Single-Cell RNA-Sequencing Analyses Identify APLNR, INS-IGF2, RGCC Genes May Be Involved in the Pathogenesis of Systemic Sclerosis Skin

Peiqiu Zhu, Weiwei Deng

1Department of Dermatology and Venereology, Beijing Jishuitan Hospital, Capital Medical University, Beijing, People’s Republic of China; 2Department of Dermatology, Dermatology Hospital of Southern Medical University, Guangzhou, People’s Republic of China

Correspondence: Weiwei Deng, Department of Dermatology, Dermatology Hospital of Southern Medical University, Guangzhou, People’s Republic of China, Tel +86 15622165048, Email 1601178683@qq.com

Background: Systemic sclerosis represents a persistent autoimmune disorder marked with fibrosis affecting both skin and other organs, which leads to a diminished quality of life and increased mortality. The affected skin provides a valuable opportunity to explore the pathogenesis of systemic sclerosis. Nevertheless, the roles of various cell populations within scleroderma remain intricate.

Methods: We conducted a comprehensive reanalysis of recently published single-cell RNA-sequencing data from skin tissue cells in scleroderma. Through the utilization of Seurat, irGSEA, AUCell packages, and WGCNA analysis, we aimed to unveil crucial genes associated with the disease’s etiological factors. Our investigation involved the characterization of heterogeneous pathway activities in both healthy and SSc-affected skin. Furthermore, we employed immunofluorescence techniques to validate the expression patterns of hub genes and differentially expressed genes.

Results: The Endothelial-to-Mesenchymal Transition (EndMT) pathway was upregulated in SSc skin. Notably, the M4 module within Endothelial cell subpopulation 1 exhibited a strong association with EndMT. Furthermore, we identified three overexpressed genes (APLNR, INS-IGF2, RGCC) that demonstrated a significant correlation with EndMT. Importantly, their expression levels were markedly higher in skin of individuals with SSc when compared to healthy controls.

Conclusion: APLNR, INS-IGF2 and RGCC serve as potential key players in the pathogenesis of SSC skin through EndMT-dependent mechanisms.

Keywords: systemic sclerosis skin, single cell RNA-sequencing analysis, the Endothelial to Mesenchymal Transition, EndMT, APLNR, INS-IGF2, RGCC

Introduction
As a sort of uncommon connective tissue disorder, scleroderma is distinguished by cutaneous sclerosis and variable systemic involvement. There are 2 main forms of systemic sclerosis(SSc): localized scleroderma (LS), also known as limited cutaneous SSc, and diffuse SSc. It is currently believed that patients who develop SSc have an underlying genetic predisposition to these conditions and are subsequently exposed to environmental factors that initiate inflammatory and fibrotic processes. The common clinical manifestations of SSc encompass skin fibrosis, digital vascular ulcers, arrhythmias in involvement with the cardiovascular system, pulmonary fibrosis, pulmonary artery hypertension, fibrotic changes in the gastrointestinal tract, endothelial cell injury in the kidneys, and even hand (finger) flexion contractures. Evaluation of sclerotic skin reveals thickened and homogenized collagen bundles in both superficial and middle dermis. Qualitative SSc nailfold capillary number has been developed to accurately determine capillary absolute number and density from nail-fold video capillarycopy images, which has been successfully validated in patients diagnosed with systemic sclerosis (SSc). Etiological factors trigger the activation of keratinocytes, T cells, and the production of various components that contribute to fibrosis. These factors induce fibroblast proliferation through toll-like receptor 4 (TLR-4) by enhancing the activity of transforming growth factor β (TGF-β). A preliminary study, albeit with a small number of
cells and samples, suggested that inhibiting TGF-β1 signaling might be one of the most promising approaches to treat skin fibrosis in scleroderma. In recent years, while the role of the TGF-β family in the pathogenesis of LS has been confirmed, the signal pathways and molecular process involved are too complex to fully elucidate. Consequently, investigating the molecular mechanisms within different types of cells, cytokines, and extracellular matrix (ECM) components can help uncover what lies beneath the surface.

Several studies have provided evidence supporting the involvement of endothelial cells (EC) in endothelial-to-mesenchymal transition (EndMT), a recently acknowledged form of cellular trans-differentiation. Involved in crucial embryonic development processes, EndMT assumes a significant role in the development of various genetically determined and acquired human diseases, including malignancies, vascular disorders, inflammatory conditions, and fibrotic disorders. During EndMT, ECs undergo phenotypical changes, losing their vascular EC markers and acquiring mesenchymal cell markers. Several studies have pinpointed myofibroblasts as the key cells instigating and progressing the fibrotic process. Prior research has explored various factors, such as the TGF-β pathway, Notch Signaling Pathway, Wnt Pathway, ET-1, NOX4, Oxidative stress, and MicroRNAs, to elucidate the molecular changes and regulatory events involved in EndMT. A newly cross-sectional study from Chiu et al suggested that EndMT and fibroblast senescence were more prevalent in SSc skin tissue. However, the precise mechanisms that dominate the EndMT process and its connection to SSc remain largely unclear.

In this study, we conducted a comprehensive reanalysis of single-cell data derived from SSc skin, building upon a prior investigation. Our goal was to unveil the diverse pathway activities and identify genes implicated in the pathogenesis of EndMT. Of utmost significance, we pinpointed the specific cell populations most intricately linked to the pathogenesis of SSc skin. Additionally, employing a weighted gene co-expression network analysis (WGCNA) algorithm, we meticulously screened genes closely associated with EndMT and their relevance to SSc skin. These findings were further confirmed through immunofluorescence, confirming that APLNR, INS-IGF2, and RGCC are potential contributors to the pathogenesis of SSc skin via the EndMT mechanism.

Materials and Methods

scRNA-Seq and RNA-Seq Datasets

The scRNA-seq data were downloaded from the GEO database (GSE195452).

Ethics Statement

All experimental procedures and protocols used in this research underwent thorough review and received approval from the committee of Dermatology Hospital of Southern Medical University (2023173). The skin donors were provided with informed consent in accordance with the Declaration of Helsinki.

Cell Subpopulation Identification

Seurat is a powerful R package widely used for the analysis of single-cell RNA-seq (scRNA-seq) data. The analysis workflow typically involves several key steps. First, the data undergoes rigorous quality control (QC) to identify and remove low-quality cells or data points. Following QC, the dataset is subjected to standardization and normalization procedures to account for variations in sequencing depth and library size. Subsequently, the data is used for the identification of single-cell subpopulations through dimensionality reduction techniques such as UMAP (Uniform Manifold Approximation and Projection) and clustering methods. This enables the grouping of cells into distinct clusters representing different cell types or states. In summary, Seurat provides a robust framework for preprocessing scRNA-seq data, identifying cell subpopulations, and facilitating downstream analyses. The statistical methods in single-cell analysis are all based on widely adopted approaches reported in the Seurat package. For example, the commonly used Wilcoxon rank-sum test method in differential gene expression analysis.
Pathway Activity Analysis
We conducted an analysis of cellular subpopulation pathway activity utilizing the irGSEA and AUCell packages, making use of the default algorithms provided by these R packages.

WGCNA Analysis
Using the scWGCNA package, we performed Weighted Gene Co-expression Network Analysis (WGCNA). Initially, we computed the soft power value using the default algorithm and subsequently proceeded to identify gene expression modules. Ultimately, we calculated the correlations between these modules and the phenotypic traits of interest.

Immunofluorescence
Sliced into 5μm thick section, formalin-fixed paraffin-embedded SSc skin tissues were deparaffinized and rehydrated. Antigens were obtained by heating and induction in pH 6.0 citrate buffer (Vector Labs, Burlingame, CA) for 10 minutes and then cooling for 10 minutes. Blocking was performed with 3% H2O2, followed by 10% goat serum (ThermoFisher Scientific, Waltham, MA) for 1 hour. The following primary antibodies were used for double antibody staining: 702069 Apelin receptor recombinant rabbit monoclonal antibody (SH5L9), MA5-17096 IGF2 monoclonal antibody 8H1, 14-9760-82 α-smooth muscle actin monoclonal antibody 1A4 and PA5-143217 CD31 polyclonal antibody. The primary antibodies were subjected to an overnight incubation at 4°C. After each step, the tissue underwent three PBS washes for 10 minutes each. Enzymatic reaction was performed for 1 hour using appropriate goat anti-rabbit or goat anti-mouse Poly-HRP-conjugated secondary antibodies. Alexa Fluor 488 or 594 labeled amide solutions were then used and completed with reaction stop solution (ThermoFisher Scientific, Waltham, MA). Samples stained with dual antibodies underwent a similar staining procedure; after the initial antibody development, the tissue went through another round of heat-induced antigen retrieval, blocking, incubation with a compatible primary antibody, multi-HRP secondary antibody, and Alexa Fluor-conjugated tyramide spectroscopy. The slides were subjected to nuclear staining using Hoechst stain (1:2000).

Finally, the slides were cover-slipped with ProLong Diamond, and Olympus FLUOVIEW FV1000 confocal laser-scanning microscope was used to capture images.

Results
Identify Pivotal Endothelial Cell Subpopulations and Pathways in SSc Skin
A total of 44 SSc skin samples and 57 healthy control samples with single-cell data, obtained from a previous study, were included in our analysis for the isolation of endothelial cells. Initially, we performed Uniform Manifold Approximation and Projection (UMAP) to cluster these cells based on the gene expression levels (Figure 1A and B). The UMAP results revealed the presence of two distinct subpopulations of endothelial cells, which we referred to as Endothelial cell subpopulation 1 and 2 (Figure 1A). We further categorized these endothelial cells into 6 subpopulations (Figure 1B). To gain insights into the functionality of different cell populations within SSc skin, we conducted pathway activity analysis for these subpopulations. A comparison of activation pathways between cell groups in normal tissues (The functional enrichment results of differentially expressed genes in each cell subpopulation relative to other cell subpopulations) revealed that EPITHELIAL-MESENCHYMAL-TRANSITION (P<0.001), ADIPOGENESIS (P<0.01), and ANGIOGENESIS (P<0.01) exhibited upregulation in the Endothelial 1 cluster (Figure 1C). Similarly, in SSc skin, EPITHELIAL-MESENCHYMAL-TRANSITION (P<0.0001) and ADIPOGENESIS (P<0.0001) (P<0.05) pathways displayed upregulation (Figure 1D). Our primary focus was on the EndMT pathway, where we observed that Endothelial 1 was associated with the EndMT direction under both normal and pathological conditions.

However, the metabolic pathway of EndMT, adipogenesis, and Interferon-α were significantly upregulated in the cluster of Endothelial 1 (E1) in SSc skin (Figure 1C and D). It’s worth noting that adipocytes are known to promote fibroblast collagen synthesis under normal conditions, while inhibiting it in fibrotic conditions. Previous studies have also implicated the activation of Type 1 IFN in the pathogenesis of SSc. In contrast, the Endothelial 1 (E1) subpopulation accounted for a higher proportion in the SSc cell population (Figure 1E), suggesting a more significant involvement in EndMT within the diseased tissue.
M4 Module Was Correlated with the EndMT Pathway

To identify key genes associated with the EndMT pathway within the E1 group, we employed Weighted Gene Co-expression Network Analysis (WGCNA). The soft power for the WGCNA analysis was determined to be 8 (Figure 2A). The E1 hdWGCNA Dendrogram illustrated the formation of 4 modules based on different gene functions, each represented by a distinct color (Figure 2B, Figure Supplementary 1). Among these 4 module clusters, M4 emerged as the most significant module in E1 and was closely related to the EndMT pathway (Figure 2C). Subsequently, 25 genes were identified as hub genes within the M4 module (Figure 2D). In our analysis using WGCNA, APLNR stood out as the most upregulated among the hub genes in E1 (Figure 3). This observation suggests that APLNR may play a specific role within the E1 subgroup. Notably, the Apelin receptor APLNR has previously been validated as a marker of EC injury in SSc skin, as identified by single-cell RNA sequencing.23

APLNR, IGF2 and RGCC Could Be the Genes with Most Relevant for the EndMT Pathway in SSc Skin

We compiled a list of the top 10 Differentially Expressed Genes (DEGs) in E1 between SSc skin and normal control, which is presented in Table 1. To explore the relationship between the APLNR gene and these top 10 DEGs in SSc skin...
with the EndMT pathways, we conducted Spearman correlation analysis. The results indicated that APLNR, IGF2, and RGCC exhibited the strongest correlations with the EndMT pathways, with correlation coefficients greater than 0.5 (Figure 4A–C). In contrast, the Spearman correlation coefficients for the other DEGs, including MIR675, CD36, FABP4, CD300LG, RBP7, AC132217.1, BTNL9, and PODXL, were below 0.5, as shown in Figure Supplementary 2. Of particular interest, RGCC has been recently identified as a novel profibrotic mediator. It possesses the ability to induce EndMT in colorectal cancer cells by activating the Smad/Sip1 signaling pathway. This suggests a potential role for RGCC in the context of EndMT in SSc skin.

**APLNR, IGF2 and RGCC Were Verified Highly Expressed in SSc Skin by Immunofluorescence Co-Localization**

Alpha-smooth muscle actin (α-SMA) serves as a specific marker expressed by endothelial cells that have transitioned into a mesenchymal phenotype during EndMT. CD31 is a well-established biomarker for human endothelial cells. In our study, we conducted immunofluorescence staining to co-localize APLNR, IGF2, and RGCC with α-SMA and CD31. This co-staining confirmed that these genes are indeed expressed by endothelial cells within the skin tissue undergoing EndMT. Our findings revealed that APLNR exhibited higher expression in SSc skin when compared to normal control...
skin (Figure 4D). Similarly, IGF2 and RGCC demonstrated elevated expression levels in SSc skin (Figure 4E and F). These results suggest a potential role for APLNR, IGF2, and RGCC in the position of EC transition to a mesenchymal phenotype during EndMT in SSc skin.

**Table 1** Top 10 Differentially Expressed Genes (DEGs) in Endothelial Subpopulation 1

<table>
<thead>
<tr>
<th>Gene</th>
<th>p_val</th>
<th>avg_log2FC</th>
<th>pct.1</th>
<th>pct.2</th>
<th>p_val_adj</th>
<th>Celltype</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIR675</td>
<td>0</td>
<td>3.91320987</td>
<td>0.469</td>
<td>0.036</td>
<td>0</td>
<td>Endothelial I</td>
</tr>
<tr>
<td>CD36</td>
<td>0</td>
<td>3.0184988</td>
<td>0.667</td>
<td>0.153</td>
<td>0</td>
<td>Endothelial I</td>
</tr>
<tr>
<td>FABP4</td>
<td>0</td>
<td>2.83665924</td>
<td>0.779</td>
<td>0.394</td>
<td>0</td>
<td>Endothelial I</td>
</tr>
<tr>
<td>CD300LG</td>
<td>0</td>
<td>2.65881267</td>
<td>0.537</td>
<td>0.052</td>
<td>0</td>
<td>Endothelial I</td>
</tr>
<tr>
<td>RBP7</td>
<td>0</td>
<td>2.64643076</td>
<td>0.684</td>
<td>0.185</td>
<td>0</td>
<td>Endothelial I</td>
</tr>
<tr>
<td>AC132217.1</td>
<td>1.11E-271</td>
<td>2.42065284</td>
<td>0.593</td>
<td>0.208</td>
<td>6.41E-267</td>
<td>Endothelial I</td>
</tr>
<tr>
<td>BTLN9</td>
<td>0</td>
<td>2.23639931</td>
<td>0.331</td>
<td>0.015</td>
<td>0</td>
<td>Endothelial I</td>
</tr>
<tr>
<td>INS-IGF2</td>
<td>3.73E-237</td>
<td>2.17469105</td>
<td>0.394</td>
<td>0.078</td>
<td>2.16E-232</td>
<td>Endothelial I</td>
</tr>
<tr>
<td>RGCC</td>
<td>6.01E-281</td>
<td>2.01376555</td>
<td>0.372</td>
<td>0.046</td>
<td>3.48E-276</td>
<td>Endothelial I</td>
</tr>
<tr>
<td>PODXL</td>
<td>0</td>
<td>1.98927416</td>
<td>0.49</td>
<td>0.089</td>
<td>0</td>
<td>Endothelial I</td>
</tr>
</tbody>
</table>
Figure 4 (A–C) Spearman correlation analysis demonstrating the relationship of APLNR, IGF2, and RGCC with the EndMT pathways, with a correlation coefficient > 0.5. (D–F) Immunofluorescent staining of formalin-fixed paraffin-embedded biopsies of SSc and control skin tissues. Merged staining with α-SMA and CD31 showed higher expression of APLNR, IGF2, and RGCC in SSc skin. The scale bar is 20 µm.
Discussion

The EndMT plays a central role in the progression of SSc skin disease. Our objective is to elucidate and characterize the transcriptomes of endothelial cell populations present in both healthy human skin and skin affected by SSc, with the aim of comprehensively understanding the mechanisms by which alterations in endothelial cell phenotypes contribute to the pathological process of EndMT. Previous investigations primarily emphasized lymphocytes, fibroblasts, and inflammatory pathways in Systemic Sclerosis (SSc), while relatively neglecting the interplay between endothelial cells and Endothelial-to-Mesenchymal Transition (EndMT). A prior sc-RNA seq study found TGFβ response predominantly in fibroblasts and smooth muscle cells in early anti-topoisomerase-1+SSc skin samples, whereas in early anti-RNA polymerase III+SSc patient skin samples, endothelial cells were the responders.25 Another study examining potential disease-associated fibroblasts and related genetic signatures in SSc skin revealed a prevalent myofibroblast-like cluster (SFRP4/PRSS23) that shared numerous upregulated genes expressed in SSc-associated myofibroblasts.26 Gaydosik et al identified a distinct cluster of recirculating CXCL13+ T cells in SSc skin, expressing a T helper follicular-like gene expression signature, that poised to promote B-cell responses within inflamed tissue.27 Here, this study sheds light on the potential involvement of APLNR, INS-IGF2, and RGCC in the pathogenesis of SSc skin through the process of EndMT (Figure Supplementary 3). We began by annotating six distinct endothelial cell subpopulations in SSc skin and identified the E1 cluster as particularly relevant to EndMT. Subsequently, we further categorized E1 into four modules based on different functional clusters. Employing WGCNA analysis, we pinpointed the genes most closely associated with EndMT within the M4 module. Among the 25 hub genes within this module, APLNR emerged as the gene with the strongest links to EndMT. To corroborate our findings, we conducted a comparative analysis of the top 10 DEGs between healthy and SSc skin. Spearman correlation analysis reinforced the importance of APLNR, INS-IGF2, and RGCC in the context of EndMT in SSc skin, with these genes demonstrating a correlation coefficient of more than 0.5. Furthermore, our immunofluorescence results validated these bioinformatics findings, confirming the elevated expression of APLNR, INS-IGF2, and RGCC in SSc skin. These findings collectively stress the potential weight of these genes in the pathogenesis of SSc skin via the process of EndMT.

The APLNR gene, namely APJ, is situated on chromosome 11q12 and stands out as an intriguing member of the G protein-coupled receptor family, classified as an orphan receptor in group A.28 As the ligand for the APJ receptor, Apelin was originally isolated from bovine stomach extracts.29 The Apelin/APJ system is diffusely distributed throughout the human body, with a particularly robust presence on endothelial cells (ECs).30 This system has been shown to promote EC proliferation, migration, and angiogenesis.31,32 Notably, APLNR was identified as one of the top markers of EC injury in SSc through single-cell RNA sequencing of patients.23 Furthermore, using human embryonic stem cells (hESCs), Slukvin et al demonstrated that mesodermal mesenchymal stem cells originate from APLNR+ precursors with angiogenic potential, known as mesenchymoangioblasts.33 Another study investigating pulmonary veno-occlusive disease, characterized by fibrosclerosis in venules and small veins, revealed that Erg inhibits the proliferation of pulmonary venous endothelial cells through the APLNR pathway.34 This discovery suggests that the ERG-APLNR axis may play a critical role in maintaining pulmonary venous endothelial homeostasis. Interestingly, while APLNR has not previously been associated with EndMT within the etiology of SSc skin, our experiments suggest that it maybe involved in this process. This leads to a twofold conjecture. Firstly, it’s possible that the high expression of APLNR during EndMT is linked to endothelial cell injury. Secondly, APLNR+ cells may possess a greater propensity to undergo EndMT during the pathogenesis of SSc skin. To validate these hypotheses, further functional tests involving knockout, downregulation, or overexpression experiments are warranted.

IGF2, located in the 11p15 region of the human genome, represents an imprinted gene with exclusive expression from the paternal allele.35 This gene plays a pivotal role in regulating growth, especially during normal fetal development, and IGF-I and IGF-II serve as the primary ligands for the type 1 IGF receptor (IGF-1R).36 The British Society for Rheumatology (BSR)/British Health Professionals in Rheumatology (BHPR) recommend that all cases of SSc undergo evaluation for lung fibrosis. Treatment decisions should be based on considerations such as the extent, severity, and likelihood of progression to severe disease.37 Importantly, elevated regional IGF-II expression has been observed in pulmonary fibrosis associated with SSc, in vitro experiments and in vivo studies, and IGF-II has been shown to induce
the production of extracellular matrix (ECM) through phosphatidylinositol-3 kinase- and Jun N-terminal kinase-dependent pathways.\textsuperscript{38} In the context of pulmonary fibrosis, IGF-II has also been associated with the upregulation of SOX9 in both human lung tissues and fibroblasts.\textsuperscript{39} Subsequently, SOX9, serves as a mediator to regulate the influence of IGF-II on TGF-\(\beta\)2 and TGF-\(\beta\)3, COL3A1, and prolyl 4-hydroxylase subunit alpha 2 in lung fibroblasts.\textsuperscript{39} In light of these findings, our data provide a fresh perspective on the role of IGF2 in cutaneous sclerosis, suggesting that IGF-II-induced ECM and TGF-\(\beta\) production may promote EndMT in localized scleroderma.

\textit{RGCC}\textsuperscript{30} was initially cloned from rat oligodendrocytes as part of a screen for genes responsive to complement activation.\textsuperscript{40} This multifaceted gene is involved in a range of physiological and pathological processes, including cell proliferation, differentiation, fibrosis, metabolic diseases, and cancer.\textsuperscript{41} Functionally, \textit{RGCC} serves as a downstream target of TGF-\(\beta\) and has been shown to mediate TGF-\(\beta\)-induced fibroblast activation by interacting with Smad3. This interaction activates the transcription programs of myofibroblast marker genes in human renal proximal tubular cells.\textsuperscript{42,43} Intriguingly, it has been proposed from a recent study that \textit{RGCC} exerts a protective function in pulmonary fibrosis, with its downregulation leading to collagen accumulation.\textsuperscript{44} \textit{RGCC} overexpression delayed early TGF-\(\beta\)-induced Smad2/3 phosphorylation, increased the expression of total and phosphorylated antifibrotic mediator STAT1, and reduced the expression of the profibrotic mediator STAT3.\textsuperscript{44} \textit{RGCC}'s roles extend beyond endothelial cells and encompass immune cells as well. In T-lymphocytes, \textit{RGCC} serves as a negative cell cycle regulator.\textsuperscript{45} Moreover, a recent viewpoint has proposed the potential emergence of lymphatic EndMT in the SSc skin, suggesting that this morpho-functional cell transition may serve as a pathogenetic connection between peripheral lymphatic network dysfunction or rarefaction in the primary phase of the disease and the progression of dermal fibrosis.\textsuperscript{46}

The sc-RNA-seq data indicates a clear connection between \textit{RGCC} and EndMT in SSc skin, as evidenced by the Spearman analysis. This implies that \textit{RGCC} may make contribution to the EndMT process of SSc skin, potentially impacting endothelial cells and lymphocytes.

**Conclusion**

Leveraging sc-RNA-seq data, our study offers a thorough characterization of endothelial cell subpopulations in SSc skin. We elucidate the activation of the EndMT pathway in SSc skin, spotlighting pivotal genes such as \textit{APLNR}, \textit{IGF2}, and \textit{RGCC}. These findings unveil novel research targets for exploring molecular mechanisms and potential therapeutic strategies.

**Acknowledgments**

Yishan Chen, participating investigators;
Contributions: served as scientific advisors, critically reviewed the revised manuscript;
Affiliation: Department of Dermatology and Venereology, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, People’s Republic of China.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**

Zhu and Deng


