Song Jing (Orcid ID: 0000-0002-8881-9664)
Cuda Carla M (Orcid ID: 0000-0002-2022-8171)
Assassi Shervin (Orcid ID: 0000-0002-8059-9978)
Frech Tracy (Orcid ID: 0000-0002-5472-3840)
Khanna Dinesh (Orcid ID: 0000-0003-1412-4453)
Perlman Harris R (Orcid ID: 0000-0003-4174-608X)
Hinchcliff Monique Evangeline (Orcid ID: 0000-0002-8652-9890)

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

Hadijat-Kubura M. Makinde PhD[1], Julia L. M. Dunn PhD[1,2], Gaurav Gadhvi MS[1], Mary Carns MS [1], Kathleen Aren MPH[1], Anh H. Chung BS[1], Lutfiyya N. Muhammad PhD[3], Jing Song, MS[3], Carla M. Cuda PhD[1], Salina Dominguez MS[1], John E. Pandolfino MD[4], Jane E. Dematte D'Amico MD [5], G. Scott Budinger MD[5], Shervin Assassi MD MS[6,7], Tracy M. Frech MD MS[6,8], Dinesh Khanna MD MS[6,9], Alex Shaeffer MD[1], Harris Perlman PhD[1,*], Monique Hinchcliff MD MS[1,6,10,*], Deborah R. Winter PhD[1,*]

1. Northwestern University, Feinberg School of Medicine Department of Medicine, Division of Rheumatology. Chicago, IL 60611

2. Cincinnati Children's Hospital Medical Center, Division of Allergy & Immunology. Cincinnati, OH 45229 (current affiliation).

3. Northwestern University, Feinberg School of Medicine Department of Preventive Medicine. Chicago, IL 60611

4. Northwestern University, Feinberg School of Medicine, Department of Medicine, Division of Gastroenterology and Hepatology. Chicago, IL 60611.

5. Northwestern University, Feinberg School of Medicine, Department of Medicine, Division of Division of Pulmonary and Critical Care. Chicago, IL 60611

6. Prospective Registry of Early Systemic Sclerosis (PRESS) consortium.

Shervin Assassi MD MS- University of Texas Health Sciences Center at Houston (TX)

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/art.42380

Elana Bernstein MD MS- Columbia University (NY)

Robyn Domsic MD MS - University of Pittsburgh (PA)

Tracy Frech MD MS - University of Utah (UT)

Jessica Gordon - Hospital for Special Surgery (NY)

Faye Hant - Medical University of South Carolina (SC)

Monique Hinchcliff – Yale School of Medicine (CT)

Dinesh Khanna MD MS - University of Michigan (MI)

Ami Shah - Johns Hopkins University (MD)

Victoria Shanmugam - George Washington University (DC)

7. University of Texas Health Science Center at Houston, Division of Rheumatology, Houston, Texas 77030.

8. Vanderbilt University, Department of Medicine, Division of Rheumatology and Immunology. Nashville, TN 37232

9. University of Michigan, Department of Medicine, Division of Rheumatology. Ann Arbor, MI 48109.

10. Yale University, School of Medicine, Section of Rheumatology, Allergy & Immunology. New Haven, CT 06520

* These authors are equal corresponding authors.

* Harris Perlman, PhD

Northwestern University, Feinberg School of Medicine

Department of Medicine, Division of Rheumatology

240 East Huron Street, Room M300

Chicago, IL 60611, USA

h-perlman@northwestern.edu

Phone: 312-503-1955

Fax: 312-503-0994

* Monique Hinchcliff, MS, MD

Yale University, School of Medicine

Section of Rheumatology, Allergy & Immunology.

New Haven, CT 06520

Email: monique.hinchcliff@yale.edu

Phone: 203-785-6855

Fax: 203-785-3229

* Deborah R Winter, PhD

Northwestern University, Feinberg School of Medicine

Department of Medicine, Division of Rheumatology

240 East Huron Street, Room M300

Chicago, IL 60611, USA

Email : deborah.winter@northwestern.edu

Phone : 312-503-0535

Fax: 312-503-0994

Acknowledgements

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 (NU) Core Center for Clinical Research (CCCR) in the Division of Rheumatology (5P30AR072579) and Northwestern University Skin Biology and Diseases Resource-based Center (5P30AR075049). This research was supported in part through the computational resources and staff contributions provided by the Genomics Compute Cluster, which is jointly supported by the Feinberg School of Medicine, the Office of the Provost, the Office for Research and Northwestern Information Technology. HMM was supported by KL2 TR001424, HL134375S1, and AR007611. LNM and JS were supported by NIH/NIAMS P30 AR072579. CMC was supported by the Lupus Research Alliance (Novel Research Grant), the Rheumatology Research Foundation (Innovative Research Grant) and Northwestern University Dixon Translational Research Initiative. JEP was supported by P01 DK117824. 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. G.R.S.B. was supported by NIH grants U19AI135964, P01AG049665, P01AG04966506S1, R01HL147575 and Veterans Affairs grant I01CX001777. SA was supported by R01AR073284, R56AR078211 and the National Scleroderma Foundation. Dinesh Khanna and the PRESS clinical database were supported by NIH/NIAMS K24 AR063120. MH was supported by funding from an NIH grants R01 AR073270 and R01 HL152677. DRW was supported by

funding from Arthritis National Research Foundation, American Federation for Aging Research, American Heart Association (18CDA34110224), National Scleroderma Foundation, American Thoracic Society, and NIH grant AI163742. HP was supported by AR074902, AR075423, CA060553, HL134375, RRF Innovative Research Grant, United States-Israel Binational Science Foundation Investigator Grant, the Precision Medicine Fund, and the Mabel Green Myers Professorship.

The authors declare no competing financial interests.

ABSTRACT

Background/Purpose: 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. Here, we focused on the two main subtypes of circulating monocytes, classical (CM) and non-classical (NCM).

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

Results: We used an unbiased approach to cluster patients into three groups (A-C) based on their transcriptional signatures of CM relative to controls. Further, each group maintained their characteristic transcriptional signature in NCM. Genes upregulated in Group C demonstrated the highest signature compared to the other groups in skin macrophages. Patients from Group B and C exhibited worse lung function than Group A, although there was no difference in skin disease at baseline. We validated our approach by applying our group classifications to published bulk monocyte RNA-seq data on SSc patients: we found that patients with no skin disease were most likely to be classified as Group A.

Conclusion: We are the first to show that transcriptional signature of CM and NCM 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 heterogenous in presentation and disease course, and may involve the skin, blood vessels, lungs, heart, kidneys, and gastrointestinal (GI) tract. SSc can be subtyped by cutaneous involvement, including limited (lcSSc), diffuse (dcSS), and non-cutaneous (ncSSc, more recently known as 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 immunological pathways [1]. Circulating monocytes exist in three main states, characterized by CD14 and CD16 in humans: classical (CM) ($CD14^{++}CD16^{-}$), they constitute the majority of monocytes, intermediate (IM) ($CD14^{+}CD16^{+}$), and non-classical (NCM) ($CD14^{low}CD16^{+}$), < 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 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-ILD [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 are detected in a subset

of SSc patients on 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 relate transcriptional signatures from peripheral blood mononuclear cells (PBMC) or whole skin of SSc patients to disease activity [5, 11]. We and others identified four 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 two groups [13, 16]. Moreover, patients with an inflammatory gene expression signature in skin are the most likely to demonstrate skin disease improvement during MMF [17]. However, there were patients classified in the inflammatory intrinsic subset who were MMF non-responders [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 CM and NCM in patients from the Prospective Registry of Early Systemic Sclerosis (PRESS) consortium, a multicenter incident cohort study of patients with diffuse cutaneous SSc, that has the goal of advancing the understanding of disease pathogenesis and identifying novel biomarkers for patients [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.

METHODS

Study participant, recruitment and sample collection:

Patients were recruited to the Prospective Registry of Early Systemic Sclerosis (PRESS) registry at Northwestern University (NU), University of Texas, University of Utah, and University of Michigan (STU00062447) as previously described [23]. Consecutive patients in the PRESS registry from four PRESS sites (University of Michigan, Northwestern University, University of Texas-Houston and University of Utah) were enrolled in this sub-study if they were immunosuppressant naïve with plans to start a new treatment for SSc from September 2015 through June 2016, although some samples were collected at a subsequent visit post-treatment. At the time of sample collection 7 out of 14 patients were on immunosuppressants: Six patients were on MMF and one patient was on rituxan 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 (STU00045513) protocol. For skin biopsies, patients and healthy controls were recruited independently from NU (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 antibody cocktail including anti-CD45 BB515 (BD Horizon), anti-CD14 PerCP-Cyanine5.5 (eBioscience), anti-HLA-DR eFlour 450 (eBioscience), anti-CD15 APC (Biolegend), anti-CD16 APC-Cy7 (BD Biosciences), anti-CD1c PE (Miltenyi Biotec), anti-CD20 Alexa Flour 700 (BD Biosciences), anti-CD56 PE-

Cy7 (BD Biosciences). Cell sorting was performed on the FACSAria III (BD Bioscience) at Northwestern University RLHCCC Flow Cytometry Core Facility.

Skin biopsy digestion and macrophage sorting:

Fat was discarded, and epidermal and dermal tissues were separated, tissue was mechanically digested with pre-set 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 (BD), anti-HLA-DR eFluor 450 (eBioscience), 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 manufacturer's instructions (Arcturus). Full-length RNA-seq libraries from CM and skin were prepared using SMART-Seq v4 Ultra Low Input RNA Kit (Clontech Laboratories) followed by Nextera XT protocol (Illumina). For NCM samples, the QuantSeq FWD (Lexogen GmbH, Vienna, Austria) was used to generate Illumina compatible libraries. Libraries were sequenced on a NextSeq 500 instrument (Illumina) with a minimum of 5 x 10^6 reads per sample and commercially available universal human RNA (uhRNA) was included for reference.

RNA-seq alignment and mapping:

Sequenced reads were de-multiplexed (bcl2fastq), trimmed (Trimmomatic0.36 for full-length, bbduk for 3'), aligned hg19 (TopHat2 2.17.1.14for full-length or STAR (version 2.5.2) for 3' data sets (NCM). Samples with fewer than 5 x 10^6 uniquely aligned reads or with complexity

(percentage of reads mapping to unique genomic positions) less than 30% were excluded due to low quality. As this included one of the 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.GR*Ch37.75* as reference. Gene coverage (5'à3') for full-length libraries were 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 dataset were defined as those with at least three samples above the minimum threshold (CM: FPKM=5, NCM: CPM=31, Skin: FPKM=3).

Identification of CM transcriptional signature in SSc:

R Studio (https://www.rstudiocom/) was used to calculate Pearson's correlation coefficients and to generate PCA plots, Venn diagrams (ggplots package), and scatterplots (ggplot2 package). PRISM (GraphPad Software LLC) was used to generate bar graphs of individual gene expression. Differential expression was calculated using the DEseq package (version 1.18.1) on 5171 expressed genes; genes with changes in expression of two-fold or greater and Benjamini Hochberg adjusted p-value < 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 with 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 standard deviation of expression for gene *i* across controls. To prevent artificially high z scores resulting from small standard deviations, where high z score may not represent true biological variability, we set the minimum σ_{Ci} to 2.5. We identified 1790 genes with $|z'| > 2$ in at least 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 whether how often the patients ended up in the same clade out of 1000 iterations with varying subset sizes. DEseq was run and DEGs defined as above on each patient group compared with controls.

Gene Set Enrichment Analysis (GSEA):

To determine the biological processes of enriched genes in each Group, 5171 expressed genes were ranked by Group based on their fold-change relative to the mean expression across controls and given as pre-ranked input for Gene Set Enrichment Analysis (GSEA v3.0, www.gsea-msigdb.org). Pre-ranked genes were compared to the biological processes (BP) of the Molecular Signature Database (MSigDB) C5v7.0 collection using pre-ranking and classic weighting. Normalized enrichment scores (NES) were reported, and significant enrichment (or depletion) was defined as adjusted p value < 0.05.

Comparison with NCM transcriptional signature:

Starting with the 1790 genes in CM clusters, we identified 1599 among the NCM expressed genes and calculated the modified z-score relative to NCM controls. Hierarchical clustering was performed on these genes based their NCM z-score and dendrogram confidence and group robustness were assessed as described above. DEseq was performed as above on 7143 expressed genes to define DEG between each patient group and controls and ggplot2 was used to compare DEG in CM vs. NCM. GSEA was run as above based on fold-change in NCM.

Expression of published monocyte genes:

We defined 431 CM and 388 NCM genes based on a publicly available microarray dataset [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 (GraphPad Software LLC) was used to the graphs of this data.

Skin macrophages:

We used a non-parametric Mann-Whitney U test to define SSc-specific (190 genes) and control-specific (138 genes) genes out of 6338 expressed genes in skin macrophages from patients compared to controls with a p-value <0.05. GENE-E was used to visualize the relative expression of these genes as a heatmap with Min-Max 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 ($P<0.05$). ROC curve (pROC package in R) was used to

determine the performance of Clusters I-IV at delineating patient vs control status. Each sample was scored as the mean of the min-max value for all genes in the given cluster; p-values were calculated by the package using Mann-Whitney U.

Statistical analysis of clinical data:

Baseline 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 in PRISM (GraphPad Software LLC). Additionally, the longitudinal FVC and mRSS data are displayed as spaghetti plots by PRISM.

Comparison with published SSc monocyte data:

All information on patient diagnosis, grouping and sample acquisition were obtained from the publicly available dataset by van der Kroef et al [27]. Starting with 1790 genes in CM cluster, we identified 1523 that were expressed in bulk monocyte RNA-seq from this dataset. For each gene, we normalized using min-max scaling and calculated the mean score across genes in each cluster for each individual. ROC curves were plotted as above to determine the performance of Clusters I-IV at delineating each SSc patient category from controls.

Results:

Classical Monocytes from SSc Patients Exhibit Greater Variability versus 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 classical (CM) and non-classical (NCM) monocytes by flow cytometry and found no significant differences in the proportion detected among $CD45^+$ immune cells between the SSc and control cohorts (**Figure 1B-C**). We performed RNA-seq to profile the gene expression of CM, the predominant monocyte population. One sample was eliminated that did not pass quality control tests, and we confirmed patient gender as an additional quality control (**Supp. Figure 1A-D**). 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, Supp 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 control" (Supp **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 **(Supp Figure 1G**). This variability did not seem to be associated with the site of sample collection **(Supp 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, Supp Table 1A,B**). The 146 genes include type I interferon signaling pathway (*OAS1*, OAS2, *IRF2*), response to cAMP (*FOS*, *FOSB, JUN*), regulation of vascular development (*VEGF*A, KLF2, SERPINE1) and regulation of leukocyte migration (*CSF1R*, *CCR2, CCL2*).

SSc Classical Monocytes Cluster into Three 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 more than 2 standard deviations from the control mean. By hierarchically clustering the SSc patients based on the 1790 genes that were differential in at least 3 patients, we identified 3 patient subgroups (Groups A-C). We also note that one patient (SSc15) was highly similar to controls and not included in any group (**Figure 2A, Supp Table 2A**). The clustering of 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 (**Supp Figure 2A-B)**. Moreover, by k-means clustering of the 1790 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 with Controls (**Figure 2A-D**). The distinct transcriptional signatures of each Group were reinforced when visualized using Principal Component Analysis (PCA) (**Figure 2E).** Patients in Group A tended to have the most differentially expressed genes and both Group A and B demonstrated overlap with the genes identified by the prior full-cohort analysis (**Figure 2F**). On the other hand, Group C patients are characterized by the expression of genes in CM 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 Groups relative to Controls for all expressed genes in the dataset to better characterize their transcriptional profiles

(**Figure 2G-H, Supp Figure 2C, E, G, Supp Table 2B-C).** We found that Group A upregulated genes were associated with type I interferon 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 re-sampling (**Supp Figure 2B**). Like Group A, Group B genes were associated with type I interferon pathways and innate immunity but exhibited higher enrichment of chemokine production and response to cAMP and corticosteroids. Both Group B and C upregulated genes associated with vasculature development and blood vessel morphogenesis, such as *VEGFA* and *HBEGF* [28]. Group C exhibited additional genes associated with TGF-β, tube formation, and BMP signaling, including *TGFB1*, *SMAD3*, *SKIL, PPARD, CXCR4* and *PTPN12*.

Patient Groups are Largely Conserved in Non-Classical Monocytes

Next, we analyzed RNA-seq data from NCM isolated from the same patients (**Supp Figure 3A**). The variation in gene expression was higher in NCM than CM in both patients and controls (**Supp Figure 3B-D**), but this may be due to technical differences in the protocols or higher sampling bias from the smaller NCM population. NCM were clustered based on the 1790 CM genes that defined the Groups A, B and C to address the conservation of groups on a global scale without introducing new genes (**Figure 3A, Supp Table 3A**). Patients in Group A and C continued to cluster next to each other, but Group B no longer formed its own clade (Figure 3A, **Supp Figure 3E-G**). As with CM, we performed pairwise analyses for DEG for each of the three Groups (**Supp Figure 3H, Supp Table 3B-C**). To complement our approach in Figure 3A, we then compared the list of DEG between the previously defined groups to determine whether individual genes were

conserved and identify genes that may be uniquely differential in NCM (**Figure 3B-F and Supp Figure 3I**). We found that genes related to interferon and innate immunity were preferentially expressed in Group A in both CM and NCM. Although Group B did not exhibit as many significantly upregulated genes in NCM as in CM, Group B NCM genes were associated with similar processes as CM. Group C patients exhibited far fewer NCM DEG but those associated with its unique transcriptional signature – such as *CXCR4*, *PTPN12* and *SKIL* – were significantly upregulated.

Monocyte identity is Altered Across Patients

It has been proposed using mouse models that a subset of CM that have egressed from the bone marrow differentiate into NCM [29]. The genes that distinguish NCM from CM 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] (**Supp Table 4)**. We found that both CM and NCM from Group A patients exhibited higher expression of their respective genes than controls – including *PLAC8, PADI4, SELL, LYZ*, and *CCR2* in CM and I*TGAL, SPN, CX3CR1*, and *CSF1R* in NCM (**Figure 4A-D**). This was a cell-type-specific effect as we did not observe increased expression of these genes in the opposite population (Supp Figure 4A-D). Group B also exhibited increased classical monocytes genes only in CM, while Group C was 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 (**Supp Figure 4E-F**). In comparing CM and NCM, we may have expected that the genome-wide transcriptional profile would be most similar between monocytes from the same patient; however, this is not the case (**Figure 4E**). The over-expression

of CM genes in Groups A and B may explains why CM gene expression from this SSc cohort is less similar compared to NCM **(Supp Figure 4E, F)**.

Skin Macrophages Upregulate SSc-Associated Monocyte Genes

Since we observed common transcriptional signatures across monocyte populations in SSc and control cohorts, we investigated whether similar gene expression profiles are observed in skin macrophages. We isolated $CD206^{+}HLADR^{+}$ macrophages from fibrotic skin of 7 SSc patient biopsies (**Figure 5A, Supp Figure 5A, B and C**) and from 7 age, sex, and race-matched healthy controls and performed RNA-seq (**Supp Figure 5D, E**). We identified 328 DEG between skin macrophages in SSc patients vs. controls (**Figure 5B and Supp Figure 5F, Supp 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 *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 Clusters I-III genes, which corresponded to upregulated genes in Groups A, B, and C in our CM analysis, were more likely to be SSc-specific than control-specific (Enrichment: I=1.54, II=1.82, III=3.27), but only Cluster III was significantly enriched (p=0.0412) (**Figure 5D, Supp Figure 5G**). Similarly, classification strategies based on the average expression of Clusters I and III genes in skin macrophages were able to differentiate between SSc patients and controls significantly better than

random (p= 0.03, 0.01) (**Figure 5E**). On the other hand, Cluster IV genes, which were largely downregulated in monocytes, were generally expressed higher in controls than patients (Enrichment: IV=0.40, p=0.029).

Patients Subgroups Differ in Disease Characteristics

Our analysis stratified SSc patients based on CM transcriptional profiles while remaining agnostic to clinical presentation such as cutaneous and pulmonary involvement. Using baseline and longitudinally collected FVC, mRSS data up to 42 months after blood sample collection, we found that Group B and C patients exhibited significant FVC decline indicating poor lung function at baseline compared to Group A patients that did not tend to improve over time (**Figure 6A, C**). In contrast, there were no significant differences in mRSS levels between groups at baseline or longitudinally (**Figure 6B, D**).

Application of Transcriptional Groups to Independent Dataset

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 artefacts, protocol choice, sequencing depth, and normalization. To test the algorithm, we used gene expression data from an unrelated study on monocytes in SSc patients performed by another group [27]. In this study, they isolated bulk monocytes for RNA-seq, and categorized patients as dcSSc, lcSSc, or non-cutaneous (ncSSc/Sine). For each patient, we calculated a score to reflect the relative level of expression of each cluster from our CM analysis (**Figure 6E, Supp Figure 6, Supp 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 four gene

Clusters and assessed their ability to correctly assign a patient to the correct SSc subtype (**Figure 6F**). As expected, Cluster IV, which was downregulated 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 on ncSSc patients. Cluster 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 results 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 CM and NCM 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 segregated into three groups based on CM transcriptional profiles, with each group demonstrating upregulation of functionally distinct gene sets including response to type I interferon, myeloid leukocyte mediated immunity and regulation of GPCR signaling pathway for group A, interleukin 1 and chemokine production, response to cAMP and response to corticosteroid for group B and tube/vessel formation, response to BMP, and TGF-ß receptor signaling pathway for group C compared to each other and to health participants. Since these three groups are largely recapitulated in analysis of NCM transcription profiling, these data suggest some conservation is the gene signatures from Groups A, B and C. Analysis of bulk skin $CD206^{+}HLADR^{+}$ macrophages revealed that genes associated with Groups A, B and C tend to increase in expression in skin macrophages from SSc patients compared with 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 replicate these results in paired skin and blood, which will

enable direct comparison of immature and mature myeloid gene expression. The Group A SSc patients display the highest FVC compared to either Group B or C but exhibits no change in mRSS. Together, our results suggest that transcriptional profiling of distinct CM and NCM subpopulations may represent a viable mechanism for identifying 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 interferon as well as the IL-1β pathway in bulk SSc monocytes (CM and NCM), neither study examined any clinical associations [27, 50]. Currently, there are only two 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 are that it clearly illustrates

the presence of disease subsets –or endotypes –in circulating immune cells. Further, our data differentiate subtypes of SSc patients in another study, which are only stratified based 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 three patient groups without apparent distinguishing clinical phenotypes, further studies will establish 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 heterogenous SSc population in the validation cohort. Future studies that include larger number of patients, match blood and skin and 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 not been established between transcriptomic intrinsic subsets and clinical characteristics, including prognosis and response to treatment. A thorough molecular understanding of SSc may help to stratify patients for enrollment in clinical trials and to inform drug selection. Further investigation of CM and NCM could determine how SSc subsets respond to targeted interventions and whether they can predict which end organ would respond to a particular therapeutic.

References

1. Assassi S, Radstake TR, Mayes MD, Martin J: **Genetics of scleroderma: implications for personalized medicine?** *BMC Med* 2013, **11**:9.
2. Hu Y, Hu Y, Xiao Y, Wen F, Zhang S, Liang D, Su L, Deng Y, Luo J, Ou J *et al*: **Genetic landscape and autoimmunity of monocytes in developing Vogt-Koyanagi-Harada disease**. *Proc Natl Acad Sci U S A* 2020, **117**(41):25712-25721.
3. Roberts ME, Barvalia M, Silva J, Cederberg RA, Chu W, Wong A, Tai DC, Chen S, Matos I, Priatel JJ *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**(10):e61-e79.
4. van der Kroef M, van den Hoogen LL, Mertens JS, Blokland SLM, Haskett S, Devaprasad A, Carvalheiro T, Chouri E, Vazirpanah N, Cossu M *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**(1):119-129.
5. Toledo DM, Pioli PA: **Macrophages in Systemic Sclerosis: Novel Insights and Therapeutic Implications**. *Curr Rheumatol Rep* 2019, **21**(7):31.
6. Misharin AV, Morales-Nebreda L, Reyfman PA, Cuda CM, Walter JM, McQuattie-Pimentel AC, Chen CI, Anekalla KR, Joshi N, Williams KJN *et al*: **Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span**. *J Exp Med* 2017, **214**(8):2387-2404.
7. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, Fernandez R, Akbarpour M, Chen CI, Ren Z *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**(12):1517-1536.

8. Higashi-Kuwata N, Jinnin M, Makino T, Fukushima S, Inoue Y, Muchemwa FC, Yonemura Y, Komohara Y, Takeya M, Mitsuya H *et al*: **Characterization of monocyte/macrophage subsets in the skin and peripheral blood derived from patients with systemic sclerosis**. *Arthritis Res Ther* 2010, **12**(4):R128.
9. Arai M, Ikawa Y, Chujo S, Hamaguchi Y, Ishida W, Shirasaki F, Hasegawa M, Mukaida N, Fujimoto M, Takehara K: **Chemokine receptors CCR2 and CX3CR1 regulate skin fibrosis in the mouse model of cytokine-induced systemic sclerosis**. *J Dermatol Sci* 2013, **69**(3):250-258.
10. Hinchcliff M, Toledo DM, Taroni JN, Wood TA, Franks JM, Ball MS, Hoffmann A, Amin SM, Tan AU, Tom K *et al*: **Mycophenolate Mofetil Treatment of Systemic Sclerosis Reduces Myeloid Cell Numbers and Attenuates the Inflammatory Gene Signature in Skin**. *J Invest Dermatol* 2018.
11. Beretta L, Barturen G, Vigone B, Bellocchi C, Hunzelmann N, De Langhe E, Cervera R, Gerosa M, Kovacs L, Ortega Castro R *et al*: **Genome-wide whole blood transcriptome profiling in a large European cohort of systemic sclerosis patients**. *Ann Rheum Dis* 2020, **79**(9):1218-1226.
12. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, Whitfield ML: **Molecular subsets in the gene expression signatures of scleroderma skin**. *PLoS One* 2008, **3**(7):e2696.

13. Pendergrass SA, Lemaire R, Francis IP, Mahoney JM, Lafyatis R, Whitfield ML: **Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies**. *J Invest Dermatol* 2012, **132**(5):1363-1373.

14. Taroni JN, Martyanov V, Huang CC, Mahoney JM, Hirano I, Shetuni B, Yang GY, Brenner D, Jung B, Wood TA *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, Wood TA, Pinckney A, Crofford LJ, Keyes-Elstein L, Furst DE, Goldmuntz E, Mayes MD *et al*: **Machine learning predicts stem cell transplant response in severe scleroderma**. *Ann Rheum Dis* 2020, **79**(12):1608-1615.

16. Assassi S, Swindell WR, Wu M, Tan FD, Khanna D, Furst DE, Tashkin DP, Jahan-Tigh RR, Mayes MD, Gudjonsson JE *et al*: **Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis**. *Arthritis Rheumatol* 2015, **67**(11):3016-3026.

17. Hinchcliff M, Huang CC, Wood TA, Matthew Mahoney J, Martyanov V, Bhattacharyya S, Tamaki Z, Lee J, Carns M, Podlusky S *et al*: **Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis**. *J Invest Dermatol* 2013, **133**(8):1979-1989.

18. Chakravarty EF, Martyanov V, Fiorentino D, Wood TA, Haddon DJ, Jarrell JA, Utz PJ, Genovese MC, Whitfield ML, Chung L: **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, Mayes MD, Runger D, Ying J, Roth MD, Hinchcliff M, Khanna D, Frech T *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**(6):1005-1013.

20. Bernstein EJ, Jaafar S, Assassi S, Domsic RT, Frech TM, Gordon JK, Broderick RJ, Hant FN, Hinchcliff ME, Shah AA *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**(11):1892-1896.
21. Frech TM, Revelo MP, Ryan JJ, Shah AA, Gordon J, Domsic R, Hant F, Assassi S, Shanmugam VK, Hinchcliff M *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, Zhang M, Hatzis C, Assassi S, Bernstein EJ, Domsic RT, Hant FN, Hinchcliff M *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**(6):791-794.
23. Jaafar S, Lescoat A, Huang S, Gordon J, Hinchcliff M, Shah AA, Assassi S, Domsic R, Bernstein EJ, Steen V *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**(1):170.
24. Skaug B, Khanna D, Swindell WR, Hinchcliff ME, Frech TM, Steen VD, Hant FN, Gordon JK, Shah AA, Zhu L *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**(3):379-386.

25. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome Biol* 2014, **15**(12):550.

26. Wong KL, Tai JJ, Wong WC, Han H, Sem X, Yeap WH, Kourilsky P, Wong SC: **Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets**. *Blood* 2011, **118**(5):e16-31.

27. van der Kroef M, Castellucci M, Mokry M, Cossu M, Garonzi M, Bossini-Castillo LM, Chouri E, Wichers CGK, Beretta L, Trombetta E *et al*: **Histone modifications underlie monocyte dysregulation in patients with systemic sclerosis, underlining the treatment potential of epigenetic targeting**. *Ann Rheum Dis* 2019, **78**(4):529-538.

28. Dao DT, Anez-Bustillos L, Adam RM, Puder M, Bielenberg DR: **Heparin-Binding Epidermal Growth Factor-Like Growth Factor as a Critical Mediator of Tissue Repair and Regeneration**. *Am J Pathol* 2018, **188**(11):2446-2456.

29. Yona S, Kim KW, Wolf Y, Mildner A, Varol D, Breker M, Strauss-Ayali D, Viukov S, Guilliams M, Misharin A *et al*: **Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis**. *Immunity* 2013, **38**(1):79-91.

30. Guilliams M, Mildner A, Yona S: **Developmental and Functional Heterogeneity of Monocytes**. 2018, **49**(4):595-613.

31. Dashti N, Mahmoudi M, Gharibdoost F, Kavosi H, Rezaei R, Imeni V, Jamshidi A, Aslani S, Mostafaei S, Vodjgani M: **Evaluation of ITGB2 (CD18) and SELL (CD62L) genes expression and methylation of ITGB2 promoter region in patients with systemic sclerosis**. *Rheumatol Int* 2018, **38**(3):489-498.

32. van Caam A, Aarts J, van Ee T, Vitters E, Koenders M, van de Loo F, van Lent P, van den Hoogen F, Thurlings R, Vonk MC *et al*: **TGFbeta-mediated expression of**

TGFbeta-activating integrins in SSc monocytes: disturbed activation of latent TGFbeta? *Arthritis Res Ther* 2020, **22**(1):42.

33. Silvan J, Gonzalez-Tajuelo R, Vicente-Rabaneda E, Perez-Frias A, Espartero-Santos M, Munoz-Callejas A, Garcia-Lorenzo E, Gamallo C, Castaneda S, Urzainqui A: **Deregulated PSGL-1 Expression in B Cells and Dendritic Cells May Be Implicated in Human Systemic Sclerosis Development**. *J Invest Dermatol* 2018, **138**(10):2123-2132.

34. Kim SN, Akindehin S, Kwon HJ, Son YH, Saha A, Jung YS, Seong JK, Lim KM, Sung JH, Maddipati KR *et al*: **Anti-inflammatory role of 15-lipoxygenase contributes to the maintenance of skin integrity in mice**. *Sci Rep* 2018, **8**(1):8856.

35. Hoober JK, Eggink LL, Cote R: **Stories From the Dendritic Cell Guardhouse**. *Front Immunol* 2019, **10**:2880.

36. Sato K, Imai Y, Higashi N, Kumamoto Y, Onami TM, Hedrick SM, Irimura T: **Lack of antigen-specific tissue remodeling in mice deficient in the macrophage galactose-type calcium-type lectin 1/CD301a**. *Blood* 2005, **106**(1):207-215.

37. Salim PH, Jobim M, Bredemeier M, Chies JA, Brenol JC, Jobim LF, Xavier RM: **Interleukin-10 gene promoter and NFKB1 promoter insertion/deletion polymorphisms in systemic sclerosis**. *Scand J Immunol* 2013, **77**(2):162-168.

38. Chen S, Bonifati S, Qin Z, St Gelais C, Kodigepalli KM, Barrett BS, Kim SH, Antonucci JM, Ladner KJ, Buzovetsky O *et al*: **SAMHD1 suppresses innate immune responses to viral infections and inflammatory stimuli by inhibiting the NF-kappaB and interferon pathways**. *Proc Natl Acad Sci U S A* 2018, **115**(16):E3798-E3807.

39. Shen YM, Zhao Y, Zeng Y, Yan L, Chen BL, Leng AM, Mu YB, Zhang GY: **Inhibition of Pim-1 kinase ameliorates dextran sodium sulfate-induced colitis in mice**. *Dig Dis Sci* 2012, **57**(7):1822-1831.

40. Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, Capon F, Ding J, Li Y, Tejasvi T, Gudjonsson JE *et al*: **Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity**. *Nat Genet* 2012, **44**(12):1341-1348.

41. Ko CY, Chang WC, Wang JM: **Biological roles of CCAAT/Enhancer-binding protein delta during inflammation**. *J Biomed Sci* 2015, **22**:6.

42. Chang LH, Huang HS, Wu PT, Jou IM, Pan MH, Chang WC, Wang DD, Wang JM: **Role of macrophage CCAAT/enhancer binding protein delta in the pathogenesis of rheumatoid arthritis in collagen-induced arthritic mice**. *PLoS One* 2012, **7**(9):e45378.

43. Mass E, Ballesteros I, Farlik M, Halbritter F, Gunther P, Crozet L, Jacome-Galarza CE, Handler K, Klughammer J, Kobayashi Y *et al*: **Specification of tissue-resident macrophages during organogenesis**. *Science* 2016, **353**(6304).

44. Li X, Zhang X, Xia J, Zhang L, Chen B, Lian G, Yun C, Yang J, Yan Y, Wang P *et al*: **Macrophage HIF-2alpha suppresses NLRP3 inflammasome activation and alleviates insulin resistance**. *Cell Rep* 2021, **36**(8):109607.

45. Wang X, Abraham S, McKenzie JAG, Jeffs N, Swire M, Tripathi VB, Luhmann UFO, Lange CAK, Zhai Z, Arthur HM *et al*: **LRG1 promotes angiogenesis by modulating endothelial TGF-beta signalling**. *Nature* 2013, **499**(7458):306-311.

46. Mehta BK, Espinoza ME, Hinchcliff M, Whitfield ML: **Molecular "omic" signatures in systemic sclerosis**. *Eur J Rheumatol* 2020, **7**(Suppl 3):S173-S180.

47. Johnson ME, Mahoney JM, Taroni J, Sargent JL, Marmarelis E, Wu MR, Varga J, Hinchcliff ME, Whitfield ML: **Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts**. *PLoS One* 2015, **10**(1):e0114017.

48. Taroni JN, Greene CS, Martyanov V, Wood TA, Christmann RB, Farber HW, Lafyatis RA, Denton CP, Hinchcliff ME, Pioli PA *et al*: **A novel multi-network approach reveals tissue-specific cellular modulators of fibrosis in systemic sclerosis**. *Genome Med* 2017, **9**(1):27.

49. Lofgren S, Hinchcliff M, Carns M, Wood T, Aren K, Arroyo E, Cheung P, Kuo A, Valenzuela A, Haemel A *et al*: **Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity**. *JCI Insight* 2016, **1**(21):e89073.

50. Kobayashi S, Nagafuchi Y, Okubo M, Sugimori Y, Shirai H, Hatano H, Junko M, Yanaoka H, Takeshima Y, Ota 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, Yang Y, Domsic R, Khanna D, Lafyatis R: **Expansion of FCGR3A(+) macrophages, FCN1(+) mo-DC, and plasmacytoid dendritic cells associated with severe skin disease in systemic sclerosis**. *Arthritis Rheumatol* 2021.

52. Gur C, Wang SY, Sheban F, Zada M, Li B, Kharouf F, Peleg H, Aamar S, Yalin A, Kirschenbaum D *et al*: **LGR5 expressing skin fibroblasts define a major cellular hub perturbed in scleroderma**. *Cell* 2022, **185**(8):1373-1388 e1320.

Figure Legends:

Figure 1: Classical monocytes from early diffuse SSc patients display transcriptional heterogeneity. (A) Panel of PRESS patients' information. Age p-value = ns, Mann-Whitney U, and sex/race p-value = ns, Fishers Exact Test. **(B)** Contour plots depicting stepwise isolation of classical monocytes (CM) and non-classical monocytes (NCM) from blood by FACS. **(C)** The percent of CM and NCM in $CD45^+$ cells from SSc compared to control samples (p = 0.14 and 0.89, respectively. Mann-Whitney). **(D)** Scatterplot showing the coefficient of variation (standard deviation/average FPKM) relative to the average expression (log_2(FPKM+1)) in SSc (purple) and control (grey) CM. **(E)** Principal Components Analysis (PCA) of gene expression in CM samples from Treated patients (purple), Untreated patients (pink), and controls (grey). **(F)** Volcano plot showing differentially expressed genes in the SSc cohort (167) compared with controls log_2 fold change >= 1 or <=-1 (significance calculated by DEseq with Benjamini Hochberg FDR correction, purple indicates genes with $P_{adj} < 0.05$ & log_2 fold change >= 1 or <=-1). **C-F** based on 5171 expressed genes in CM.

Figure 2: Early diffuse SSc patients stratify into 3 groups based on the transcriptional profile of CM. (A) Heatmap of adapted z-score of CM gene expression relative to controls, depicting dendrogram resulting from unsupervised hierarchical clustering of samples into 3 groups (Group A=coral; Group B=cyan; Group C=green) and K-means clustering of 1790 differential genes by individual (|adapted z-score| > 2 in at least 3 patients) (rows) (k=4). Expression of representative genes from **(B)** Cluster I (592), **(C)** Cluster II (261), **(D)** Cluster III (500), and Cluster IV (437) in each patient group as well as controls. * Indicates $p_{adj} < 0.05$ calculated by DEseq with Benjamini Hochberg FDR correction (**E**) PCA of gene expression in CM samples color-coded based on

patient groups. **(F)** Number of genes with adapted Z scores greater than 2 or less than -2 in individual CM samples organized by patient groups. **(G)** Venn Diagram showing overlap of upregulated (($log_2FC > 1$, and p-adj < 0.05) or downregulated (($log_2FC < -1$, and p-adj < 0.05) genes in each group compared to controls. **(H)** Heatmap showing the FDR based on GSEA of biological processes associated with differential genes in the three groups. **A-H** based on 5171 expressed genes in CM.

Figure 3: Patient groups from CM exhibit distinct transcriptional profiles in NCM. **(A)** Heatmap of adapted z-score of NCM gene expression relative to controls using 1599 genes from CM clusters in 2A. **(B-D)** Expression of representative genes from clusters I-III respectively. * Indicates $p_{adj} < 0.05$ calculated by DEseq with Benjamini Hochberg FDR correction. **(E)** Scatterplot of Log_2FC of gene expression in each patient group compared to controls in CM (x-axis) versus NCM (y-axis). Significant genes (Log2FC =>1 or <= -1and P_{adj}<=0.05) are colored if they are shared between CM and NCM (purple), only NCM (pink), or only in CM (blue). **(F)** Bar graph of the Normalized Enrichment Scores (NES) based on GSEA of biological processes from Figure 2H in CM and NCM**.** Dashed lines indicate the 90% threshold for NES in CM (blue) and NCM (pink). **A-F** based on 7143 expressed genes in NCM.

Figure 4: Variation in expression of monocyte gene signature in SSc CM and NCM. (A) Average expression level of all CM genes in CM samples of patient groups and controls. **(B)** Average expression level of all NCM genes in NCM samples of patient groups and controls. * Indicates $p < 0.05$ Mann-Whitney nonparametric test. **(C)** Expression levels of representative CM genes in patient CM samples by group. * Indicates $p_{adj} < 0.05$ calculated by DEseq with Benjamini Hochberg FDR correction. **(D)** Expression levels of representative NCM genes in patient NCM

samples by group. * Indicates $p_{adj} < 0.05$ calculated by DEseq with Benjamini Hochberg FDR correction. **(E)** Pearson correlation of average gene expression between CM and NCM samples in patient groups and controls.

Figure 5: Clusters I and III genes are upregulated in SSc skin macrophages compared to control. (A) Panel of SSc patients and control subject information. Age p-value = ns, Mann-Whitney U, and sex/race p-value = ns, Fishers Exact Test. **(B**) Gating scheme depicting the purification of macrophages from skin. **(C)** Heatmap of 190 SSc-specific and 138 Control-specific genes based on significantly differential 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. **(E)** Number of genes from **B** that overlap CM Clusters I-IV in **2A**. * indicates $p < 0.05$ by Fisher's exact test. **(F)** ROC curve based on the sensitivity and specificity of distinguishing skin macrophages from SSc patients vs. controls using the average normalized expression of genes from CM Clusters I-IV (**2A).** B-E based on 6338 expressed genes in skin macrophages.

Figure 6: SSc patient groups differ 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)** modified Rodnan Skin Score (mRSS) by patient group at baseline (time of blood sample). **(D)** mRSS of each patient colored by group longitudinally over 42 months from baseline (time 0). **(E)** Average normalized expression of genes from CM Clusters I-IV (**2A**) in bulk monocytes from SSc patients categorized by disease phenotype (dataset described in van der Kroef, et al). **(F)** ROC

curve based on the sensitivity and specificity of distinguishing bulk monocytes from SSc patient categories vs. controls using the average normalized expression of genes from CM Clusters I-IV (**2A)** (dataset described in van der Kroef, et al).

Supplemental Figures.

Supplemental Figure 1: Quality control analysis of classical monocyte RNA-seq. (A) Histogram indicating the number of genes expressed at the given Log_2(FPKM) in each CM sample. Red line indicates the threshold for 5171 expressed genes (FPKM = 5) in at least 3 samples. **(B)** Normalized 5'-3' gene coverage across length of genes. Sum of FPKM mapped to **(C)** Y-chromosome and **(D)** X-chromosome genes for each SSc CM sample. **(E)** Coefficient of variation histogram between control and SSc samples. **F)** Correlation (Pearson's coefficient) of CM expression in each SSc patient relative to the average gene expression across control samples. **(G)** Pairwise Pearson correlation between gene expression of CM samples from SSc patients. **(H)** PCA of control and SSc samples are color coded based on sample collection sites. MI= University of Michigan. (green), NU, Northwestern University (purple), UT= University of Texas at Houston (orange).

Supplemental Figure 2: Differentially expressed genes vary by group in CM. Validation of the hierarchical clustering of CM gene expression in SSc patients. **(A)** Resampling of 1790 genes from 2A to calculate approximately unbiased (AU) probability (red) and bootstrap probability (BP) (green) of each edge (order in grey). **(B)** The percent of subset samples at various sizes that successfully recapitulate each patient group (A, B, C) or all patient groups. Volcano plot showing differential expression of genes in **(C)** Group A, 421 genes up regulated, and 433 genes down regulated **(E)** Group B, 161 genes upregulated, and 47 genes down regulated and **(G)** Group C, 263 genes up regulated and 66 genes down regulated vs control. FDR was calculated independently for each group taking into account the total number of expressed genes. Purple dots are $P_{adj} < 0.05$

& log_2 fold change >= 1 or <=-1. Representative enrichment plot from GSEA software for **(D)** Group A, **(F)** Group B and **(H)** Group C.

Supplemental Figure 3: CM and NCM share transcriptional features in SSc patients. (A) Histogram indicating the number of genes expressed at the given Log_2(CPM) in each NCM sample. Red line indicates the threshold for 7143 expressed genes (CPM = 31) in at least 3 samples. **(B)** Coefficient of variation histogram between control and SSc NCM samples. **(C-D)** Coefficient of variation histogram between CM and NCM of **(C)** SSc patients and **(D)** Controls. Validation of the hierarchical clustering of NCM gene expression in SSc patients. **(E)** Resampling of 1599 genes from 3A to calculate approximately unbiased (AU) probability (red) and bootstrap probability (BP) (green) of each edge (order in grey). **(F)** The percent of subset samples at various sizes that successfully recapitulate each patient group (A, B, C) or all patient groups. **(G)** Comparison of the hierarchical clustering dendrograms for SSc CM and NCM gene expression. **(H)** Venn Diagram showing overlap of upregulated (log_2FC >1 and p-adj < 0.05) or downregulated (log_2FC < -1 and p-adj < 0.05) genes in each patient group compared to controls. **(I)** Adapted z-score of relative expression in CM and NCM for representative genes associated with conserved processes patient groups.

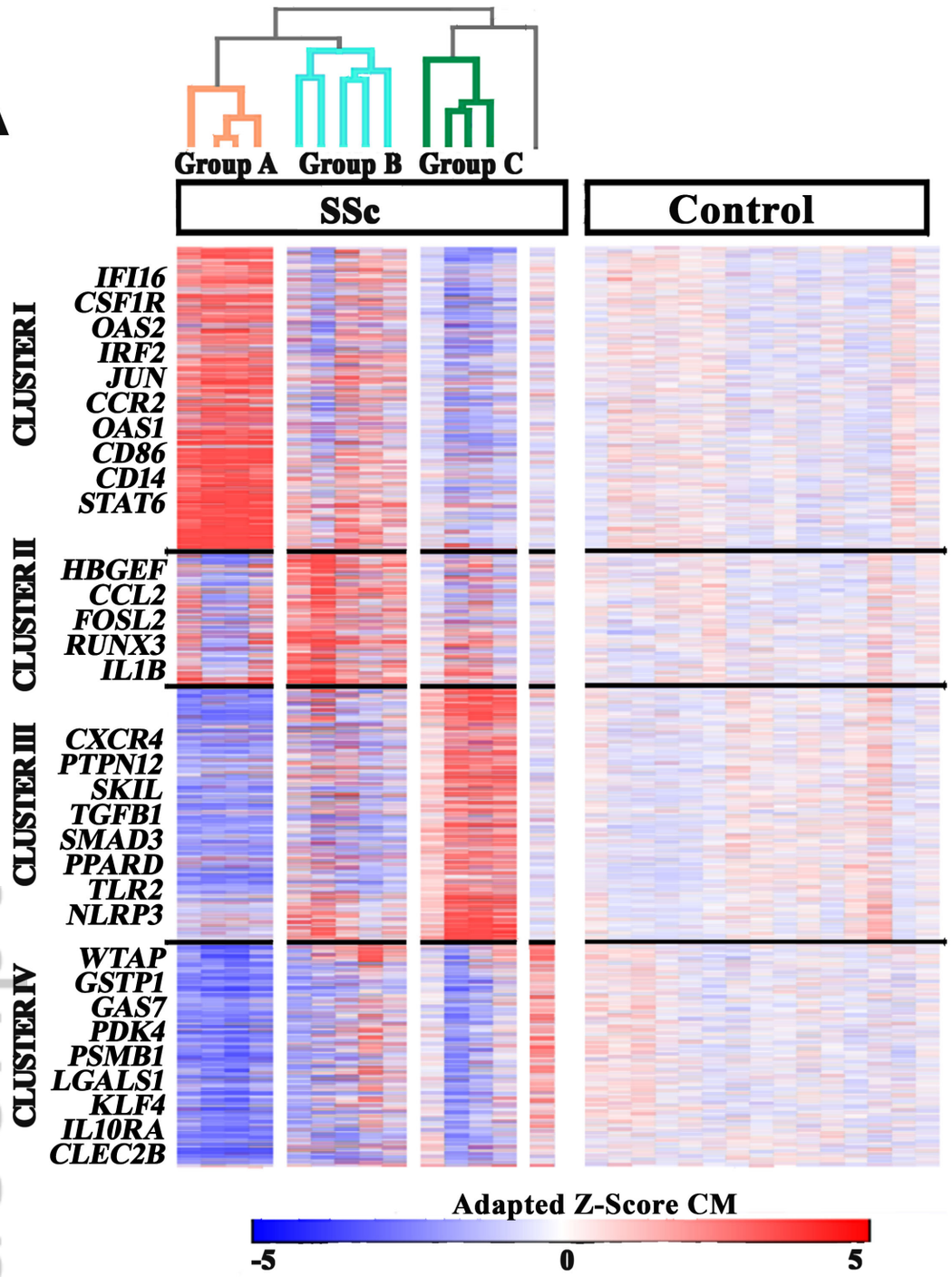
Supplemental Figure 4: CM and NCM retain their individual cell type signature. (A) Average expression level of all CM genes in NCM of patients and controls. **(B)** Average expression level of all NCM genes in CM of patients and controls. **(C)** Expression of representative NCM genes in patient CM samples by group. **(D)** Expression of representative CM genes in patient NCM samples by group. **E)** Percentage of CM genes with higher average CM expression in each patient group

than control average. (**F)** Percentage of NCM genes with higher average NCM expression in each patient group than control average. * Indicates $p < 0.05$, Fisher's exact test.

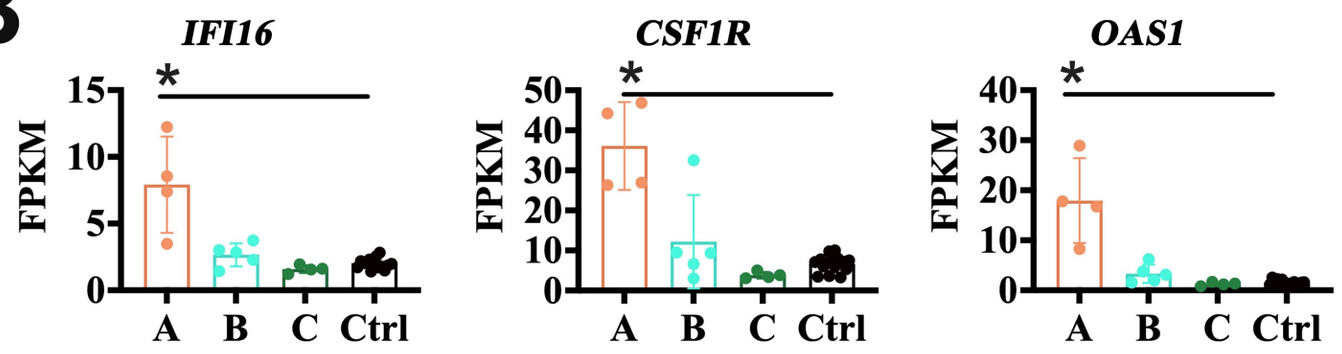
Supplemental Figure 5: **Gene expression in skin macrophages. (A)** FACS of enzymatic digestions (Dig. 1-4) performed on mammoplasty surgical discards. **(B)** Flow cytometric analyses of sorted biopsies obtained from HC and SSc patients. **(C)** Numbers of macrophages (46-700, avg 240) FACSorted from two 4mm punch biopsies for HC vs. SSc patients (healthy n=4, SSc n=5). Dermal macrophages are denoted by $CD45^+CD14^+HLADR^+CD11b^+CD163^+CD206^+$. **(D)** Histogram indicating the number of genes expressed at the given Log_2(FPKM) in each skin macrophage sample. Red line indicates the threshold for 6339 expressed genes (FPKM = 3) in at least 3 samples. **(E)** 5'→3' gene coverage in each patient and control sample. **(F)** PCA of skin macrophages samples annotated by SSc or control. **(G)** Bar graph showing the enrichment of SSc-specific vs. Control-specific genes in the overlap with CM Clusters compared to the expected ratio (190/138); dotted line).

Supplemental Figure 6: Heatmap of normalized expression of genes per patient from CM Clusters I-IV in bulk monocytes from SSc patients categorized by disease phenotype (dataset described in van der Kroef, et al).

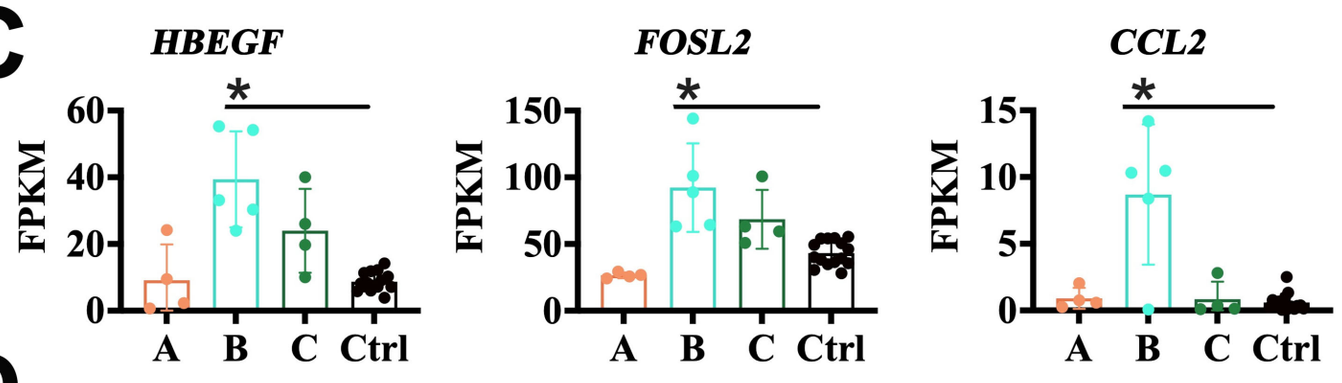
A



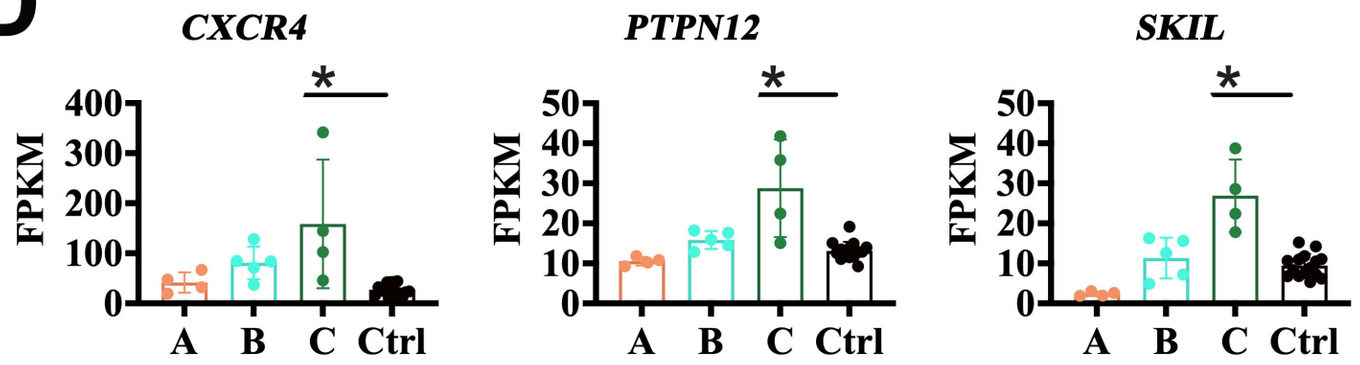
B



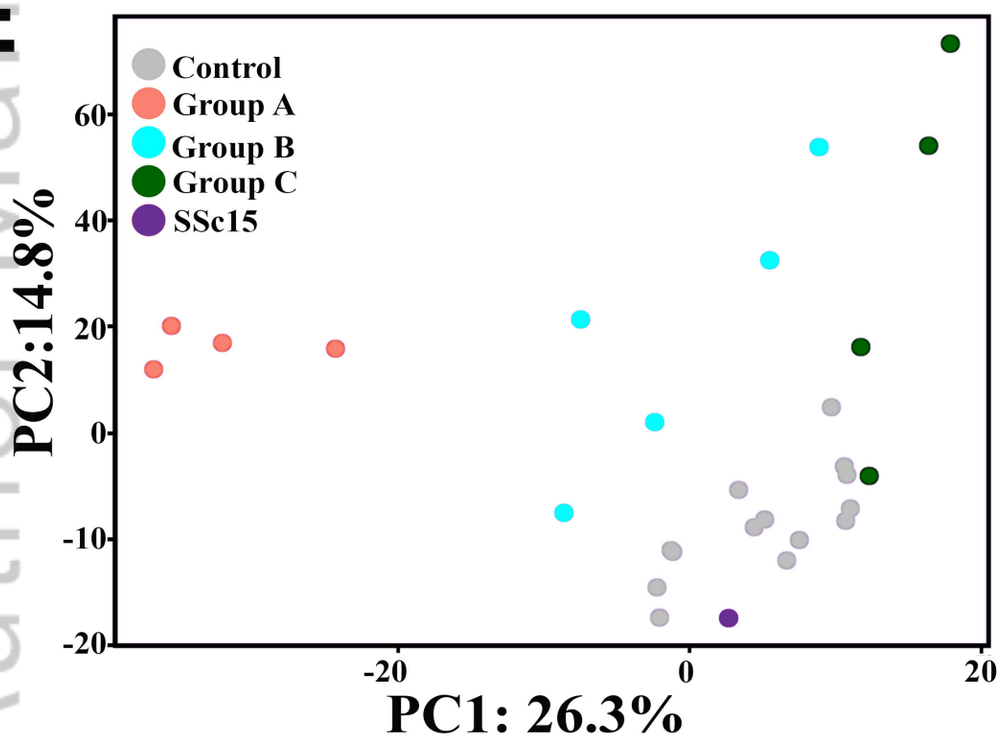
C



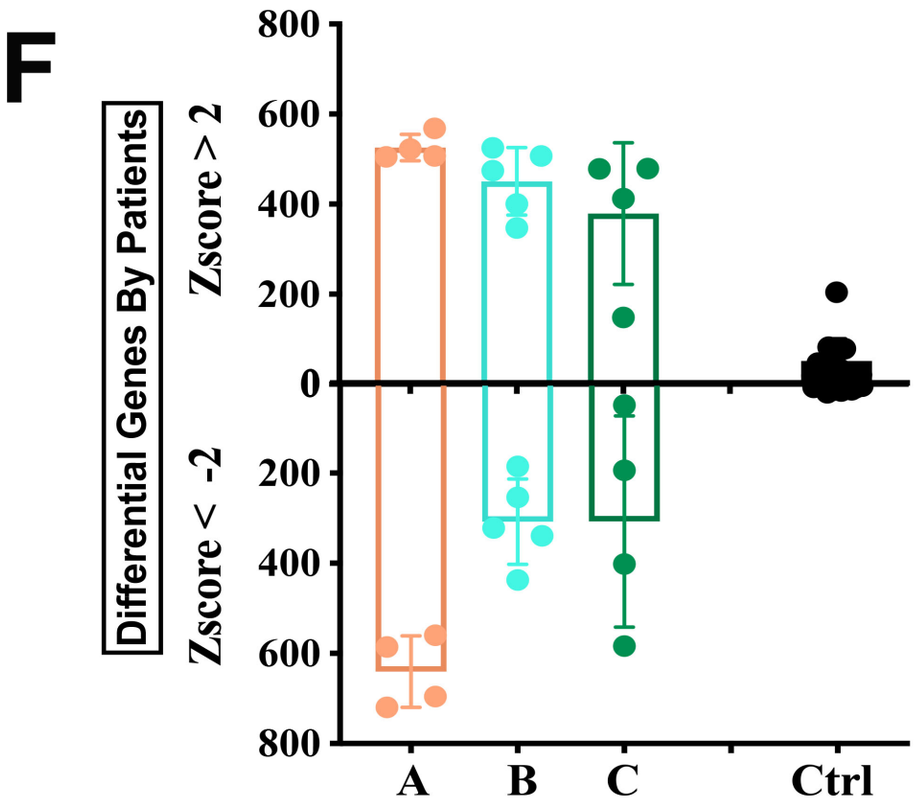
D



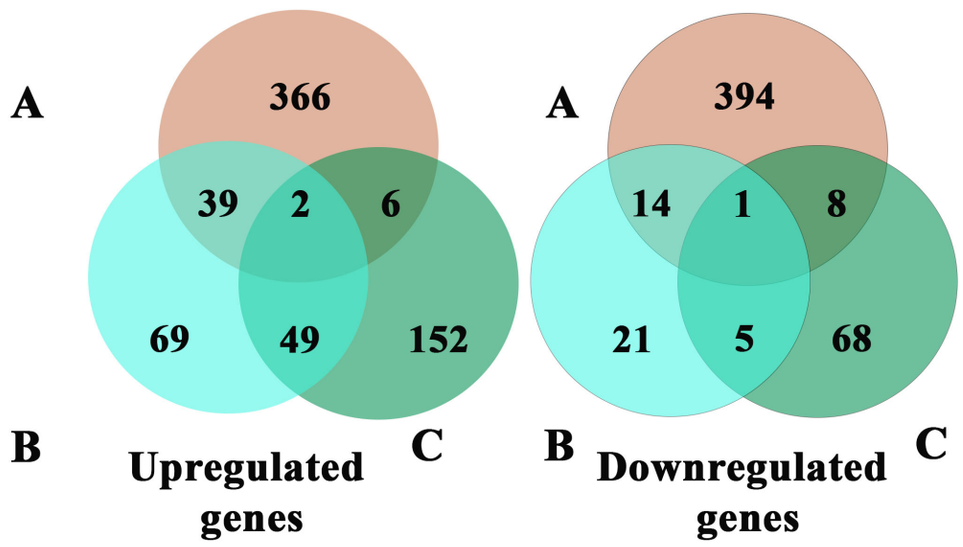
E



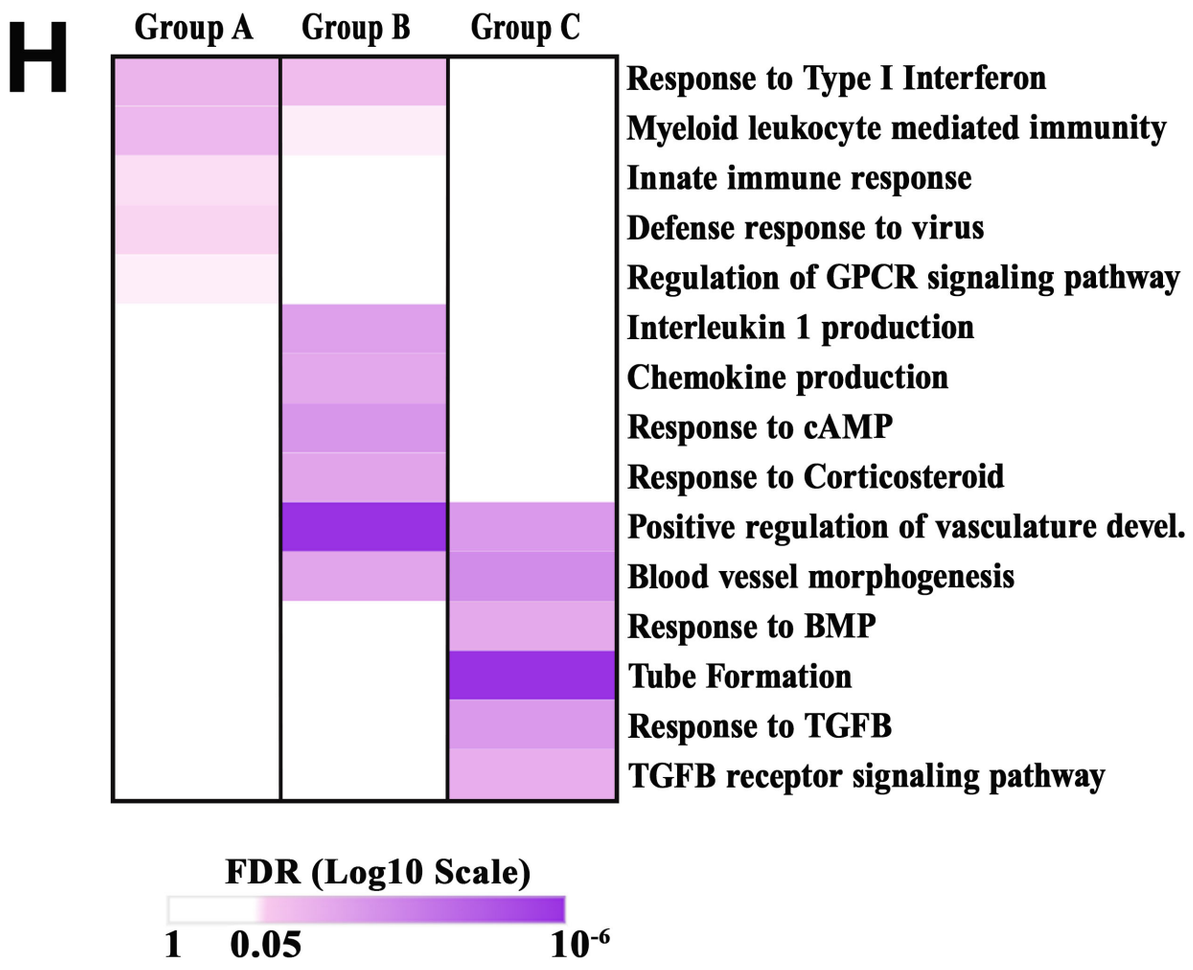
F

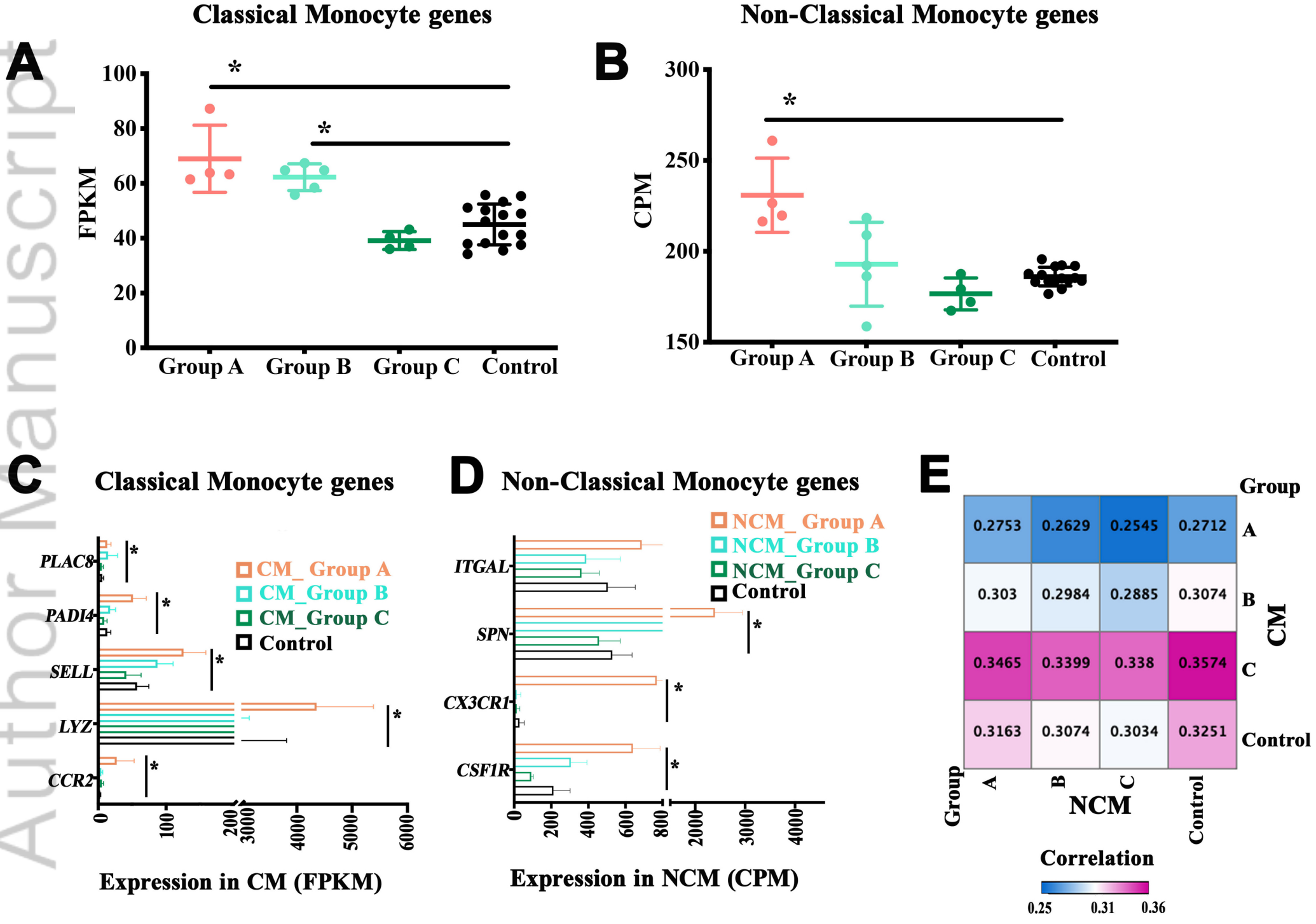


G



H

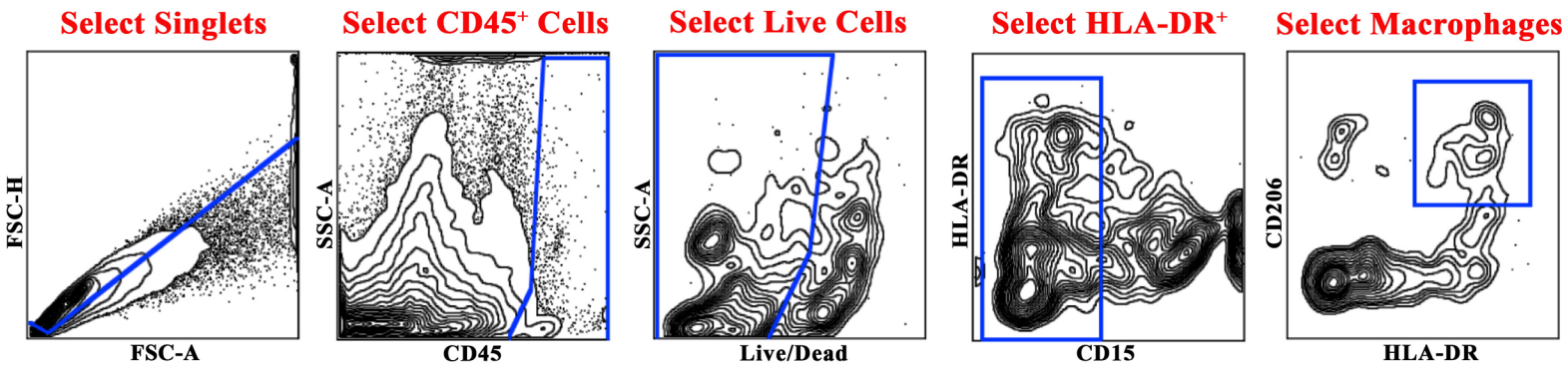




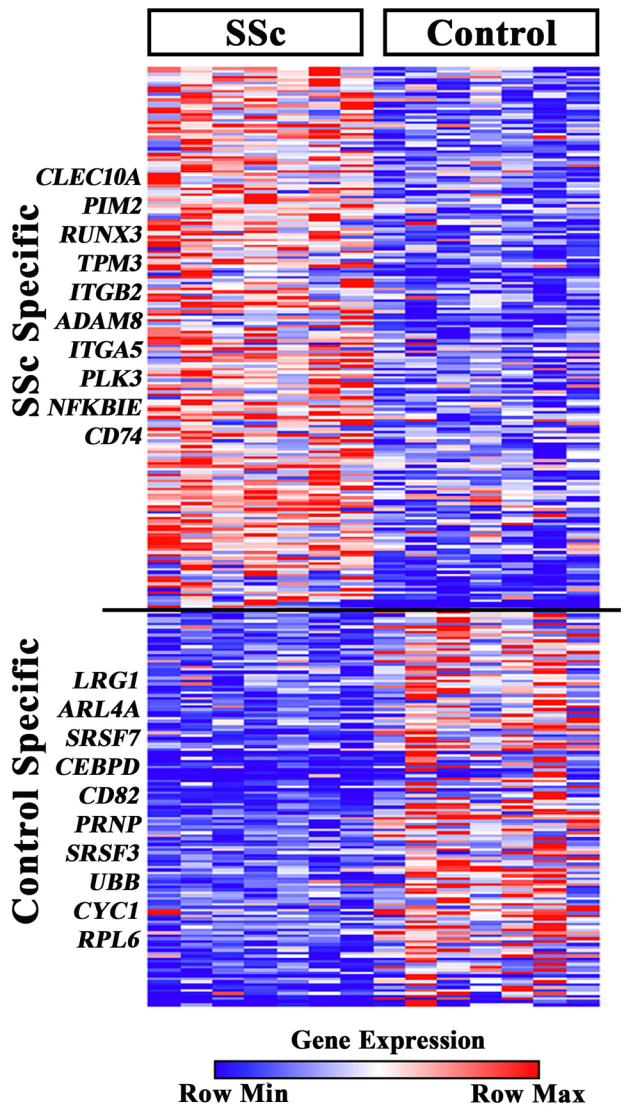
A

Characteristic	Sex, n (% women)	Race, n (% white)	Age, mean years (SD)	mRSS, mean (SD)	SSc disease duration, mean months (SD)	Scl-70, n (% positive)	RNA Polymerase III, n (% positive)	Current immune suppression, n (%)	Prior treatment, n (%)
SSc patients (n=7)	5 (71)	6 (86)	55 (13)	14 (6)	87 (104)	NA	NA	4 (57)	3 (48)
Control (n=7)	5 (71)	6 (86)	45 (15)	NA	NA	NA	NA	NA	NA
p - val2	ns	ns	ns	NA	NA	NA	NA	NA	NA

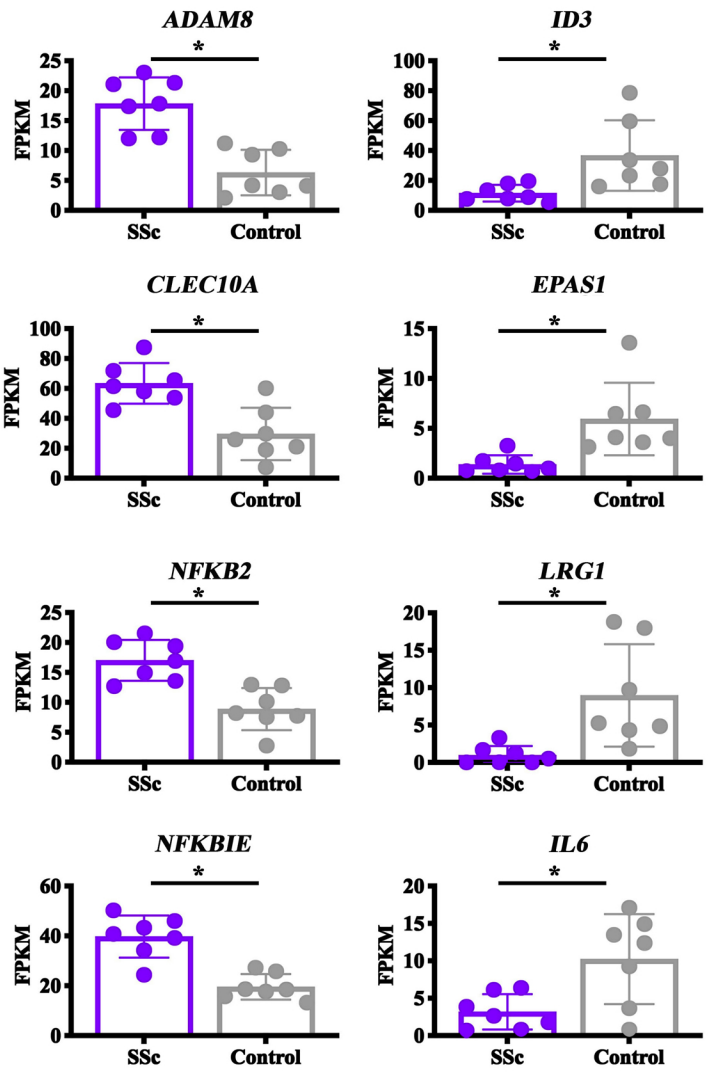
B



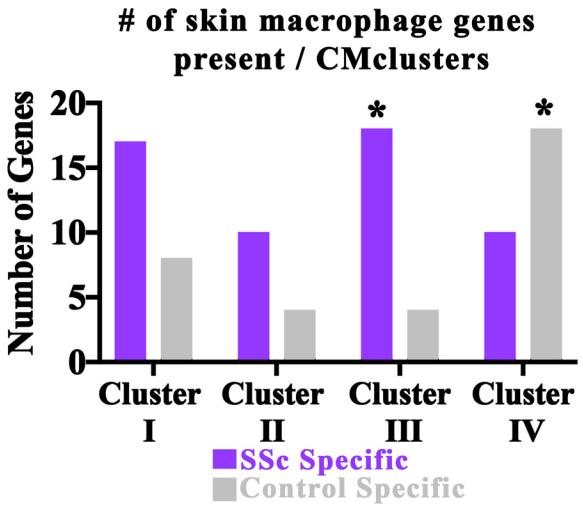
C



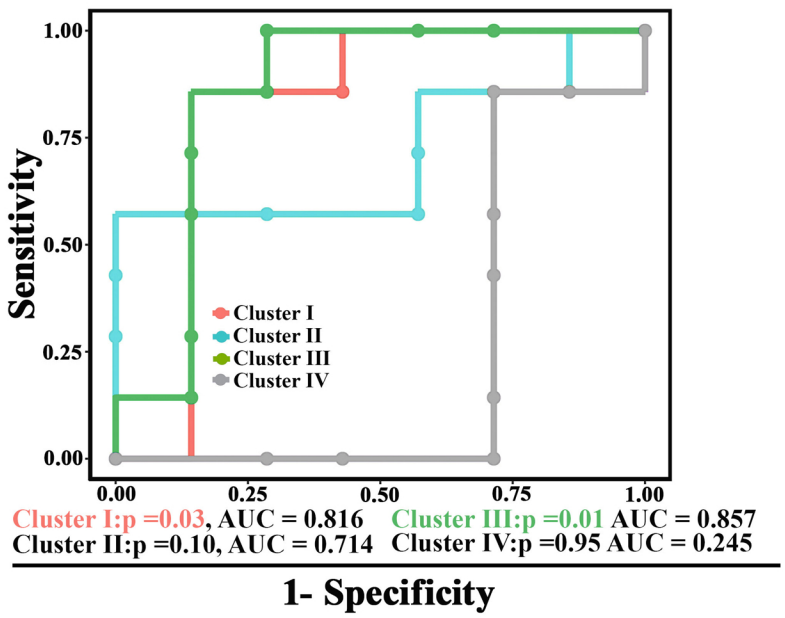
D

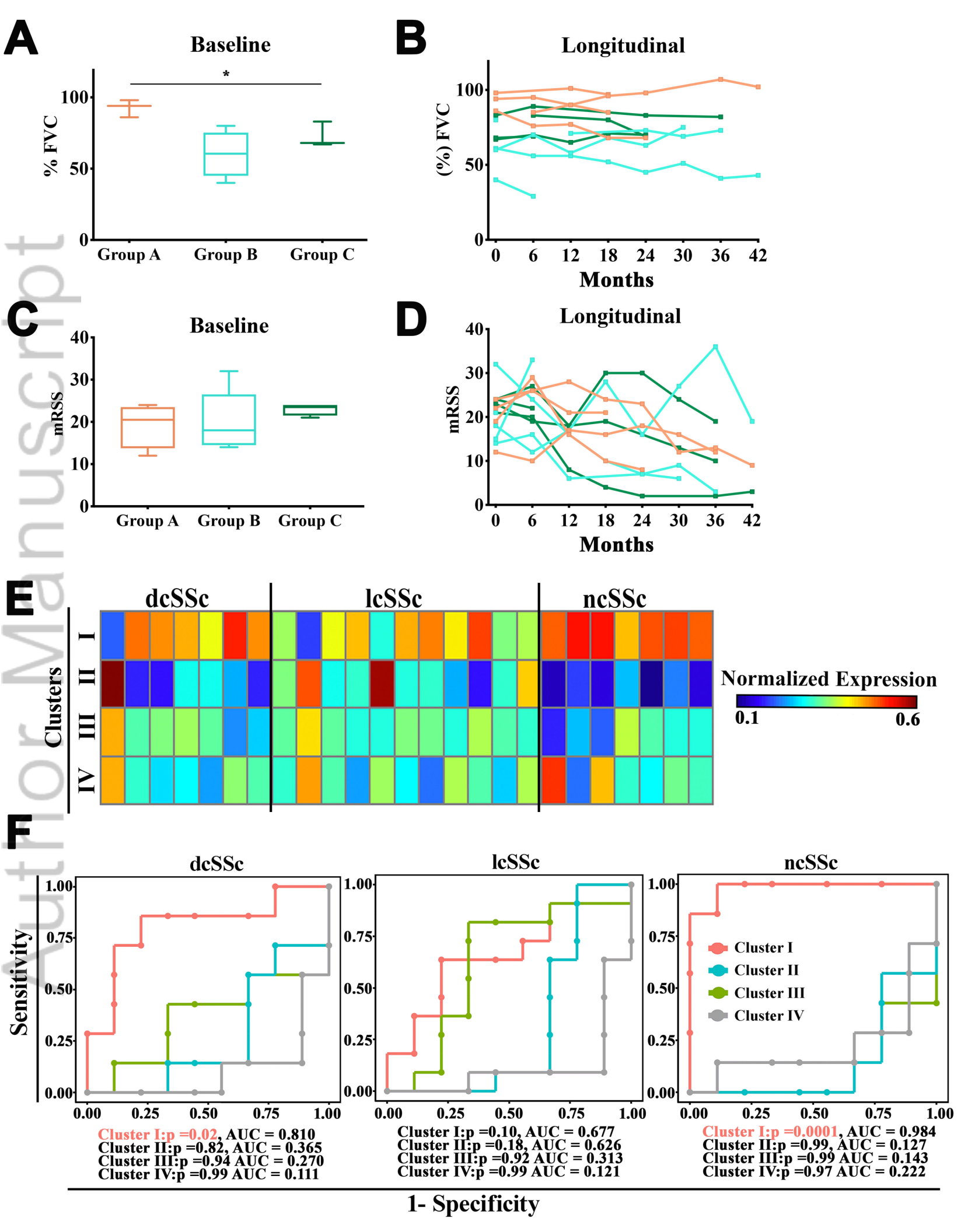


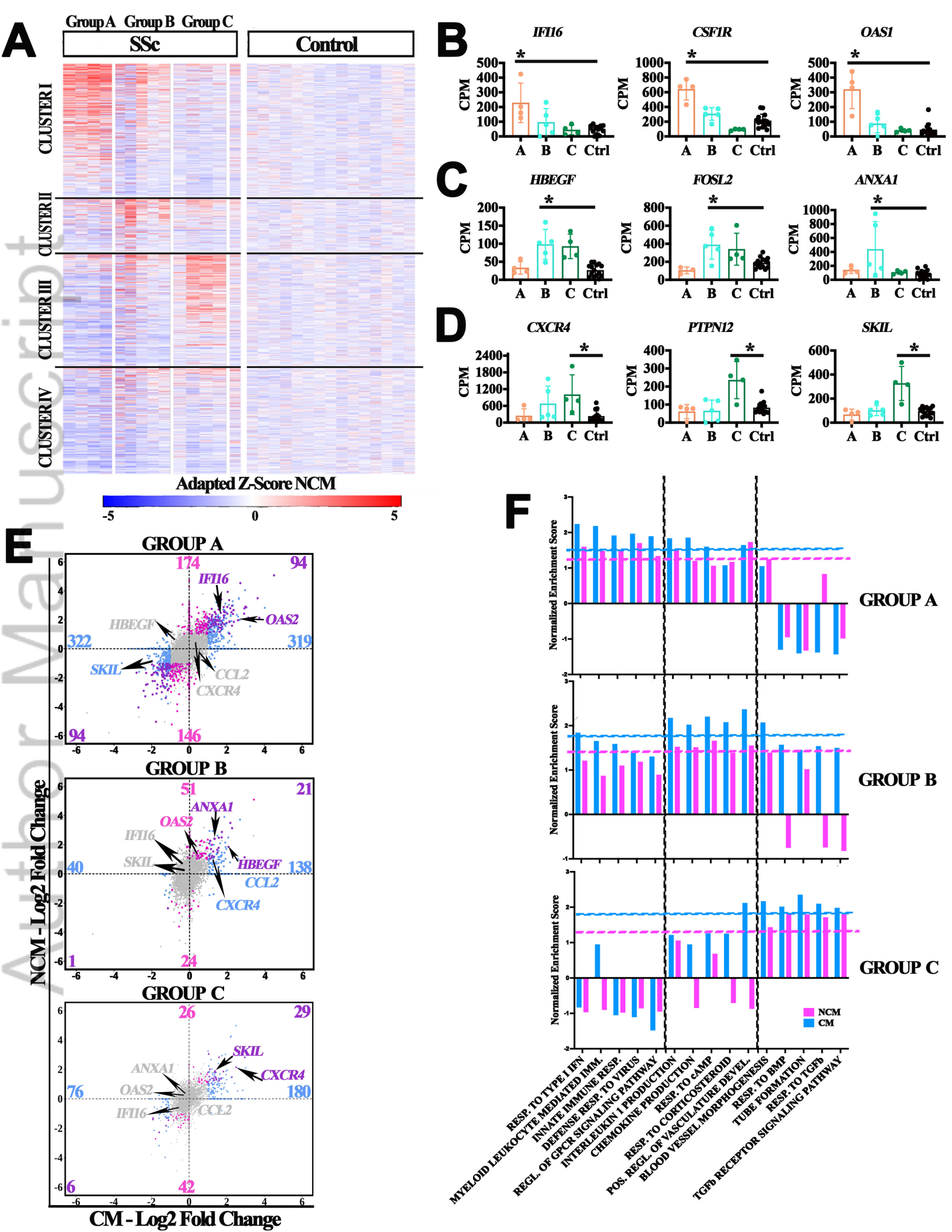
E



F







A

Characteristic	Sex, n (% women)	Race, n (% white)	Age, mean years (SD)	mRSS, mean (SD)	SSc disease duration, mean months (SD)	Scl-70, n (% positive)	RNA Polymerase III, n (% positive)	Current immune suppression, n (%)
SSc patients (n=14)	10 (71)	8 (57)	53 (13)	18 (5)	15 (6)	3 (21)	3 (21)	7 (50)
Control (n=15)	11 (73)	10 (67)	51 (13)	NA	NA	NA	NA	NA
p - val	ns	ns	ns	NA	NA	NA	NA	NA