








Three Distinct Transcriptional Profiles of Monocytes Associate with Disease Activity in Scleroderma Patients

Hadijat-Kubura M. Makinde,¹ Julia L. M. Dunn,² Gaurav Gadhvi,¹ Mary Carns,¹ Kathleen Aren,¹ Anh H. Chung,¹ Lutfiyya N. Muhammad,³ Jing Song,³  Carla M. Cuda,¹  Salina Dominguez,¹ John E. Pandolfino,⁴ Jane E. Dematte D'Amico,⁵ G. Scott Budinger,⁵ Shervin Assassi,⁶  Tracy M. Frech,⁷  Dinesh Khanna,⁸  Alex Shaeffer,¹ Harris Perlman,¹  Monique Hinchcliff,⁹  and Deborah R. Winter¹

Objective. Patients with diffuse cutaneous systemic sclerosis (dcSSc) display a complex clinical phenotype. Transcriptional profiling of whole blood or tissue from patients are affected by changes in cellular composition that drive gene expression and an inability to detect minority cell populations. We undertook this study to focus on the 2 main subtypes of circulating monocytes, classical monocytes (CMs) and nonclassical monocytes (NCMs) as a biomarker of SSc disease severity.

Methods. SSc patients were recruited from the Prospective Registry for Early Systemic Sclerosis. Clinical data were collected, as well as peripheral blood for isolation of CMs and NCMs. Age-, sex-, and race-matched healthy volunteers were recruited as controls. Bulk macrophages were isolated from the skin in a separate cohort. All samples were assayed by RNA sequencing (RNA-seq).

Results. We used an unbiased approach to cluster patients into 3 groups (groups A–C) based on the transcriptional signatures of CMs relative to controls. Each group maintained their characteristic transcriptional signature in NCMs. Genes up-regulated in group C demonstrated the highest expression compared to the other groups in SSc skin macrophages, relative to controls. Patients from groups B and C exhibited worse lung function than group A, although there was no difference in SSc skin disease at baseline, relative to controls. We validated our approach by applying our group classifications to published bulk monocyte RNA-seq data from SSc patients, and we found that patients without skin disease were most likely to be classified as group A.

Conclusion. We are the first to show that transcriptional signatures of CMs and NCMs can be used to unbiasedly stratify SSc patients and correlate with disease activity outcome measures.

INTRODUCTION

Systemic sclerosis/scleroderma (SSc) is a complex, autoimmune-mediated connective tissue disease characterized

by widespread vascular damage, chronic inflammation, and fibrosis. SSc is highly heterogeneous in presentation and disease course, and may involve the skin, blood vessels, lungs, heart, kidneys, and gastrointestinal tract. SSc can be subtyped by

Dr. Makinde's work was supported by Northwestern University (grants KL2-TR001424, HL134375S1, and AR007611). Dr. Muhammad's and Dr. Song's work was supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases [NIAMS] grant P30-AR072579). Dr. Cuda's work was supported by the Lupus Research Alliance (Novel Research Grant), the Rheumatology Research Foundation (Innovative Research Grant), and Northwestern University Dixon Translational Research Initiative. Dr. Pandolfino's work was supported by the NIH (grant P01-DK-117824). Dr. Budinger's work was supported by the NIH (grants U19-AI-135964, P01-AG-049665, P01-AG-04966506S1, and R01-HL-147575) and Veterans Affairs (grant I01-CX-001777). Dr. Assassi's work was supported by the NIH (NIAMS grants R01-AR-073284 and R56-AR-078211) and the National Scleroderma Foundation. Dr. Khanna's work and the PRESS clinical database were supported by the NIH (NIAMS grant K24-AR-063120). Dr. Perlman's work was supported by the NIH (grants AR-074902, AR-075423, CA-060553, and HL-134375), the Rheumatology Research Foundation Innovative Research Grant, the United States-Israel Binational Science Foundation Investigator Grant, the Precision Medicine Fund, and the Mabel Green Myers Professorship. Dr. Hinchcliff's work was supported by the NIH (grants R01-AR-073270 and R01-HL-152677). Dr. Winter's work was supported by the Arthritis National Research Foundation, the American Federation for Aging Research, the

American Heart Association (grant 18CDA34110224), the National Scleroderma Foundation, the American Thoracic Society, and the NIH (grant AI-163742).

¹Hadijat-Kubura M. Makinde, PhD, Gaurav Gadhvi, MS, Mary Carns, MS, Kathleen Aren, MPH, Anh H. Chung, BS, Carla M. Cuda, PhD, Salina Dominguez, MS, Alex Shaeffer, MD, Harris Perlman, PhD, Deborah R. Winter, PhD: Division of Rheumatology, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; ²Julia L. M. Dunn PhD: Division of Rheumatology, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, and Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; ³Lutfiyya N. Muhammad, PhD, Jing Song, MS: Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; ⁴John E. Pandolfino, MD: Division of Gastroenterology and Hepatology, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; ⁵Jane E. Dematte D'Amico, MD, G. Scott Budinger, MD: Division of Division of Pulmonary and Critical Care, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; ⁶Shervin Assassi, MD, MS: Prospective Registry of Early Systemic Sclerosis (PRESS) consortium, and Division of Rheumatology, University of Texas Health Science Center at

cutaneous involvement, including limited cutaneous SSc (lcSSc), diffuse cutaneous SSc (dcSSc), and noncutaneous SSc (ncSSc, more recently known as SSc sine scleroderma). While clinically defined SSc cutaneous subsets are associated with internal organ complications and death, they cannot predict disease course on a per-patient basis or inform treatment decisions. Thus, identification of disease endophenotypes and their longitudinal course will facilitate precision medicine implementation.

The role of immune cells in SSc pathogenesis is supported by the fact that the vast majority of genetic susceptibility loci in SSc belong to immunologic pathways (1). Circulating monocytes exist in 3 main states, characterized by CD14 and CD16 in humans: classical monocytes (CMs) (CD14+CD16-), which constitute the majority of monocytes; intermediate monocytes (IMs) (CD14+CD16+); and nonclassical monocytes (NCMs) (CD14+CD16+), which constitute <5% of total monocytes. However, recent studies have identified additional monocyte subsets with definitions beyond expression of CD14 and CD16 (2,3). The numbers of CD14+ monocytes in the blood of patients with SSc are higher compared to healthy controls (4,5) and are associated with reduced survival (5). Our group was the first to identify a causal role for monocyte-derived alveolar macrophages in a murine model of fibrosis and to confirm the presence of these cells in humans with SSc-related interstitial lung disease (6,7). Increased numbers of dermal macrophages have also been identified in skin from SSc patients (5,8) and in murine models of SSc-like disease (9). Moreover, reductions in skin macrophage numbers have been detected in a subset of SSc patients receiving mycophenolate mofetil (MMF) (10). These data suggest that the monocyte/macrophage population is crucial for SSc development. However, little is known regarding individual populations of monocytes in SSc.

There are numerous studies that identify relationships between transcriptional signatures from peripheral blood mononuclear cells (PBMCs) or whole skin of SSc patients and disease activity (5,11). We and others identified 4 molecular pathway-centric “intrinsic SSc subsets” using whole-genome microarray analysis of PB, whole skin, and esophageal biopsies from SSc patients (12–15), which were validated by 2 groups (13,16). Moreover, patients with an inflammatory gene expression signature in skin are the most likely to demonstrate skin disease improvement during MMF treatment (17). However, there were patients classified in the inflammatory intrinsic subset who were MMF nonresponders (17,18), which suggests that an understanding of the

transcriptional profile of individual immune populations may be necessary to discern which patients are more susceptible to a particular therapy.

In this study, we profiled the transcriptional signature of CMs and NCMs in patients from the Prospective Registry of Early Systemic Sclerosis (PRESS) consortium, a multicenter incident cohort study of patients with diffuse cutaneous SSc, which has the goal of advancing the understanding of disease pathogenesis and identifying novel biomarkers to inform patient care (19–24). Our study demonstrates the potential for transcriptional profiling of individual monocyte populations as a useful tool to stratify patients prior to therapeutic intervention.

PATIENTS AND METHODS

Study participants, recruitment, and sample collection. Patients were recruited to the PRESS registry at Northwestern University (NU), University of Texas, University of Utah, and University of Michigan (protocol no. STU00062447), as previously described (22). Consecutive patients in the PRESS registry from 4 PRESS sites (University of Michigan, NU, University of Texas-Houston, and University of Utah) were enrolled in this substudy if they were naive to immunosuppressant treatment with plans to start a new treatment for SSc from September 2015 through June 2016, although some samples were collected at posttreatment initiation. At the time of sample collection, 7 out of 14 patients were receiving immunosuppressants: 6 patients were receiving MMF and 1 patient was receiving rituximab and prednisone. Clinical information including modified Rodnan skin score (MRSS) and pulmonary function tests were obtained. Age-, sex-, and ethnicity-matched controls were recruited at NU through the Control Blood Acquisition for Rheumatology Research (no. STU00045513) protocol. For skin biopsies, patients and healthy controls were recruited independently from NU (no. STU00002669). Two side-by-side 4-mm dermal punch biopsies from the dorsal surface of the forearm midway between the ulnar styloid and the olecranon were sampled.

Blood processing and fluorescence-activated cell sorting (FACS). Blood and tissue were processed, stained, and sorted via FACS using an antibody cocktail including BB515-conjugated anti-CD45 (BD Horizon), PerCP-Cyanine5.5-conjugated anti-CD14 (eBioscience), eFluor 450-conjugated anti-HLA-DR (eBioscience), allophycocyanin (APC)-conjugated anti-CD15

Houston; ⁷Tracy M. Frech, MD, MS: PRESS consortium, and Division of Rheumatology and Immunology, Department of Medicine, Vanderbilt University, Nashville, Tennessee; ⁸Dinesh Khanna, MD, MS: PRESS consortium, and Division of Rheumatology, Department of Medicine, University of Michigan, Ann Arbor; ⁹Monique Hinchcliff, MD, MS: Division of Rheumatology, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, PRESS consortium, and Section of Rheumatology, Allergy & Immunology, School of Medicine, Yale University, New Haven, Connecticut.

Author disclosures are available at <https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fart.42380&file=art42380-sup-0001-Disclosureform.pdf>.

Address correspondence via email to Harris Perlman, PhD, at h-perlman@northwestern.edu; to Monique Hinchcliff, MS, MD, at monique.hinchcliff@yale.edu; or to Deborah R Winter, PhD, at deborah.winter@northwestern.edu.

Submitted for publication February 28, 2022; accepted in revised form October 6, 2022.

allophycocyanin (Biolegend), APC–Cy7–conjugated anti-CD16 (BD Biosciences), phycoerythrin-conjugated anti-CD1c (PE; Miltenyi Biotec), Alexa Fluor 700-conjugated anti-CD20 (BD Biosciences), and PE–Cy7–conjugated anti-CD56 PE–Cy7 (BD Biosciences). Cell sorting was performed on the FACSria III (BD Biosciences) at the NU Robert H. Lurie Cancer Center Flow Cytometry Core Facility.

Skin biopsy digestion and macrophage sorting. Fat was discarded, epidermal and dermal tissues were separated, and tissue was mechanically digested with preset gentleMACS program *m_lung_01*. Cells were labeled with viability dye, treated with Fc block, and stained with the following cocktail: anti-CD45 PE–Cy7, anti-HLA–DR eFluor 450, anti-CD206 APC (BD Biosciences), anti-CD16 PE (BD Biosciences), anti-CD56 Alexa 700 (Biolegend), and anti-CD15 Alexa700 (BD Biosciences).

Preparation of RNA library. RNA from sorted cells was extracted using the PicoPure RNA isolation kit per instructions of the manufacturer (Arcturus). Full-length RNA-seq libraries from CMs and skin were prepared using SMART-Seq v4 Ultra Low Input RNA Kit (Clontech Laboratories) followed by Nextera XT protocol (Illumina). For NCM samples, a QuantSeq FWD (Lexogen GmbH) was used to generate Illumina compatible libraries. Libraries were sequenced on a NextSeq 500 instrument (Illumina) with a minimum of 5×10^6 reads per sample, and commercially available universal human RNA was included for reference.

RNA-seq alignment and mapping. Sequenced reads were de-multiplexed (*bcl2fastq*), trimmed (*Trimomatic0.36* for full-length, *bbduk* for 3'), and aligned to hg19 (*TopHat2* 2.17.1.14 for full-length or *STAR*, version 2.5.2, for 3' data sets [NCMs]). Samples with fewer than 5×10^6 uniquely aligned reads or with complexity (percentage of reads mapping to unique genomic positions) $<30\%$ were excluded due to low quality. As this included 1 of the original 15 SSc CM samples, we also removed the paired NCM sample for this patient. Aligned reads were mapped to genes and counted using *HTseq* with *Homo_sapiens.GRCh37.75* as reference. Gene coverage (5' and 3') for full-length libraries was calculated using the *RSeQC* package. Gene counts were normalized by calculating fragments per kilobase per million (FPKM) for CM and skin macrophage samples or counts per million (CPM) for NCM samples. Expressed genes in each data set were defined as those with ≥ 3 samples above the minimum threshold (FPKM = 5 for CMs, CPM = 31 for NCMs, FPKM = 3 for skin).

Identification of CM transcriptional signature in SSc. R Studio (<https://www.rstudiocom/>) was used to calculate Pearson's correlation coefficients and to generate principal components analysis (PCA) plots, Venn diagrams (*ggplots* package),

and scatterplots (*ggplot2* package). GraphPad Prism was used to generate bar graphs of individual gene expression. Differential expression was calculated using the *DEseq* package (version 1.18.1) on 5,171 expressed genes; genes with changes in expression of ≥ 2 -fold and with Benjamini-Hochberg-adjusted *P* values less than 0.05 were considered differentially expressed genes (DEGs) between the SSc cohort and controls (25). Gene Ontology (GO) analysis was performed on up and down DEGs independently using *GOrilla* (cbl-gorilla.cs.technion.ac.il) with expressed genes (5171) as background. To define differential genes in individual SSc patients compared to healthy controls, we calculated a modified Z score (z') for each gene based on its expression distribution in the control cohort:

$$z'_{ij} = \frac{(X_{ij} - \mu_{Ci})}{\sigma_{Ci}}$$

Where x is the expression level of gene i in patient j , μ_{Ci} is the mean expression for gene i across controls, and σ_{Ci} is the SD of expression for gene i across controls. To prevent artificially high Z scores resulting from small SDs, where a high Z score may not represent true biologic variability, we set the minimum σ_{Ci} to 2.5. We identified 1,790 genes with $|z'| > 2$ in ≥ 3 patients. *GENE-E* (<https://software.broadinstitute.org/GENE-E/>) was used to perform hierarchical clustering of samples and K-means clustering of genes ($k = 4$), resulting in patient groups A–C and gene clusters I–IV. *Pvclust* was used to assess the confidence in the dendrogram resulting from hierarchical clustering. We used resampling method (without replacement) of the genes to evaluate the robustness of the three groups by determining how often the patients ended up in the same clade out of 1,000 iterations with varying subset sizes. *DEseq* was run, and DEGs were defined as above in each patient group compared to controls.

Gene set enrichment analysis (GSEA). To determine the biologic processes of enriched genes in each group, 5,171 expressed genes were ranked by group based on their fold change relative to the mean expression across controls and given as preranked input for *GSEA* version 3.0 (www.gsea-msigdb.org). Preranked genes were compared to the biologic processes of the Molecular Signature Database (MSigDB) C5 version 7.0 collection, using preranking and classic weighting. Normalized enrichment scores were reported, and significant enrichment (or depletion) was defined as having an adjusted *P* value less than 0.05.

Comparison to NCM transcriptional signature. Starting with the 1,790 genes in CM clusters, we identified 1,599 among the NCM expressed genes and calculated the modified Z score relative to NCM controls. Hierarchical clustering was performed on these genes based on their NCM Z score, and dendrogram confidence and group robustness were assessed as

described above. DEseq was performed as above on 7,143 expressed genes to define DEGs between each patient group and controls, and ggplot2 was used to compare DEGs in CMs versus NCMs. GSEA was run as described above based on fold-change in NCMs.

Expression of published monocyte genes. We defined 431 CM and 388 NCM genes based on a publicly available microarray data set (26). To determine whether these gene sets were differentially expressed in our patient groups, we calculated the average expression in each individual and compared the distribution in patient groups to the control cohort. We calculated the average expression of each gene by patient groups and determined the percentage that exhibited greater expression than in controls. Significance was calculated by Fisher's exact test ($P < 0.05$). Prism was used to graph these data.

Skin macrophages. We used a nonparametric Mann-Whitney U test to define SSc-specific genes ($n = 190$) and control-specific genes ($n = 138$) out of 6,338 expressed genes in skin macrophages from patients compared to controls with a P value less than 0.05. GENE-E was used to visualize the relative expression of these genes as a heatmap with minimum-to-maximum scaling. The number of these genes that overlap with those defined in each of the CM clusters were calculated, and the significance of enrichment was determined using Fisher's exact test ($P < 0.05$). Receiver operating characteristic (ROC) curve (pROC package in R) was used to determine the performance of clusters I–IV at delineating status in patients versus controls. Each sample was scored as the mean of the minimum-to-maximum value for all genes in the given cluster; P values were calculated by the package using Mann-Whitney U test.

Statistical analysis of clinical data. Baseline forced vital capacity percent predicted (FVC%) and MRSS differences among the groups identified in the cluster analysis were assessed using the Kruskal-Wallis test. Baseline distributions of FVC and MRSS by groups identified in the cluster analysis were illustrated using box plots using Prism. Additionally, the longitudinal FVC and MRSS data are displayed as spaghetti plots using Prism.

Comparison to published SSc monocyte data. All information on patient diagnosis, grouping, and sample acquisition was obtained from the publicly available data set by van der Kroef et al (27). Starting with 1,790 genes across the CM clusters, we identified 1,523 that were expressed in bulk monocyte RNA-seq from this data set. For each gene, we normalized using minimum-to-maximum scaling and calculated the mean score across genes in each cluster for each individual. ROC curves were

plotted as described above to determine the performance of clusters I–IV at delineating each SSc patient category versus controls.

RESULTS

Greater variability in CMs from SSc patients versus those from healthy controls. We obtained blood samples from 14 dcSSc patients and 15 age-, sex-, and race-matched healthy volunteers as controls (Figure 1A). We then isolated CMs and NCMs using flow cytometry and found no significant differences in the proportion detected among CD45+ immune cells between the SSc and control cohorts (Figures 1B and C). We performed RNA-seq to profile the gene expression of CM, the predominant monocyte population. One sample that did not pass quality control tests was eliminated, and we confirmed patient sex as an additional quality control (Supplementary Figures 1A–D, available on the *Arthritis & Rheumatology* site at <https://onlinelibrary.wiley.com/doi/10.1002/art.42380>). From the remaining 29 samples, we found that the expression of genes tended to be more variable across SSc patients than across controls (Figure 1D and Supplementary Figure 1E). By calculating an average transcriptional profile for the control cohort, we determined that SSc patients exhibited a range in their similarity to healthy controls (Supplementary Figure 1F). Moreover, while the CM gene expression of patients that most resembled controls was necessarily similar to each other, the remaining patients exhibited substantial variability in their transcriptional profiles (Supplementary Figure 1G). This variability did not seem to be associated with the site of sample collection (Supplementary Figure 1H), although patients receiving immunosuppressive therapies tended to more closely resemble controls (Figure 1E).

We next performed pairwise differential expression analysis between the SSc and control cohorts and defined 146 and 21 genes as up- and down-regulated in SSc, respectively (Figure 1F and Supplementary Tables 1A and B, <https://onlinelibrary.wiley.com/doi/10.1002/art.42380>). The 146 genes include a type I interferon (IFN) signaling pathway (*OAS1*, *OAS2*, *IRF2*), response to cAMP (*FOS*, *FOSB*, *JUN*), regulation of vascular development (*VEGFA*, *KLF2*, *SERPINE1*), and regulation of leukocyte migration (*CSF1R*, *CCR2*, *CCL2*).

Clustering SSc CMs into 3 transcriptional subgroups.

In order to capture the transcriptional dysregulation in individual SSc patients, we devised an alternative approach to define genes associated with disease. For each patient, we considered a given gene to be differential if its expression was >2 SDs from the control mean. By hierarchically clustering the SSc patients based on the 1,790 genes that were differential in ≥ 3 patients, we identified 3 patient subgroups (groups A–C). We also noted that 1 patient (no. SSc15) was highly similar to controls and not included in any group (Figure 2A Supplementary Table 2A, <https://onlinelibrary.wiley.com/doi/10.1002/art.42380>). The clustering of

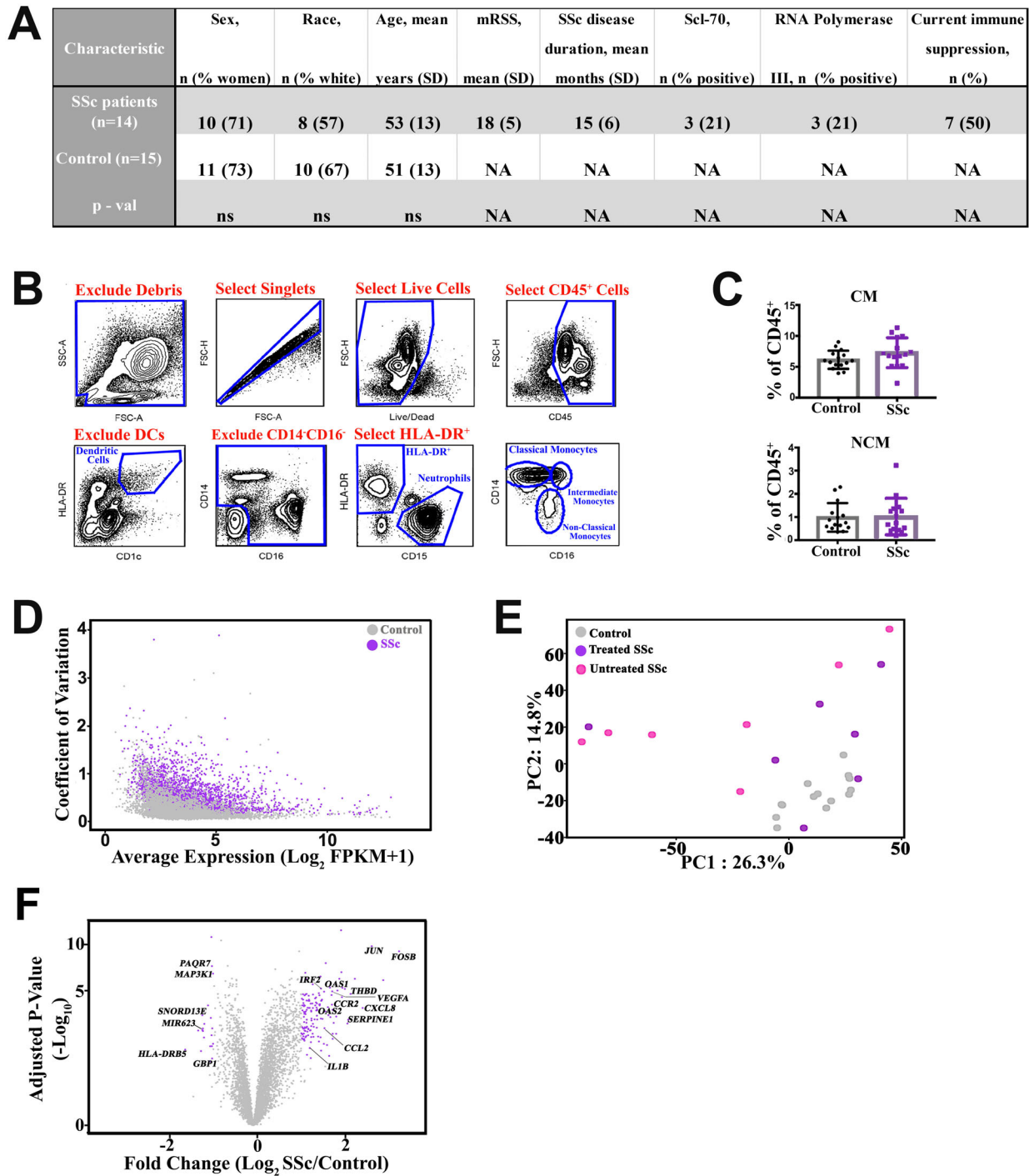


Figure 1. Classical monocytes (CMs) from patients with early diffuse systemic sclerosis (SSc) displayed transcriptional heterogeneity. **A**, Characteristics of the patients in the Prospective Registry of Early Systemic Sclerosis and the controls. *P* value for age (by Mann-Whitney U test) was not significant (ns), and *P* values for sex and race (by Fisher’s exact test) were not significant. **B**, Contour plots depicting stepwise isolation of CMs and nonclassical monocytes (NCMs) from blood by fluorescence-activated cell sorting. **C**, Percentage of CMs and NCMs in CD45+ cells from SSc compared to control samples (*P* = 0.14 and *P* = 0.89, respectively, by Mann-Whitney U test). Symbols represent individual samples; bars show the mean ± SD. **D**, Scatterplot showing the coefficient of variation (SD/average) for expression (fragments per kilobase per million (FPKM) relative to the average expression (log₂[FPKM + 1]) in SSc and control CMs. **E**, Principal Components Analysis (PCA) of gene expression in CM samples from treated SSc patients, untreated SSc patients, and controls. **F**, Volcano plot showing differentially expressed genes in the SSc cohort (n = 167) compared to controls, based on log₂ fold change ≥ 1 or ≤ -1 (significance calculated by DEseq with Benjamini-Hochberg false discovery rate correction). Purple indicates genes with *P*_{adj} < 0.05 and log₂ fold change ≥ 1 or ≤ -1. Data shown in **C–F** were based on 5,171 expressed genes in CMs. MRSS = modified Rodnan skin thickness score; DCs = dendritic cells; NA = not applicable.

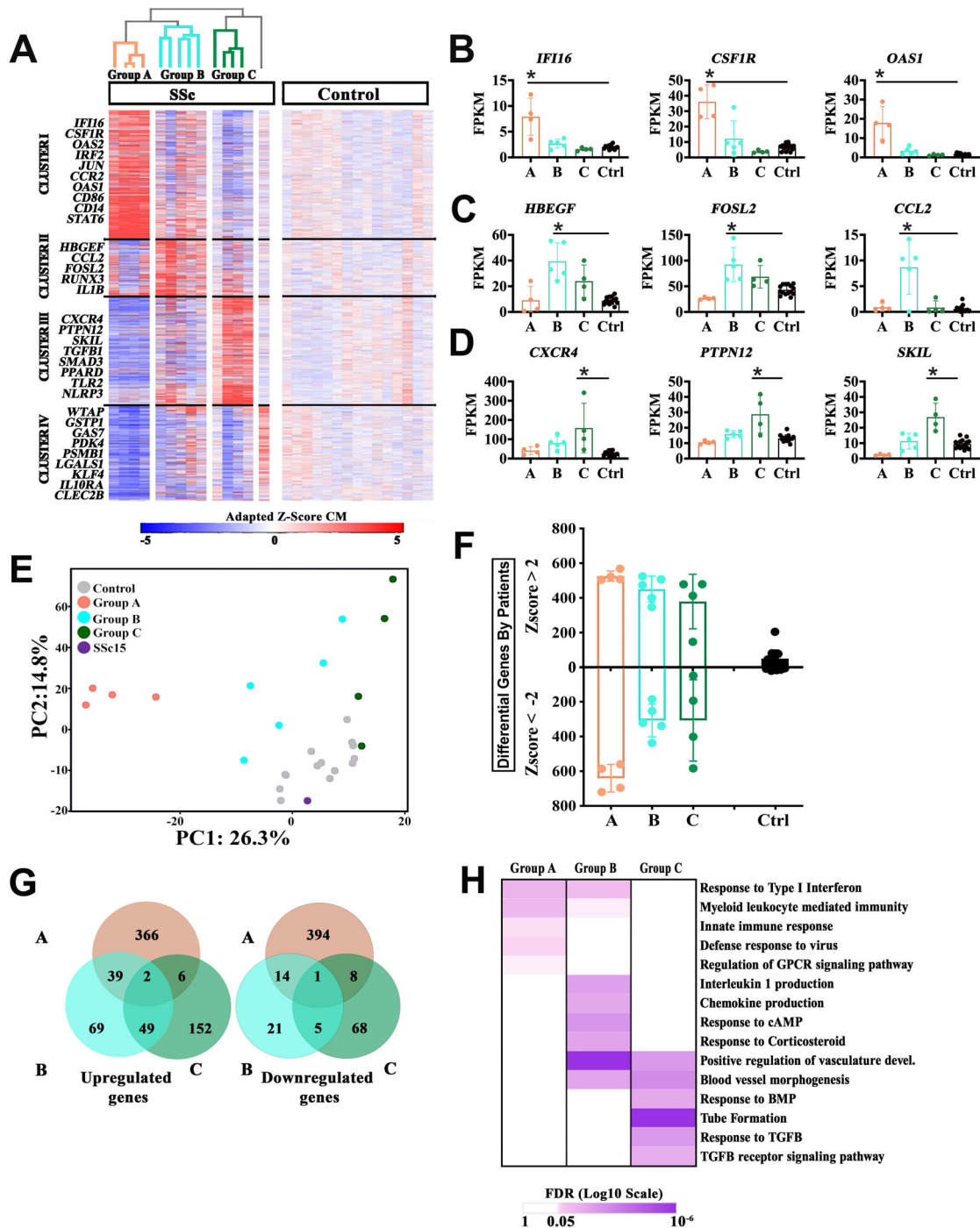


Figure 2. Patients with early diffuse SSc stratified into 3 groups based on the transcriptional profile of CMs. **A**, Heatmap of adapted Z scores of CM gene expression relative to controls, depicting dendrogram resulting from unsupervised hierarchical clustering of samples into 3 groups and K-means clustering of 1,790 differentially expressed genes (DEGs) by individual [adapted Z score] >2 in ≥ 3 patients) (rows) ($k = 4$). **B–D**, Expression of representative genes from cluster I ($n = 592$) (**B**), cluster II ($n = 261$) (**C**), and cluster III ($n = 500$) (**D**), in each patient group as well as controls (Ctrl). Symbols represent individual samples; bars show the mean \pm SD. * = $P_{\text{adj}} < 0.05$, by DEseq with Benjamini-Hochberg false discovery rate (FDR) correction. **E**, PCA of gene expression in CM samples color-coded based on patient groups. **F**, Number of genes with adapted Z scores >2 or <-2 in individual CM samples organized by patient groups. **G**, Venn diagram showing overlap of up-regulated (\log_2 fold change >1 and $P_{\text{adj}} < 0.05$) or down-regulated (\log_2 fold change <-1 and $P_{\text{adj}} < 0.05$) genes in each group (groups A–C) compared to controls. **H**, Heatmap showing the FDR based on gene set enrichment analysis of biologic processes associated with DEGs in the 3 groups. Data shown in **A–H** were based on 5,171 expressed genes in CMs. GPCR = G protein-coupled receptor; BMP = bone morphogenetic protein; TGFB = transforming growth factor β (see Figure 1 for other definitions).

patients was robust: while individual branches of the dendrogram changed, the identity of the groups remained largely constant regardless of the subset of genes used for clustering analyses (Supplementary Figures 2A and B). Moreover, using K-means clustering of the 1,790 genes, we characterized 4 gene expression patterns associated with these subgroups. Cluster I is primarily expressed in group A, cluster II in group B, cluster III in group C, and cluster IV is down-regulated in the majority of patients compared to controls (Figures 2A–D). The distinct transcriptional signatures of each group were reinforced when visualized using PCA (Figure 2E). Patients in group A tended to have the most DEGs, and both groups A and B demonstrated overlap with the genes identified in the prior full-cohort analysis (Figure 2F). On the other hand, group C patients were characterized by the expression of genes in CMs that have hitherto been undetected and may represent dysregulated pathways that have been missed in earlier studies as well.

We performed pairwise differential analysis between each patient group relative to controls for all expressed genes in the data set to better characterize their transcriptional profiles (Figures 2G and H, Supplementary Figures 2C, E, and G, and Supplementary Tables 2B and C, <https://onlinelibrary.wiley.com/doi/10.1002/art.42380>). We found that up-regulated genes in group A were associated with type I IFN, such as *IFI16*, *OAS1*, *OAS2*, and *IRF2*, as well as innate immunity in general and G protein-coupled receptor (GPCR) signaling, including *CSFR1*, *CCR2*, and *CD86*. Group B had the fewest DEGs and the most overlap with other groups. This may explain why group B was the least conserved in resampling (Supplementary Figure 2B). Like group A, group B genes were associated with type I IFN pathways and innate immunity but exhibited higher enrichment of chemokine production and response to cAMP and glucocorticoids. Both groups B and C showed up-regulated genes associated with vasculature development and blood vessel morphogenesis, such as *VEGFA* and *HBEGF* (28). Group C exhibited additional genes associated with transforming growth factor β (TGF β), tube formation, and bone morphogenetic protein (BMP) signaling, including *TGFB1*, *SMAD3*, *SKIL*, *PPARD*, *CXCR4*, and *PTPN12*.

Patient groups were largely conserved in NCM analyses. Next, we analyzed RNA-seq data from NCMs isolated from the same patients (Supplementary Figure 3A, <https://onlinelibrary.wiley.com/doi/10.1002/art.42380>). The variation in gene expression was higher in NCMs than CMs in both patients and controls (Supplementary Figures 3B–D), but this may be due to technical differences in the protocols or higher sampling bias from the smaller NCM population. NCMs were clustered based on the 1,790 CM genes that defined groups A, B, and C to address the conservation of groups on a global scale without introducing new genes (Figure 3A and Supplementary Table 3A). Patients in groups A and C continued to cluster next to each

other, but group B no longer formed its own clade (Figure 3A and Supplementary Figures 3E–G). As with CMs, we performed pairwise analyses for DEGs for each of the 3 groups (Supplementary Figure 3H and Supplementary Tables 3B and C). To complement our approach in Figure 3A, we then compared the list of DEGs between the previously defined groups to determine whether individual genes were conserved and identify genes that may be uniquely differential in NCMs (Figures 3B–F and Supplementary Figure 3I). We found that genes related to IFN pathways and innate immunity were preferentially expressed in group A in both CMs and NCMs. Although group B did not exhibit as many significantly up-regulated genes in NCMs as in CMs, group B NCM genes were associated with similar processes as CMs. Group C patients exhibited far fewer NCM DEGs, but those associated with its unique transcriptional signature, such as *CXCR4*, *PTPN12*, and *SKIL*, were significantly up-regulated.

Comparison of monocyte identity across patient subgroups. It has been proposed using mouse models that a subset of CMs that have egressed from the bone marrow differentiate into NCMs (29). The genes that distinguish NCMs from CMs on a transcriptional level have been well established and are conserved between mice and humans (30). To determine whether monocyte identity is altered in SSc, we generated a list of CM-specific and NCM-specific genes based on recent studies (26) (Supplementary Table 4). We found that both CMs and NCMs from group A patients exhibited higher expression of their respective population genes than controls, including *PLAC8*, *PADI4*, *SELL*, *LYZ*, and *CCR2* in CMs and *ITGAL*, *SPN*, *CX3CR1*, and *CSF1R* in NCMs (Figures 4A–D). This was a cell type-specific effect, as we did not observe increased expression of these genes in the opposite population (Supplementary Figures 4A–D, <https://onlinelibrary.wiley.com/doi/10.1002/art.42380>). Group B also exhibited increased CM gene expression only in CMs, while group C levels were slightly lower. These observations were driven by more than a few genes, as a significant proportion were expressed above the average level of the control cohort (Supplementary Figures 4E and F). In comparing CMs and NCMs, we may have expected that the genome-wide transcriptional profile would be most similar between monocytes from the same patient; however, this was not the case (Figure 4E). The overexpression of CM genes in groups A and B may explain why CM gene expression among this SSc cohort is less similar compared to NCMs (Supplementary Figures 4E and F).

Up-regulation of SSc-associated monocyte genes by skin macrophages. Since we observed common transcriptional signatures across monocyte populations in SSc and control cohorts, we investigated whether similar gene expression profiles were observed in skin macrophages. We isolated CD206+HLA-DR+ macrophages from fibrotic skin from 7 SSc patient biopsy samples (Figure 5A and Supplementary Figures 5A–C, <https://>

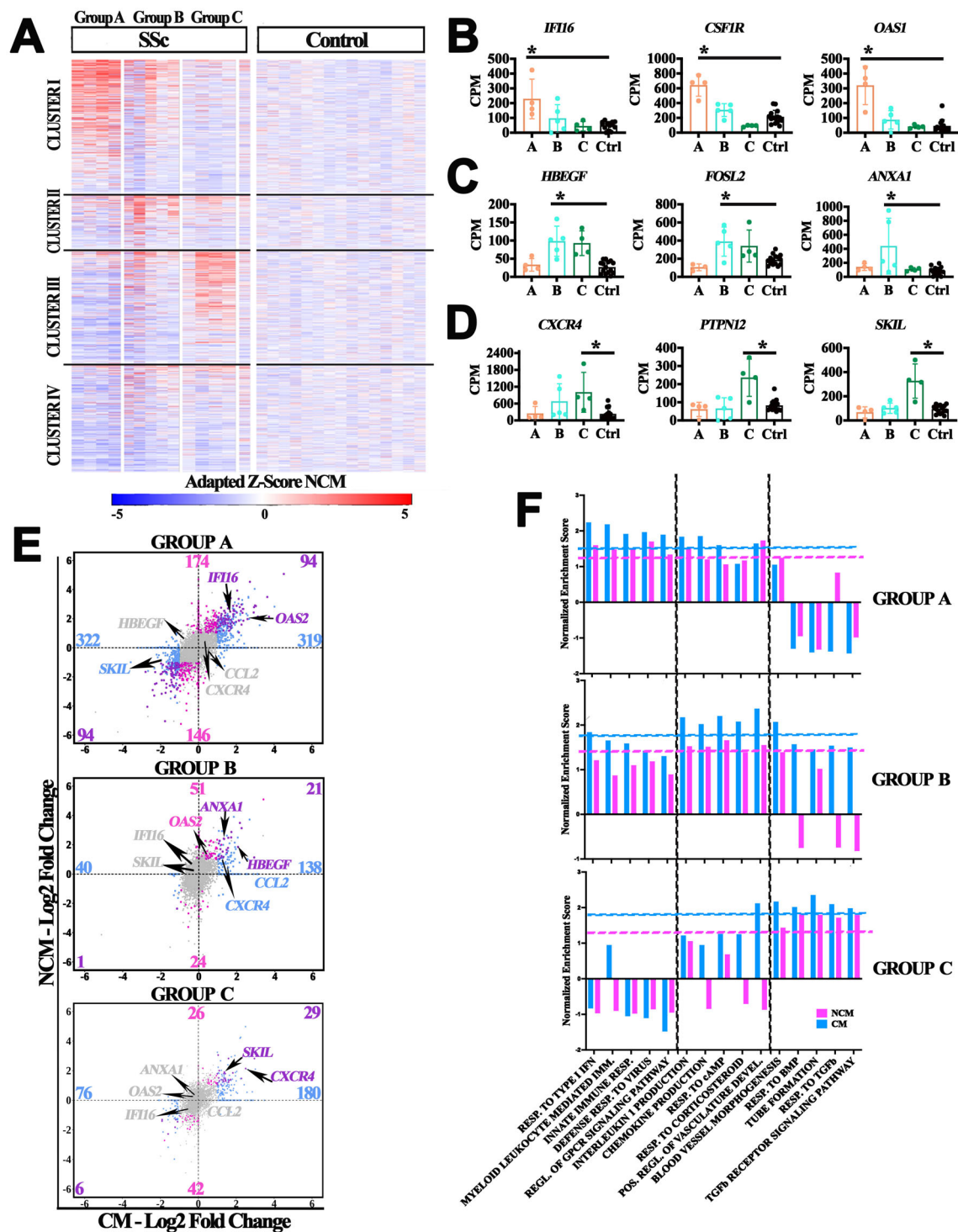


Figure 3. Patient groupings based on CMs exhibited distinct transcriptional profiles in NCMs. **A**, Heatmap of adapted Z scores of NCM gene expression relative to controls using 1,599 genes from CM clusters in Figure 2A. **B–D**, Expression of representative genes from cluster I (**B**), cluster II (**C**), and cluster III (**D**), in each patient group as well as controls (Ctrl). Symbols represent individual samples; bars show the mean \pm SD. $*$ = $P_{adj} < 0.05$, by DESeq with Benjamini-Hochberg false discovery rate correction. **E**, Scatterplots of \log_2 fold change of gene expression in each patient group compared to controls for CMs versus NCMs. Significant genes, as described in Figure 2, were color-coded if they were shared between CMs and NCMs (purple), only in NCMs (pink), or only in CMs (blue). **F**, Bar graph of the normalized enrichment scores (NES) based on gene set enrichment analysis of biologic processes from Figure 2H in CMs and NCMs. Dashed lines indicate the 90% threshold for NES in CMs and NCMs. Vertical lines separate GO terms that relate to each group. Data shown in **A–F** were based on 7,143 expressed genes in NCMs. CPM = counts per million; IFN = interferon; GPCR = G protein-coupled receptor; BMP = bone morphogenetic protein; TGFB = transforming growth factor β (see Figure 1 for other definitions).

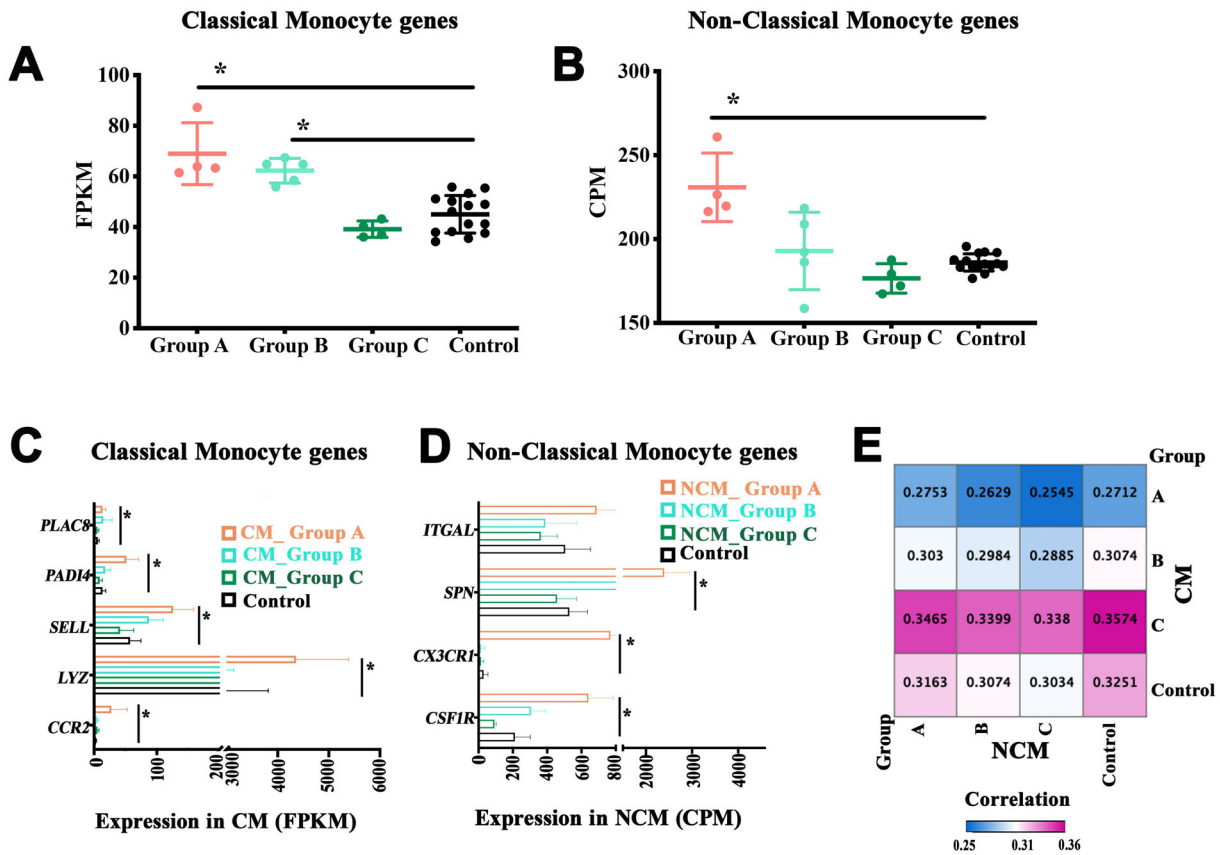


Figure 4. Variation in expression of monocyte gene signatures in SSc CMs and NCMs. **A**, Average expression level of all CM genes in CM samples from patient groups and controls. **B**, Average expression level of all NCM genes in NCM samples from patient groups and controls. **C**, Expression levels of representative CM genes in patient CM samples by group. **D**, Expression levels of representative NCM genes in patient NCM samples by group. Symbols represent individual samples; bars show the mean ± SD. * = $P < 0.05$ by Mann-Whitney nonparametric test (**A** and **B**) and $P_{adj} < 0.05$ by DEseq with Benjamini-Hochberg false discovery rate correction (**C** and **D**). **E**, Pearson correlation of average gene expression between CM and NCM samples in patient groups and controls. CPM = counts per million (see Figure 1 for other definitions).

onlinelibrary.wiley.com/doi/10.1002/art.42380) and from 7 age-, sex-, and race-matched healthy controls for RNA-seq analysis (Supplementary Figures 5D and E). We identified 328 DEGs between skin macrophages in SSc patients versus controls (Figure 5B, Supplementary Figure 5F, and Supplementary Table 5). The 190 SSc-specific genes included those that have been previously implicated in SSc and/or macrophages, such as *ITGB2* (31), *ITGA5* (32), *ADAM8* (33), *ALOX15* (34), *CLEC10A* (35,36), *NFKB2* (37), *NFKBIE* (38,39), and *RUNX3* (40). The 138 control-specific genes included transcription factors, such as *CEBPD*, which potentiates macrophage inflammatory response and angiogenesis (41,42), *ID3*, which is necessary for macrophage specification in liver (43), *EPAS1* (HIF2a), which suppresses the *NLRP3* inflammasome (44), and *LRG1* (45) and *IL6*, which play a role in angiogenesis (Figure 5C). Of note, *ITGB2*, *ITGA5*, *ADAM8*, *CEBPD*, and *NFKB2* were all differentially expressed in CM SSc patients compared to controls.

We found that cluster I–III genes, which corresponded to up-regulated genes in groups A, B, and C in our CM analysis, were more likely to be SSc-specific than control-specific

(enrichment: cluster I = 1.54, cluster II = 1.82, cluster III = 3.27), but only cluster III (group C) was significantly enriched ($P = 0.0412$) (Figure 5D and Supplementary Figure 5G, <https://onlinelibrary.wiley.com/doi/10.1002/art.42380>). Similarly, classification strategies based on the average expression of cluster I and III genes in skin macrophages were able to differentiate between SSc patients and controls significantly better than at random ($P = 0.03$ and $P = 0.01$, respectively) (Figure 5E). On the other hand, cluster IV genes, which were largely down-regulated in monocytes, were generally expressed more highly in controls than in SSc patients (enrichment: cluster IV = 0.40; $P = 0.029$).

Differing disease characteristics across patient subgroups. Our analysis stratified SSc patients based on CM transcriptional profiles, while remaining agnostic to clinical presentation such as cutaneous and pulmonary involvement. We collected baseline and longitudinal MRSS and FVC data up to 42 months after blood sample collection, we found that patients in groups B and C exhibited significant FVC

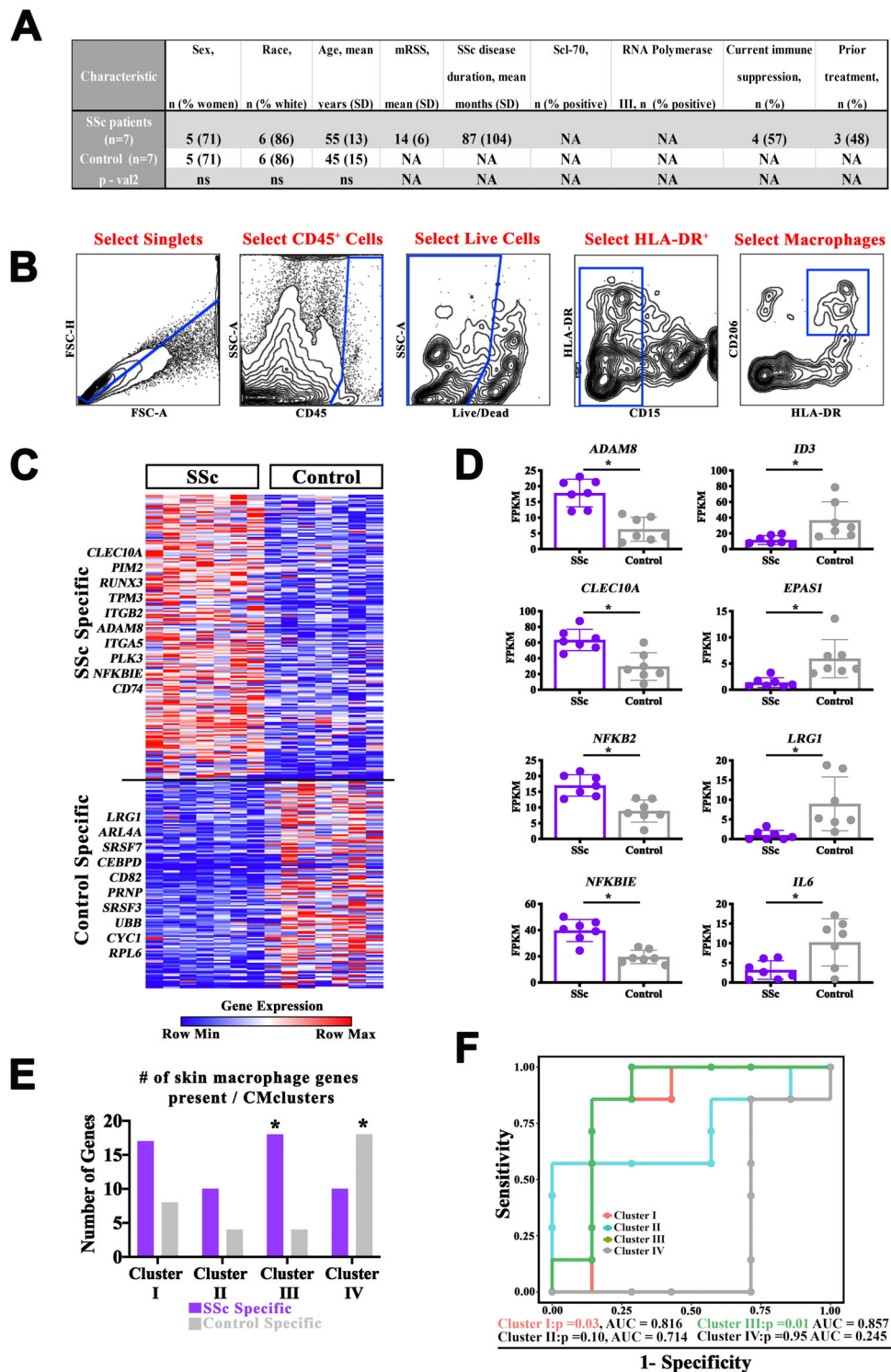


Figure 5. SSc-associated monocyte genes were up-regulated in SSc skin macrophages compared to controls. **A**, Characteristics of the patients in the Prospective Registry of Early Systemic Sclerosis and the controls. P values for age (by Mann-Whitney U test) and for sex and race (by Fisher's exact test) were not significant. **B**, Gating scheme depicting the purification of macrophages from skin. **C**, Heatmap of 190 SSc-specific and 138 control-specific genes based on significantly differing expression in skin macrophages ($P < 0.05$ by Mann-Whitney U test). **D**, Expression of representative genes from heatmap in skin macrophages from SSc patients and controls. Symbols represent individual samples; bars show the mean \pm SD. * = $P_{\text{adj}} < 0.05$. **E**, Number of genes from **B** that overlap CM clusters I-IV in Figure 2A. * = $P < 0.05$ by Fisher's exact test. **F**, Receiver operating characteristic curve based on the sensitivity and specificity of distinguishing skin macrophages from SSc patients versus controls using the average normalized expression of genes from CM clusters I-IV. Significant areas under the curve (AUCs) are colored. Data shown in **B-E** were based on 6,338 expressed genes in skin macrophages. See Figure 1 for other definitions.

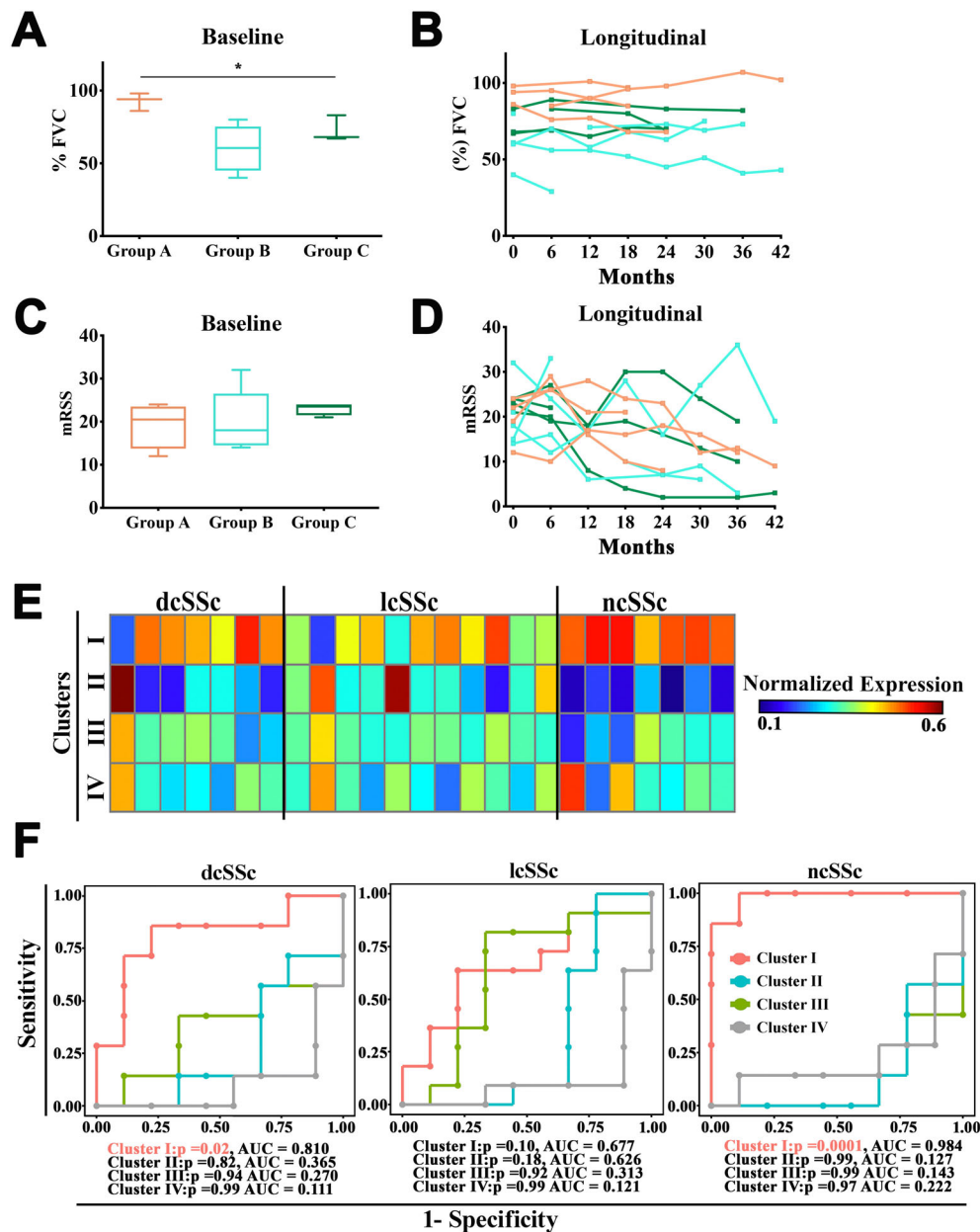


Figure 6. SSc patient groups differed across clinical phenotypes. **A**, Forced vital capacity (FVC) by patient group at baseline (time of blood sample). * = $P < 0.05$ by Kruskal-Wallis test for 3 groups. **B**, FVC of each patient colored by group longitudinally over 42 months from baseline (time = 0). **C**, MRSS patient group at baseline (time of blood sample). In **A** and **C**, each box extends from the 25th to the 75th percentile. Lines inside the boxes represent the median. Whiskers represent the minimum to maximum values. **D**, MRSS of each patient colored by group longitudinally over 42 months from baseline. **E**, Average normalized expression of genes from CM clusters I-IV (Figure 2A) in bulk monocytes from SSc patients categorized by disease phenotype (data set described in ref. 27). **F**, Receiver operating characteristic curve based on the sensitivity and specificity of distinguishing bulk monocytes from SSc patient categories versus controls, using the average normalized expression of genes from CM clusters I-IV. Significant areas under the curve (AUCs) are colored. dcSSc = diffuse cutaneous SSc; lcSSc = limited cutaneous SSc; ncSSc = noncutaneous SSc; (see Figure 1 for other definitions).

reduction, indicating poor lung function at baseline, compared to group A patients who had better lung function that tended to remain stable over time (Figures 6A and C). In contrast, there were no significant differences in MRSS levels between groups at baseline or longitudinally (Figures 6B and D).

Application of transcriptional groups to independent data set. In order to assess the potential utility of our CM classification for future clinical studies, we developed an unbiased algorithm that is tolerant to data differences resulting from technical artifacts, protocol choice, sequencing depth, and normalization. To test the algorithm, we used gene expression data from

an unrelated study of monocytes in having SSc patients performed by another group (27). In this study, they isolated bulk monocytes for RNA-seq and categorized patients as having dcSSc, lcSSc, or ncSSc/sine scleroderma. For each patient, we calculated a score to reflect the relative level of expression of each cluster from our CM analysis (Figure 6E, Supplementary Figure 6, and Supplementary Table 6). We found that there was variability in the transcriptional signature of each patient that did not align with their disease subtype. We then built classifiers for each of the 4 gene clusters and assessed their ability to correctly assign a patient to the correct SSc subtype (Figure 6F). As expected, cluster IV, which was down-regulated in the majority of our patients, worked better as a negative classifier in each subtype. Cluster I (associated with group A) accurately distinguished dcSSc from controls ($P = 0.02$) but performed particularly well in ncSSc patients. Clusters I and II both exhibited high but not significant AUC values in lcSSc patients, suggesting that this category is split between these transcriptional identities. Although individual patients exhibited high expression of cluster III genes, it did not perform well as a classifier in any subtype. These findings suggest that the transcriptional signatures we defined can be identified in independent data sets, but more investigation is needed to determine the full clinical significance of our patient groups.

DISCUSSION

Multiple studies have quantified gene expression in whole peripheral blood or skin from SSc patients (46). Like all bulk microarray or RNA-seq experiments, these studies are subject to changes in cellular composition that can drive gene expression signatures and a loss of the ability to detect biologically important transcriptional changes within minority cell populations. As such, they rely on complex analyses, such as deconvolution and network analysis, and a priori knowledge from extant gene expression databases (e.g., GEO). Here, we are the first to show that unbiased transcriptomic analysis of CMs and NCMs can stratify SSc patients and associate with disease activity outcome measures independent of treatment. One of the strengths of focusing on circulating monocytes is that they are composed of only 2–3 populations, which allows for bulk separation.

Patients were segregated into 3 groups based on CM transcriptional profiles, with each group demonstrating up-regulation of functionally distinct gene sets, including response to type I IFN, myeloid leukocyte-mediated immunity, and regulation of GPCR signaling pathway for group A; interleukin-1 (IL-1) and chemokine production, response to cAMP, and response to glucocorticoids for group B; and tube/vessel formation, response to BMP, and TGF β receptor signaling pathway for group C. They were compared to each other and to other participants. Since these 3 groups are largely recapitulated in the analysis of NCM transcription profiling, these data suggest some conservation in the gene signatures from groups A, B, and C.

Analysis of bulk skin CD206+HLA-DR+ macrophages revealed that genes associated with groups A, B, and C tended to increase in expression in skin macrophages from SSc patients compared to controls. Many of these genes were previously associated with disease processes in monocytes and macrophages. These results suggest a connection between circulating precursors and mature tissue-resident cells, and ongoing studies are underway to further investigate these results in paired skin and blood, which will enable direct comparison of immature and mature myeloid gene expression. The SSc patients in group A displayed the highest baseline lung function compared to either group B or C, but exhibited no difference in MRSS. Taken together, our results suggest that transcriptional profiling of distinct CM and NCM subpopulations may represent a viable mechanism for stratifying patients and potentially their response to therapeutics.

Over the past decade, numerous studies have utilized transcriptional profiling to understand SSc disease pathology or generate new patient stratification methods. DNA microarray of skin biopsies from patients with SSc has demonstrated four SSc “intrinsic” subsets that may be detectable in PBMCs and esophageal biopsies (14,47,48); however, it had not been established that this type of differentiation is detectable outside of diseased end organs, although one study suggested a link between the immune component and fibrosis across multiple end organs (48). Further, an inflammatory gene signature associated with macrophages in bulk skin using RNA-seq provides additional support for the crucial roles that macrophages play in the development of SSc. Moreover, a change in skin severity score that includes a 415-gene signature has been shown to correlate with MRSS (49), although to date no studies have corroborated the skin severity score. Recently, a few groups examined gene expression specifically in individual immune populations such as monocytes or macrophages using bulk or single-cell RNA-seq (27,50,51). While these studies support the role for the type I IFN as well as the IL-1 β pathway in bulk SSc monocytes (CMs and NCMs), neither study examined any clinical associations (27,50).

Currently, there are only 2 single-cell RNA-seq studies of human SSc skin which examined the proportion of populations and their respective gene signature but did not examine association with disease outcome measures (51,52). The importance of our CM and NCM data is that they clearly illustrate the presence of disease subsets—or endotypes—in circulating immune cells. Further, our data differentiate subtypes of SSc patients in another study, which were only stratified based on their clinical cutaneous classification, i.e., dcSSc, lcSSc, and ncSSc (27). These data suggest that comparison of variability across populations which can be discerned via CM or NCM RNA-seq analysis may influence the traditional analyses and classification of SSc patients. Although our study clearly delineated 3 patient groups without apparent distinguishing clinical phenotypes, further studies will examine the reproducibility of this grouping and expand the

sampling to definitively test any association with clinical parameters. Additional limitations to our study include the sample size, an inability to acquire match blood and skin over time, the potential that circulating factors may affect monocyte subpopulation numbers as well as gene expression at various stages of disease, the focus on tissue resident macrophages, and the heterogeneous SSc population in the validation cohort. Future studies that include a larger number of patients, match blood and skin, and include longitudinal analysis are needed to determine whether these groups are static with and without treatment.

Clinical heterogeneity in SSc presentation has been well characterized for decades, and differential gene expression has been demonstrated in fibrotic tissue; however, reproducible connections have only begun to be established between transcriptional intrinsic subsets and clinical characteristics (53), including prognosis and response to treatment (17,18,54,55). A thorough molecular understanding of SSc may help to stratify patients for enrollment in clinical trials and to inform drug selection. Further investigation of CMs and NCMs could determine how SSc subsets respond to targeted interventions and whether they can predict which end organ would respond to a particular therapeutic.

ACKNOWLEDGMENTS

We thank the Northwestern University Lurie Cancer Center Flow Cytometry Core Facility, which is supported by NCI Cancer Center Support Grant P30 CA060553 awarded to the Robert H. Lurie Comprehensive Cancer Center. We also thank the Next Generation Sequencing Core, the Division of Pulmonary and Critical Care and Rheumatology Sequencing Core, NUCATS Biostatistics and Collaboration Center at Northwestern University, Northwestern University Core Center for Clinical Research in the Division of Rheumatology (project no. 5P30AR072579), and Northwestern University Skin Biology and Diseases Resource-based Center (project no. 5P30AR075049). This research was supported in part through the computational resources and staff contributions provided by the Genomics Compute Cluster on QUEST, which is jointly supported by the Feinberg School of Medicine, the Office of the Provost, the Office for Research, and Northwestern Information Technology. In addition, this work was made possible by a grant from the Northwestern Digestive Health Foundation and gifts from Joe and Nives Rizza and the Todd and Renee Schilling Charitable Fund.

AUTHOR CONTRIBUTIONS

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. Dr. Perlman had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Makinde, Carns, Aren, Cuda, Pandolfino, Dematte D'Amico, Budinger, Assassi, Frech, Khanna, Perlman, Hinchcliff, Winter.

Acquisition of data. Makinde, Dunn, Gadhvi, Carns, Aren, Cuda, Dominguez, Pandolfino, Dematte D'Amico, Budinger, Assassi, Frech, Khanna, Shaeffer, Perlman, Hinchcliff, Winter.

Analysis and interpretation of data. Makinde, Dunn, Gadhvi, Carns, Aren, Chung, Muhammad, Song, Cuda, Dominguez, Pandolfino, Dematte D'Amico, Budinger, Assassi, Frech, Khanna, Shaeffer, Perlman, Hinchcliff, Winter.

REFERENCES

1. Assassi S, Radstake TR, Mayes MD, et al. Genetics of scleroderma: implications for personalized medicine? *BMC Med* 2013;11:9.
2. Hu Y, Hu Y, Xiao Y, et al. Genetic landscape and autoimmunity of monocytes in developing Vogt-Koyanagi-Harada disease. *Proc Natl Acad Sci U S A* 2020;117:25712–21.
3. Roberts ME, Barvalia M, Silva J, et al. Deep phenotyping by mass cytometry and single-cell RNA-sequencing reveals LYN-regulated signaling profiles underlying monocyte subset heterogeneity and lifespan. *Circ Res* 2020;126:e61–79.
4. Van der Kroef M, van den Hoogen LL, Mertens JS, et al. Cytometry by time of flight identifies distinct signatures in patients with systemic sclerosis, systemic lupus erythematosus and Sjogrens syndrome. *Eur J Immunol* 2020;50:119–29.
5. Toledo DM, Pioli PA. Macrophages in systemic sclerosis: novel insights and therapeutic implications. *Curr Rheumatol Rep* 2019;21:31.
6. Misharin AV, Morales-Nebreda L, Reyfman PA, et al. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J Exp Med* 2017;214:2387–404.
7. Reyfman PA, Walter JM, Joshi N, et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am J Respir Crit Care Med* 2019;199:1517–36.
8. Higashi-Kuwata N, Jinnin M, Makino T, et al. Characterization of monocyte/macrophage subsets in the skin and peripheral blood derived from patients with systemic sclerosis. *Arthritis Res Ther* 2010;12:R128.
9. Arai M, Ikawa Y, Chujo S, et al. Chemokine receptors CCR2 and CX3CR1 regulate skin fibrosis in the mouse model of cytokine-induced systemic sclerosis. *J Dermatol Sci* 2013;69:250–8.
10. Hinchcliff M, Toledo DM, Taroni JN, et al. Mycophenolate mofetil treatment of systemic sclerosis reduces myeloid cell numbers and attenuates the inflammatory gene signature in skin. *J Invest Dermatol* 2018;138:1301–10.
11. Beretta L, Barturen G, Vigone B, et al. Genome-wide whole blood transcriptome profiling in a large European cohort of systemic sclerosis patients. *Ann Rheum Dis* 2020;79:1218–26.
12. Milano A, Pendergrass SA, Sargent JL, et al. Molecular subsets in the gene expression signatures of scleroderma skin. *PLoS One* 2008;3:e2696.
13. Pendergrass SA, Lemaire R, Francis IP, et al. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. *J Invest Dermatol* 2012;132:1363–73.
14. Taroni JN, Martyanov V, Huang CC, et al. Molecular characterization of systemic sclerosis esophageal pathology identifies inflammatory and proliferative signatures. *Arthritis Res Ther* 2015;17:194.
15. Franks JM, Martyanov V, Wang Y, et al. Machine learning predicts stem cell transplant response in severe scleroderma. *Ann Rheum Dis* 2020;79:1608–15.
16. Assassi S, Swindell WR, Wu M, et al. Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis. *Arthritis Rheumatol* 2015;67:3016–26.
17. Hinchcliff M, Huang CC, Wood TA, et al. Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis. *J Invest Dermatol* 2013;133:1979–89.
18. Chakravarty EF, Martyanov V, Fiorentino D, et al. Gene expression changes reflect clinical response in a placebo-controlled randomized trial of abatacept in patients with diffuse cutaneous systemic sclerosis. *Arthritis Res Ther* 2015;17:159.
19. Assassi S, Li N, Volkmann ER, et al. Predictive significance of serum interferon-inducible protein score for response to treatment in systemic sclerosis-related interstitial lung disease. *Arthritis Rheumatol* 2021;73:1005–13.

20. Bernstein EJ, Jaafar S, Assassi S, et al. Performance characteristics of pulmonary function tests for the detection of interstitial lung disease in adults with early diffuse cutaneous systemic sclerosis. *Arthritis Rheumatol* 2020;72:1892–6.
21. Frech TM, Revelo MP, Ryan JJ, et al. Cardiac metabolomics and autopsy in a patient with early diffuse systemic sclerosis presenting with dyspnea: a case report. *J Med Case Rep* 2015;9:136.
22. Gordon JK, Girish G, Berrocal VJ, et al. Reliability and validity of the tender and swollen joint counts and the modified Rodnan skin score in early diffuse cutaneous systemic sclerosis: analysis from the Prospective Registry of Early Systemic Sclerosis Cohort. *J Rheumatol* 2017;44:791–4.
23. Jaafar S, Lescoat A, Huang S, et al. Clinical characteristics, visceral involvement, and mortality in at-risk or early diffuse systemic sclerosis: a longitudinal analysis of an observational prospective multicenter US cohort. *Arthritis Res Ther* 2021;23:170.
24. Skaug B, Khanna D, Swindell WR, et al. Global skin gene expression analysis of early diffuse cutaneous systemic sclerosis shows a prominent innate and adaptive inflammatory profile. *Ann Rheum Dis* 2020;79:379–86.
25. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
26. Wong KL, Tai JJ, Wong WC, et al. Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets. *Blood* 2011;118:e16–31.
27. Van der Kroef M, Castellucci M, Mokry M, et al. Histone modifications underlie monocyte dysregulation in patients with systemic sclerosis, underlining the treatment potential of epigenetic targeting. *Ann Rheum Dis* 2019;78:529–38.
28. Dao DT, Anez-Bustillos L, Adam RM, et al. Heparin-binding epidermal growth factor-like growth factor as a critical mediator of tissue repair and regeneration. *Am J Pathol* 2018;188:2446–56.
29. Yona S, Kim KW, Wolf Y, et al. Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* 2013;38:79–91.
30. Williams M, Mildner A, Yona S. Developmental and functional heterogeneity of monocytes. *Immunity* 2018;49:595–613.
31. Dashti N, Mahmoudi M, Gharibdoost F, et al. Evaluation of ITGB2 (CD18) and SELL (CD62L) genes expression and methylation of ITGB2 promoter region in patients with systemic sclerosis. *Rheumatol Int* 2018;38:489–98.
32. Van Caam A, Aarts J, van Ee T, et al. TGF β -mediated expression of TGF β -activating integrins in SSc monocytes: disturbed activation of latent TGF β ? *Arthritis Res Ther* 2020;22:42.
33. Silvan J, Gonzalez-Tajuelo R, Vicente-Rabaneda E, et al. Deregulated PSGL-1 expression in B cells and dendritic cells may be implicated in human systemic sclerosis development. *J Invest Dermatol* 2018;138:2123–32.
34. Kim SN, Akindehin S, Kwon HJ, et al. Anti-inflammatory role of 15-lipoxygenase contributes to the maintenance of skin integrity in mice. *Sci Rep* 2018;8:8856.
35. Hooper JK, Eggink LL, Cote R. Stories from the dendritic cell guard-house. *Front Immunol* 2019;10:2880.
36. Sato K, Imai Y, Higashi N, et al. Lack of antigen-specific tissue remodeling in mice deficient in the macrophage galactose-type calcium-type lectin 1/CD301a. *Blood* 2005;106:207–15.
37. Salim PH, Jobim M, Bredemeier M, et al. Interleukin-10 gene promoter and NFKB1 promoter insertion/deletion polymorphisms in systemic sclerosis. *Scand J Immunol* 2013;77:162–8.
38. Chen S, Bonifati S, Qin Z, et al. SAMHD1 suppresses innate immune responses to viral infections and inflammatory stimuli by inhibiting the NF- κ B and interferon pathways. *Proc Natl Acad Sci U S A* 2018;115: E3798–807.
39. Shen YM, Zhao Y, Zeng Y, et al. Inhibition of Pim-1 kinase ameliorates dextran sodium sulfate-induced colitis in mice. *Dig Dis Sci* 2012;57: 1822–31.
40. Tsoi LC, Spain SL, Knight J, et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet* 2012;44:1341–8.
41. Ko CY, Chang WC, Wang JM. Biological roles of CCAAT/Enhancer-binding protein δ during inflammation. *J Biomed Sci* 2015;22:6.
42. Chang LH, Huang HS, Wu PT, et al. Role of macrophage CCAAT/enhancer binding protein delta in the pathogenesis of rheumatoid arthritis in collagen-induced arthritic mice. *PLoS One* 2012;7: e45378.
43. Mass E, Ballesteros I, Farlik M, et al. Specification of tissue-resident macrophages during organogenesis. *Science* 2016;353:aaf4238.
44. Li X, Zhang X, Xia J, et al. Macrophage HIF-2 α suppresses NLRP3 inflammasome activation and alleviates insulin resistance. *Cell Rep* 2021;36:109607.
45. Wang X, Abraham S, McKenzie JA, et al. LRG1 promotes angiogenesis by modulating endothelial TGF- β signalling. *Nature* 2013;499: 306–11.
46. Mehta BK, Espinoza ME, Hinchcliff M, et al. Molecular “omic” signatures in systemic sclerosis. *Eur J Rheumatol* 2020;7 Suppl:S173–80.
47. Johnson ME, Mahoney JM, Taroni J, et al. Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts. *PLoS One* 2015;10:e0114017.
48. Taroni JN, Greene CS, Martyanov V, et al. A novel multi-network approach reveals tissue-specific cellular modulators of fibrosis in systemic sclerosis. *Genome Med* 2017;9:27.
49. Lofgren S, Hinchcliff M, Carns M, et al. Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity. *JCI Insight* 2016;1:e89073.
50. Kobayashi S, Nagafuchi Y, Okubo M, et al. Integrated bulk and single-cell RNA-sequencing identified disease-relevant monocytes and a gene network module underlying systemic sclerosis. *J Autoimmun* 2021;116:102547.
51. Xue D, Tabib T, Morse C, et al. Expansion of FCGR3A(+) macrophages, FCN1(+) mo-DC, and plasmacytoid dendritic cells associated with severe skin disease in systemic sclerosis. *Arthritis Rheumatol* 2022;74:329–41.
52. Gur C, Wang SY, Sheban F, et al. LGR5 expressing skin fibroblasts define a major cellular hub perturbed in scleroderma. *Cell* 2022;185: 1373–88.
53. Yang M, Goh V, Lee J, et al. Clinical phenotypes of patients with systemic sclerosis with distinct molecular signatures in skin. *Arthritis Care Res* 2023. doi: [10.1002/acr.24998](https://doi.org/10.1002/acr.24998).
54. Spiera RF, Gordon JK, Mersten JN, et al. Imatinib mesylate (Gleevec) in the treatment of diffuse cutaneous systemic sclerosis: results of a 1-year, phase IIa, single-arm, open-label clinical trial. *Ann Rheum Dis* 2011;70:1003–9.
55. Gordon JK, Martyanov V, Magro C, et al. Nilotinib (Tasigna) in the treatment of early diffuse systemic sclerosis: an open-label, pilot clinical trial. *Arthritis Res Ther* 2015;17:213.