.
[20] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21.
[21] S. Anders, P.T. Pyl, W. Huber, HTSeq–a Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2015) 166–169.
[22] M.D. Robinson, D.J. McCarthy, G. K. Smyth edgeR, A Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
[23] D. Risso, J. Ngai, T.P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control genes or samples, Nat. Biotechnol. 32 (2014) 896–902.
[24] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.
[25] L.D. Goldstein, Y.J. Chen, J. Dunne, A. Mir, H. Hubschle, J. Guillory, et al., Massively parallel nanowell-based single-cell gene expression profiling, BMC Genom. 18 (2017) 519.
[26] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21.
[27] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W.M. Mauck 3rd, et al., Comprehensive integration of single-cell data, Cell 177 (2019), 1888-902.e21.
[28] T. Veremeyko, A.W.Y. Yung, D.C. Anthony, T. Strekalova, E.D. Ponomarev, Early growth response gene-2 is essential for M1 and M2 macrophage activation and plasticity by modulation of the transcription factor CEBPbeta, Front. Immunol. 9 (2018) 2515.
[29] S. Assassi, W.R. Swindell, M. Wu, F.D. Tan, D. Khanna, D.E. Furst, et al., Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis, Arthritis Rheum. 67 (2015) 3016–3026.
[30] J.N. Taroni, C.S. Greene, V. Martyanov, T.A. Wood, R.B. Christmann, H.W. Farber, et al., A novel multi-network approach reveals tissue-specific cellular modulators of fibrosis in systemic sclerosis, Genome Med. 9 (2017) 27.
[31] L. Ziegler-Heitbrock, P. Ancuta, S. Crowe, M. Dalod, V. Grau, D.N. Hart, et al., Nomenclature of monocytes and dendritic cells in blood, Blood 116 (2010) e74–80.
[32] J. Yang, L. Zhang, C. Yu, X.F. Yang, H. Wang, Monocyte and macrophage differentiation: circulation inflammatory monocyte as biomarker for inflammatory diseases, Biomark Res 2 (2014) 1.
[33] C. Auffray, D. Fogg, M. Garfa, G. Elain, O. Join-Lambert, S. Kayal, et al., Monitoring of blood vessels and tissues by a population of monocytes with patrolling behavior, Science 317 (2007) 666–670.
[34] A. Lescoat, V. Lecureur, M. Roussel, B.L. Sunnaram, A. Ballerie, G. Coiffier, et al., CD16-positive circulating monocytes and fibrotic manifestations of systemic sclerosis, Clin. Rheumatol. 36 (2017) 1649–1654.
[35] H. Umehara, S. Kumagai, M. Murakami, T. Suginoshita, K. Tanaka, S. Hashida, et al., Enhanced production of interleukin-1 and tumor necrosis factor alpha by cultured peripheral blood monocytes from patients with scleroderma, Arthritis Rheum. 33 (1990) 893–897.
[36] M.J. Park, S.J. Moon, E.J. Lee, K.A. Jung, E.K. Kim, D.S. Kim, et al., IL-1-IL-17 signaling Axis contributes to fibrosis and inflammation in two different murine models of systemic sclerosis, Front. Immunol. 9 (2018) 1611.
[37] C. Muangchan, J.E. Pope, Interleukin 6 in systemic sclerosis and potential implications for targeted therapy, J. Rheumatol. 39 (2012) 1120–1124.
[38] D. Aran, A.P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, et al., Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage, Nat. Immunol. 20 (2019) 163–172.
[39] S. Chakarov, H.Y. Lim, L. Tan, S.Y. Lim, P. See, J. Lum, et al., Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches, Science (2019) 363.
[40] J.C. Mantero, N. Kishore, J. Ziemek, G. Stifano, C. Zammitti, D. Khanna, et al., Randomised, double-blind, placebo-controlled trial of IL1-trap, rilonacept, in systemic sclerosis. A phase I/II biomarker trial, Clin. Exp. Rheumatol. 36 (Suppl 113) (2018) 146–149.
[41] B. Spittau, K. Krieglstein, Klf10 and Klf11 as mediators of TGF-beta superfamily signaling, Cell Tissue Res. 347 (2012) 65–72.
[42] J.X. DiMario, KLF10 gene expression modulates fibrosis in dystrophic skeletal muscle, Am. J. Pathol. 188 (2018) 1263–1275.
[43] M. Manetti, Y. Allanore, L. Revillod, C. Fatini, S. Guiducci, G. Cuomo, et al., A genetic variation located in the promoter region of the UPAR (CD87) gene is associated with the vascular complications of systemic sclerosis, Arthritis Rheum. 63 (2011) 247–256.
[44] K. Palumbo, P. Zerr, M. Tomcik, S. Vollath, C. Dees, A. Akhmetshina, et al., The transcription factor JunD mediates transforming growth factor {beta}-induced

fibroblast activation and fibrosis in systemic sclerosis, Ann. Rheum. Dis. 70 (2011) 1320–1326.

[45] M. Ponticos, I. Papaioannou, S. Xu, A.M. Holmes, K. Khan, C.P. Denton, et al., Failed degradation of JunB contributes to overproduction of type I collagen and development of dermal fibrosis in patients with systemic sclerosis, Arthritis Rheum. 67 (2015) 243–253.

[46] T. Odani, S. Yasuda, Y. Ota, Y. Fujieda, Y. Kon, T. Horita, et al., Up-regulated expression of HLA-DRB5 transcripts and high frequency of the HLA-DRB5*01:05 allele in scleroderma patients with interstitial lung disease, Rheumatology 51 (2012) 1765–1774.

[47] T.L.W. Muskardin, T.B. Niewold, Type I interferon in rheumatic diseases, Nat. Rev. Rheumatol. 14 (2018) 214–228.

[48] F.J. Barrat, M.K. Crow, L.B. Ivashkiv, Interferon target-gene expression and epigenomic signatures in health and disease, Nat. Immunol. 20 (2019) 1574–1583.

[49] M.D. Ah Kioon, C. Tripodo, D. Fernandez, K.A. Kirou, R.F. Spiera, M.K. Crow, et al., Plasmacytoid dendritic cells promote systemic sclerosis with a key role for TLR8, Sci. Transl. Med. 10 (2018).

[50] C.M. Black, A.J. Silman, A.I. Herrick, C.P. Denton, H. Wilson, J. Newman, et al., Interferon-alpha does not improve outcome at one year in patients with diffuse cutaneous scleroderma: results of a randomized, double-blind, placebo-controlled trial, Arthritis Rheum. 42 (1999) 299–305.