nature portfolio

Peer Review File

Uncovering a Fibroblast Differentiation-Based Keloid Classification by Integration of Single-cell and Bulk RNA Sequencing

Corresponding Author: Professor Shizhao Ji

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript proposes a novel keloid molecular typing grounded on the differentiation status of fibroblasts, utilizing both single-cell and bulk sequencing data. However, the results pertaining to fibroblast trajectories appear inconsistent with previous findings, thereby casting doubt on the identified DEFDRGs that form the basis of the proposed molecular typing. Furthermore, while the authors suggest that these DEFDRGs correlate with keloid clinical outcomes, key prognostic parameters such as recurrence in keloids are conspicuously absent, limiting the clinical applicability of this typing system.

Several critical concerns and queries arise:
1.Both single-cell sequencing and bulk sequencing belong to transcriptomics. Therefore, the term "multi-omics" is inappropriate here, as it typically denotes analyses that integrate at least two distinct omics datasets, such as transcriptomics and epigenomics.
2.The term "prognosis" in keloids predominantly refers to recurrence post-treatment, given that keloids do not metastasize or lead to mortality. The manuscript claims "Patients in the HFIC group had a worse prognosis," yet lacks data substantiating this prognosis.
3.In line 90-91, previous revious studies delineating the molecular classification of keloid pathogenesis should be incorporated into the introduction for comprehensive context.
4.The manuscript utilizes "IntegrateData" for single-cell RNA-seq data merging without batch effect correction, leading to evident batch effects in Figure 1B. Such effects also appear to influence fibroblast subtypes, as observed with sample KL1 composed solely of fibroblast 6 (Figure 2B).
5.The manuscript uses two distinct control tissues—normal scars for scRNA-seq and normal skin for bulk RNA-seq—yet intersects these comparisons without justification for such choice.
6.The top 5 DEGs in fibroblast subtype 4 are indicative of endothelial cells (SELE) and immune cells (CD74, HLA-DRA). This prompts an investigation into whether this cluster represents doublets, as no doublet detection or removal was conducted.
7.Fibroblast 2 expresses genes like POSTN, ASPN, and COMP, consistent with previous scRNA studies identifying this cluster as a major contributor to keloid formation and potentially terminal cells in keloids. However, characterizing fibroblast 2 as the origin of differentiation in the study contradicts prior findings (Liu et al. 2022).
8.While 25 DEFDRGs were identified with correlations to keloids or healthy individuals, these associations lack clinical relevance without considering recurrence—a pivotal factor absent from the analysis.
9.The manuscript's claim that "FIBKC was significantly correlated with patient disease diagnosis and diagnostic time" does not substantiate the conclusion that "cluster 1 in FIBKC had a better prognosis, while cluster 3 was the opposite."

Several minor corrections:
10.Line 63 should replace "topical steroid injection" with the correct description.
11.Line 76 should read "at a single-cell level".
12.Line 124 should be "single-cell RNA-sequencing was performed on samples".
13.The term "muscle fiber cell" is unconventional in skin scRNA-seq studies; "smooth muscle cells" or "mural cells" would be more common.
14.Consistency in color representation of different fibroblast subtypes and different cell states across different panels (Figure

2B, Figure 3E, and Figure S5) should be maintained to avoid confusion.
15.The manuscript fails to specify which branches cell fates 1 or 2 represent in Figure 3.

References:
Liu X, Chen W, Zeng Q, Ma B, Li Z, Meng T, et al. Single-Cell RNA-Sequencing Reveals Lineage-Specific Regulatory Changes of Fibroblasts and Vascular Endothelial Cells in Keloids. Journal of Investigative Dermatology. 2022;142(1):124-135.e11

Reviewer #2

(Remarks to the Author)
This manuscript provides a state-of-the-art analysis of transcriptomic profiles in keloid samples. The authors propose a set of 25 clinically relevant genes through comprehensive analysis of single-cell and bulk RNA-seq data. While the bioinformatics approach is well-suited for such comprehensive studies, there are several issues that could be addressed by the authors to enhance the manuscript.

1. The authors utilize scRNA-seq data published by Deng et al. in Nature Communications, where the keloid fibroblast was divided into four subpopulations (Mesenchymal, Papillary, Reticular, and Inflammatory fibroblasts) after removing the epidermis and assessing the dermal component of keloids.
Given the use of the same data, it raises questions about the significance of introducing a new classification system. Additionally, the potential presence of doublet cells in Fibroblast 4 should be addressed, especially considering the expression of endothelial cell markers like SELE and ACKR1. There is a question regarding whether Fibroblast6 soley comprised Keloid1.

2. In recent analyses, there have been findings suggesting that interactions and diversity within immune cells and endothelial cells have an impact on prognosis rather than solely relying on the characteristics of fibroblasts. Is it feasible to distinguish risk based solely on the features of fibroblasts? However, the identification of markers like TPSB2 or PECAM1, seemingly unrelated to fibroblasts, warrants an explanation for this discrepancy.

3. In the immunohistochemistry (IHC) results (Figure 7A), TPSB2, a mast cell marker, shows staining in mast cells rather than fibroblasts.
It is worth noting a recent scRNA-Seq study on keloids published in Frontiers in Immunology by Yeo et al., which investigates the roles of mast cells in keloid disease activity. The authors should also address findings from this study, titled "Revisiting roles of mast cells and neural cells in keloid: exploring their connection to disease activity" (Yeo et al., Front Immunol. 2024).

4. The authors state: "Furthermore, it was confirmed that FIBKC was significantly correlated with patient disease diagnosis and diagnostic time by chi-square test ($P < 0.001$) (Figure 4D-E)."
The significance of "diagnostic time" and the meaning of Day 0 and Day 42 require clarification in Figure 4D-E.

5. While fibroblast-related genes may correlate with keloid characteristics, predicting treatment responses is crucial. Without indicators of treatment response or post-surgical recurrence, discussing high risk/low risk is challenging, making it difficult to assert that HFIC correlates with a worse prognosis.

6. Claiming Ouabain as a potential inhibitor without supporting cell experiments lacks substantiation and should be approached with caution.

Minor
1. In Figure 5D, it's challenging to distinguish between male and female colors. Additionally, it's advisable to use "sex" instead of "gender."
2. The transition from Cluster 1/Cluster 2/Cluster 3 to LFIC/MFIC/HFIC in the figure is confusing. Consider adding parentheses next to LFIC and labeling it as Cluster 1 in the figure.
3. Considering the flow of the paper, it might be beneficial to promote Figure S6A to the main figure to enhance clarity and follow the paper's progression.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
In this revised manuscript, the authors have addressed most of the concerns raised in the previous review. However, a few minor but important issues remain:
1. The HPA database identifies PECAM1 as an endothelial cell marker, and its apparent expression in fibroblasts may be due to doublets or ambient RNA contamination. Figure S2 suggests doublets in the data, likely due to single-cell analysis limitations. The claim that PECAM1 is expressed in skin fibroblasts based solely on database results lacks strong support.

Co-staining PECAM1 with fibroblast-specific markers would be essential to confirm this expression.
2. The expression of SELE, ACKR1, CD74, and HLA-DRA in Fibroblast cluster 4 raises the possibility that this cluster may consist of doublets. Moreover, the GSE163973 study by Deng et al. does not report a similar cluster, which calls for clarification regarding this discrepancy.
3. Terms like LFDC, MFDC, and HFDC are used in the abstract without explanation, which may confuse readers unfamiliar with these abbreviations. Providing definitions in the abstract would improve clarity.

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
Where are the results of RNA velocity since authors mentioned this in the results content of paragraph 3.4 (Line 295)? How do you set the starting point of fibroblast pseudotime trajectory? What is your rationale in Figure 3A? What is pseudotime direction in Figure 3E?

In Figure 7A, the images from MMP1 staining in LFDC, FOXC2 staining in MFDC, KLK6 and TPSB2 staining in HFDC are like a normal region, not a true keloid. Please change the representative images and quantify the results again.

I agree with the reviewer 2 that PECAM1 is a very specific marker for endothelial cells. In addition, TPSB2 is a specific marker for mast cells in the skin scRNA-seq dataset. Even though authors used the software to remove doublet cells, sometimes free form mRNA can be captured within one droplet. So, when we identify some markers, we usually confirm whether the markers are specific for fibroblasts. If not, we check the references first that also showed similar results and try to explain the functions of those markers in fibroblast subgroups. However, there is almost no paper showing that PECAM1 and TPSB2 are expressed by fibroblasts. It's appreciated that authors used ICC in fibroblast cell line to verify that fibroblasts can also express a little PECAM1 protein. It's suggested that authors should use RNA FISH to demonstrate that PECAM1 and TPSB2 mRNA are also expressed in fibroblasts of keloid tissues. If authors can't verify this, please remove them from the manuscript.

Version 3:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
Please incorporate the FISH data into the manuscript. Most of the questions have been addressed.

Ms. Aylin Bircan. No.: COMMSBIO-24-1782-T
Manuscript title: Uncovering a Novel Keloid Molecular Typing Based on Fibroblast Differentiation by Integration of Single-cell Sequencing and Multi-omics Analysis
Communications Biology
12th October 2024

Dear reviewers

The interactive review started on: 14th May 2024

Thank you very much for your letter and advice. We have revised the paper carefully, and would like to re-submit it for your consideration. We have addressed the comments raised by the reviewers. Our point-by-point responses are appended to this letter and the amendments are highlighted in red in the revised manuscript.

We would like to express our sincere gratitude to the reviewers for their constructive and positive comments. Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches.

Furthermore, we have refined the abstract, introduction, results and discussion sections of the manuscript according to the checklist for the revised manuscript submission file.

We hope that the revision is acceptable. We look forward to hearing from you soon.

Sincerely yours,
Runzhi Huang, Zhaofan Xia, Shizhao Ji
12th October 2024

Reviewer #1:

Comment 1: Both single-cell sequencing and bulk sequencing belong to transcriptomics. Therefore, the term "multi-omics" is inappropriate here, as it typically denotes analyses that integrate at least two distinct omics datasets, such as transcriptomics and epigenomics.

Response: We appreciate the reviewer's thorough examination and valuable feedback. Based on the integrated analysis of single-cell sequencing and bulk sequencing data, our research finally identified 25 genes related to fibroblast differentiation and clinical diagnosis, and constructed a keloid classification model. The results were verified by immunohistochemistry. Therefore, we modified the title of this paper to "**Uncovering a Novel Fibroblast Differentiation-Based Keloid Classification by Integration of Single-cell and Bulk-RNA Sequencing**", which would be more relevant. What's more, thanks to your precious suggestion, we collected keloids, hypertrophic scars, and normal skin samples from clinical patients for proteomics sequencing, and analyzed the differential protein levels of 25 key genes to further verify our classification. 16 protein expression of 25 DEFDRGs were detected, among which 6 proteins showed obvious expression differences **(Figure S9C in revised manuscript)**. Baseline information of patients was illustrated in **Supplementary table 1 in revised manuscript** (n = 27).

Corresponding revisions in the manuscript are as follows:

Method section (Page 9, Line 251-259)

2.18 Proteomics Sequencing

Keloid, hypertrophic scar, and normal skin samples were collected from clinical patients. After protein extraction, the BCA kit was employed to measure the protein concentration. Equal amounts of proteins from each sample were subjected to enzymatic hydrolysis to obtain peptides. The peptides were dissolved using liquid chromatography and loaded onto Evotip as per the instructions. Subsequently, the Evosep One ultra-high-performance liquid system was utilized for separation. After that, the peptides were injected into the capillary ion source for ionization and then entered the timsTOF Pro2 mass spectrometer for data acquisition to achieve a quantitative analysis of protein expression in the samples.

Result section (Page 13-14, Line 384-393)

The expression of 25 DEFDRGs in three FDBKC clusters was analyzed by the ssGSEA algorithm. Significant differences in gene expression levels among clusters were found (Figure 4G). Almost all DEFDRGs were highly expressed in cluster 1, consistent with the above results (Figure 4A). **In addition, proteomics sequencing results of clinical keloids, hypertrophic scars, and normal skin identified 16 protein expressions of 25 DEFDRGs, among which 6 proteins showed obvious expression differences (Figure S9C), the baseline information of patients was in Supplementary table 1.** To make our results more solid, we additionally selected five datasets for verification, including GSE188952, GSE7890, and GSE44270. Almost all DEFDRGs can be identified in these gene sets, and many genes have statistical differences (Figure S9D-F and Supplementary table 2).

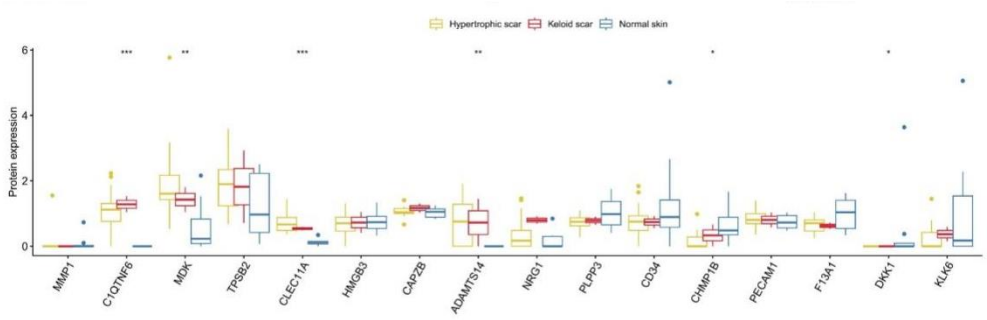


Figure S9C in revised manuscript The differential expression of proteins encoded by 25 DEFDRGs in normal skin, hypertrophic scar, and keloid.

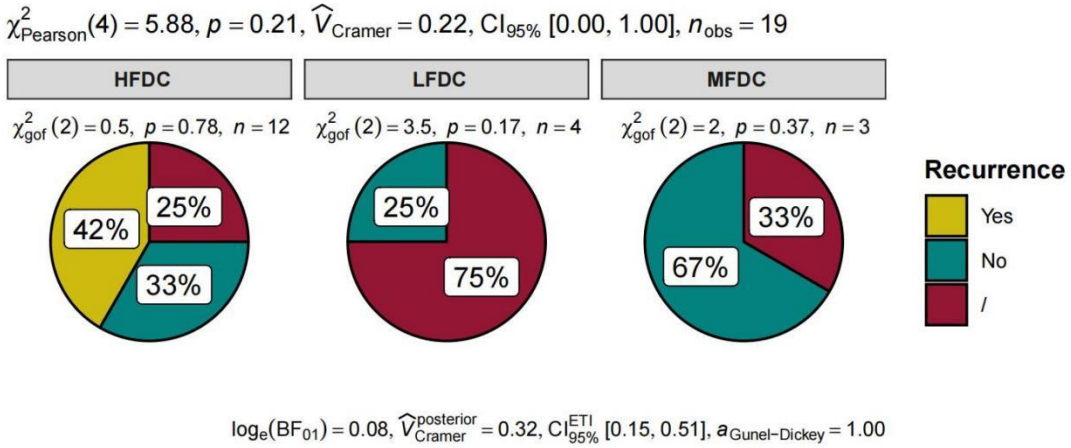
Sample-ID	Group	Age	Sex	Location
SC211-9	Hypertrophic scar	6	male	right lower extremity
SC211-7	Hypertrophic scar	6	male	right lower extremity
SC116-7	Hypertrophic scar	54	male	right elbow
SC144-5	Hypertrophic scar	60	male	right upper extremity
SC141-7	Hypertrophic scar	60	male	right upper extremity
SC196-4	Hypertrophic scar	42	male	abdomen
SC168-4	Hypertrophic scar	8	male	left lower abdomen
SC154-2	Hypertrophic scar	25	male	right elbow；right axilla
SC158-3	Hypertrophic scar	22	female	left thigh
SC55-4	Hypertrophic scar	57	female	right lower extremity
SC70-6	Hypertrophic scar	34	male	right lower extremity
SC68-3	Hypertrophic scar	34	male	right abdomen
SC215-4	Hypertrophic scar	18	female	lumbar and dorsal region
SC216-6	Hypertrophic scar	60	male	neck
SC171-2	Hypertrophic scar	12	male	perineum
SC66-6	Hypertrophic scar	49	female	right upper extremity
SC203-3	Hypertrophic scar	37	male	palm of left hand
NOR202-1	Normal skin	44	female	right thigh
NOR208-4	Normal skin	20	male	prepuce
NOR207-2	Normal skin	51	male	prepuce
NOR197-3	Normal skin	28	male	abdomen
NOR58-12	Normal skin	87	male	left thigh
NOR164-9	Normal skin	72	male	left calf
NOR164-16	Normal skin	72	male	left calf
NOR205-2	Normal skin	/	/	prepuce
SC167-4	keloid	73	male	chest
SC136-3	keloid	67	male	right upper cubital fossa

Supplementary table 1 in revised manuscript Baseline information of clinical keloid, hypertrophic scars, and normal skin patients, including sex, age of patients, and sample location.

Comment 2: The term "prognosis" in keloids predominantly refers to recurrence post-treatment, given that keloids do not metastasize or lead to mortality. The manuscript claims "Patients in the HFIC group had a worse prognosis," yet lacks data substantiating this prognosis.

Response: Thank you very much for your constructive advice. We are very sorry for the mistakes in the statement. We have checked the manuscript and revised it carefully. Due to the lack of sequencing data related to keloid recurrence in public database, our analysis is limited, which leads to limitations in predicting the recurrence and other prognosis of keloids.

Therefore, this study mainly focuses on the occurrence and development of keloids in order to provide insights into the formation of keloids. In addition, during the clinical relevance verification of the keloid classification, we attempted to analyze the recurrence data. However, despite intense efforts, it's hard for all patients to return visit during the pandemic. For the limited enrolled patients (n=19) and high percentage of patients withdrawn, the results of the clinical relevance of keloid recurrence and keloid classification had no statistical difference **(Response letter Figure 1)**. With your suggestion, we have described this limitation in the discussion. In addition, to better reflect the close correlation between the novel keloid classification and the fibroblast differentiation state, we have renamed the classification as FDBKC (Fibroblast Differentiation-based Keloid Classification), which is divided into LFDC, MFDC, and HFDC to correspond to LFIC, MFIC, and HFIC. Modifications and replacements are made throughout the text.



Response letter Figure 1 The pie charts with Chi-square test revealed no significant differences in the recurrence among HFDC, LFDC, and MFDC, three cluster in the novel keloid classification. The results may be caused by limited enrolled patients (n = 19) and high percentage of patients withdrawn.

Corresponding revisions in the manuscript are as follows:

Abstract section (Page 1, Line 3-15)

Keloids are dermal fibroproliferative skin disorders caused by abnormal wound healing, resulting in impaired skin function and aesthetic defects. Abnormal fibroblast proliferation and excessive collagen deposition are involved in keloid formation. This study investigated the role of fibroblast differentiation in the development of keloid. Single-cell and bulk RNA sequencing data of keloids were comprehensively analyzed and 25 clinically relevant fibroblast-differentiation-related genes were finally identified. A novel keloid diagnostic classification, including LFDC, MFDC, and HFDC, was constructed, indicating that this set of genes could serve as new therapeutic targets for keloids. Besides, multiple microarray data, protein sequencing data and immunohistochemistry of key markers in clinical keloid samples were used for further verification. In conclusion, this study constructed keloid molecular classification based on fibroblast differentiation for the first time, contributing to the further understanding of the keloid pathogenesis, which might provide new insights for keloid

diagnosis and treatment.

Result section (Page 12-13, Line 356-372; Page 17, Line 481-489)

3.6. Construction FDBKC by Consistent Clustering Based on DEFDRGs

We constructed FDBKC based on the expression levels of 25 DEFDRGs through a consistent clustering algorithm, which was divided into three clusters (Figure S9A). The delta-area diagram showed changes in the area under the Cumulative Distribution Function (CDF) curve with different *k* values. The *P* value of $k = 3$ is the smallest, indicating that $k = 3$ has the best clustering statistic expression (Figure S9B). We illustrated the expression of 25 key DEFDRGs in various dimensions, including consensus cluster, state, sex, age, and diagnosis (Figure 4A). Gene expression levels in three consensus clusters were relevant to clinicopathological characteristics. Strikingly, strong connections between fibroblast state and consensus clusters were observed. The overall expression levels of fibroblast fate determinant genes were higher in cluster 1, and lower in cluster 2 and cluster 3 with different expression patterns, demonstrating obvious heterogeneity in the obtained clusters (Figure 4A).

Furthermore, it was confirmed that FDBKC was significantly correlated with keloid diagnosis and time by chi-square test ($P < 0.001$) (Figure 4B-C). Keloid time represented day 0 and day 42 after injury in keloid-prone patients. The stacking bar chart showed that healthy samples were mainly in cluster 1 and keloid samples (including day 0 and day 42) were mostly in cluster 3, indicating that cluster 3 has a more severe pathological state.

3.11 Pathological manifestations of different clinical subtypes of keloids

According to the clinical manifestations of patients with different subtypes, the average VSS score in the HFDC group was significantly higher than that in the other two groups (Figure 7B). Keloids in the HFDC group were observed to be thicker than those in the other two groups, nearly 83% greater than 2 mm (Figure 7C). Furthermore, the keloids in the HFDC group were also found to be harder, possibly due to the abnormal distribution of fibroblasts and collagen (Figure 7D). The proportions of keloids with pain, itching, and abnormal blood vessels were relatively high in the HFDC group (Figure 7E-G), reflecting the more severe clinical manifestations of keloids in the HFDC group. These findings were consistent with the results of sc-RNA seq and bulk-RNA seq analyses and proved the efficacy of our novel keloid classification.

Discussion section (Page 21, Line 607-616):

However, this study has limitations. MCs and endothelial cells have also been reported to be closely associated with keloid formation and activity[41,42]. However, in this study, considering easier clinical translation, keloid classification is only carried out based on the differentiation of fibroblasts. Besides, due to biases in pure bioinformatics analysis, it cannot fully generalize fibroblast lineagetrajectories and cell differentiation states. Further cellular and animal experiments are needed to validate the correlation between identified DEFDRGs and phenotypes and to verify the clinical relevance of the new keloid classification. **Furthermore, due to the lack of sequencing data related to keloid prognosis, especially recurrence, in public databases, this keloid classification is only for interpretation of keloid development and cannot predict the prognosis, which partially limits the clinical application of this classification.**

Comment 3: In line 90-91, previous studies delineating the molecular classification of keloid pathogenesis should be incorporated into the introduction for comprehensive context.

Response: We concur with the reviewer's suggestion. There are few studies on molecular classification of keloid in previous studies. We systematically reviewed the literature and related studies and found two types of keloids: a. the H/N molecular classification based on immune cells [1]; b. keloid classification based on glycosphingolipid metabolism-related genes [2], which were helpful to further understand the pathogenesis of keloid and important for the treatment of keloid. However, there are few studies based on molecular classification of keloid, which is the code of the study. We incorporated these two molecular classifications of keloid into the introduction.

Reference

[1] Xu H, Zhu Z, Hu J, Sun J, Wo Y, Wang X, Zou H, Li B, Zhang Y. Downregulated cytotoxic CD8+ T-cell identifies with the NKG2A-soluble HLA-E axis as a predictive biomarker and potential therapeutic target in keloids. Cell Mol Immunol. 2022 Apr;19(4):527-539. doi: 10.1038/s41423-021-00834-1. Epub 2022 Jan 17.

[2] Song B, Zheng Y, Chi H, Zhu Y, Cui Z, Chen L, Chen G, Gao B, Du Y, Yu Z, Song B. Revealing the roles of glycosphingolipid metabolism pathway in the development of keloid: a conjoint analysis of single-cell and machine learning. Front Immunol. 2023 Apr 24;14:1139775. doi: 10.3389/fimmu.2023.1139775.

Corresponding revisions in the manuscript are as follows:

Introduction section (Page 2, Line 35-42):

Based on these clinical dilemmas, exploring the complex cellular and molecular mechanisms, and decoding precise molecular classification of keloids might be favorable to finding keloid pathogenesis and potential treatment. However, there are few studies focused on constructing diagnostic typing of keloids. A previous study found that the downregulated cytotoxic CD8+ T-cell regulated by the NKG2A-soluble HLA-E axis may function as a predictive biomarker and potential therapeutic target in keloids, and then proposed the H/N classification[7]. Besides, one study proposed a keloid classification based on glycosphingolipid metabolism-related genes[8].

Comment 4: The manuscript utilizes "IntegrateData" for single-cell RNA-seq data merging without batch effect correction, leading to evident batch effects in Figure 1B. Such effects also appear to influence fibroblast subtypes, as observed with sample KL1 composed solely of fibroblast 6 (Figure 2B).

Response: Thank you for your valuable advice. Since this study was an analysis of single-cell sequencing data in public databases, we performed canonical correlation analysis (CCA) between different public data according to the analysis method reported in the literature (Deng, C.-C. et al. Single-cell RNA-seq reveals fibroblast heterogeneity and increased mesenchymal fibroblasts in human fibrotic skin diseases. Nat Commun 12, 3709 (2021)). Besides, the reviewer's insightful suggestion has been incorporated as requested. We performed batch effect correction with harmony algorithm, and then performed subsequent analysis (**Figure S7 and S8 in revised manuscript**). The final fibroblast differentiation related

gene set also included 25 key genes identified by "IntegrateData", indicating the reliability of our study.

Corresponding revisions in the manuscript are as follows：
Result section (Page 12, Line 351-355):

Moreover, the harmony algorithm was employed in this study to mitigate the batch effect. Subsequently, fibroblast subtype analysis and trajectory analysis were conducted to identify the DEG set related to fibroblast differentiation, named the Harmony-Monocle-sig set. This set contains the 25 DEFDRGs determined previously, further validating the reliability of our analysis (Figure S7-S8).

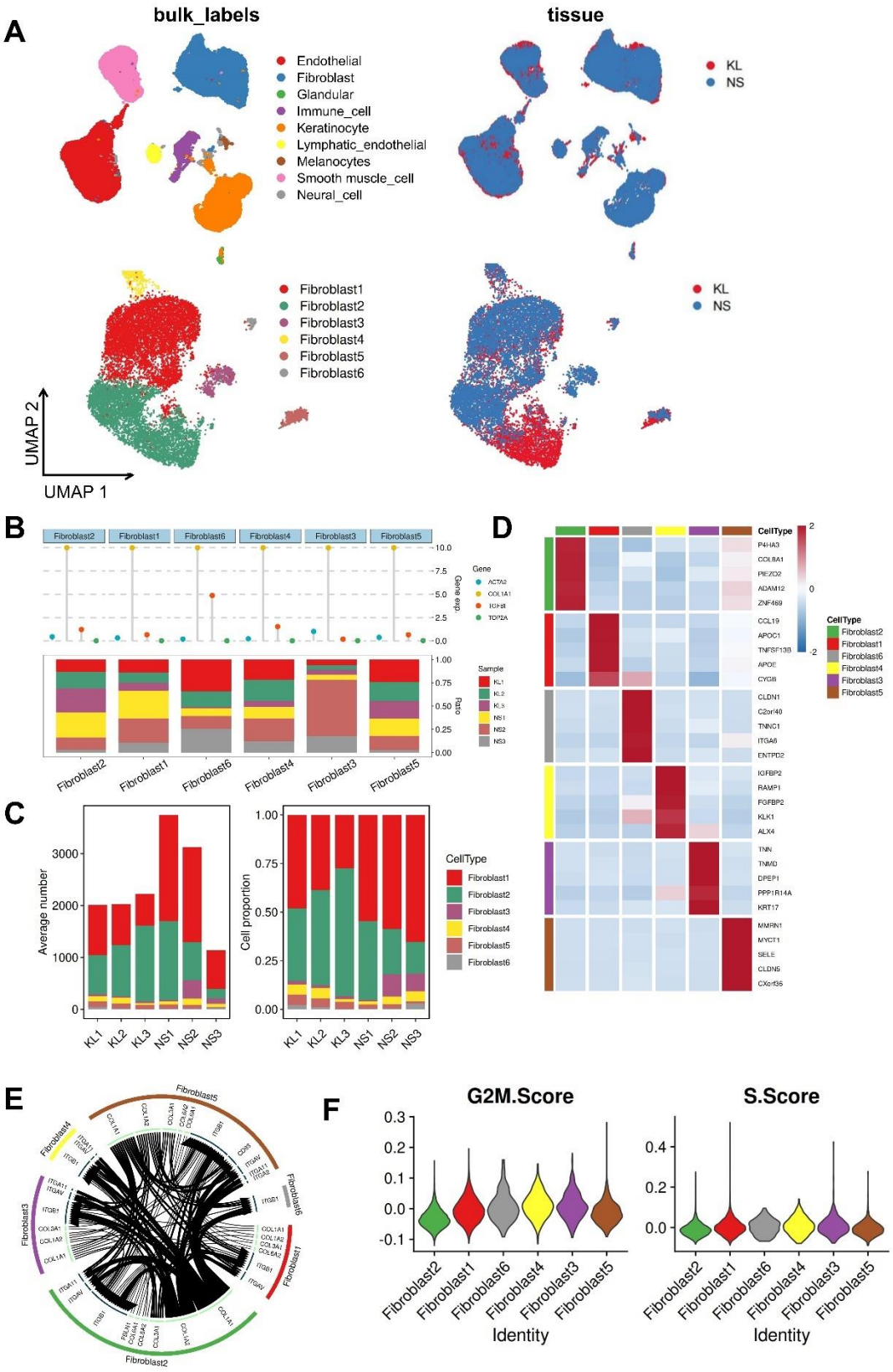


Figure S7 in revised manuscript Sc-RNA seq and fibroblast subtype analysis adjusted by harmony. (A) Nine cell types were identified. Among them, fibroblasts can be divided into six subtypes. (B) The Cleveland dot plot shows 4 classical markers' expression (ACTA2, COL1A1, TGFBI, and TOP2A) in fibroblasts. The bar plot displays the proportion of each fibroblast cluster

in 6 samples. (C) The histogram shows the average number and proportion of fibroblast subtypes in different samples. (D) Top 5 marker genes of each fibroblast subtype. (E) Cell communication between fibroblast subtypes. (F) The cell cycle status of the fibroblast subtypes.

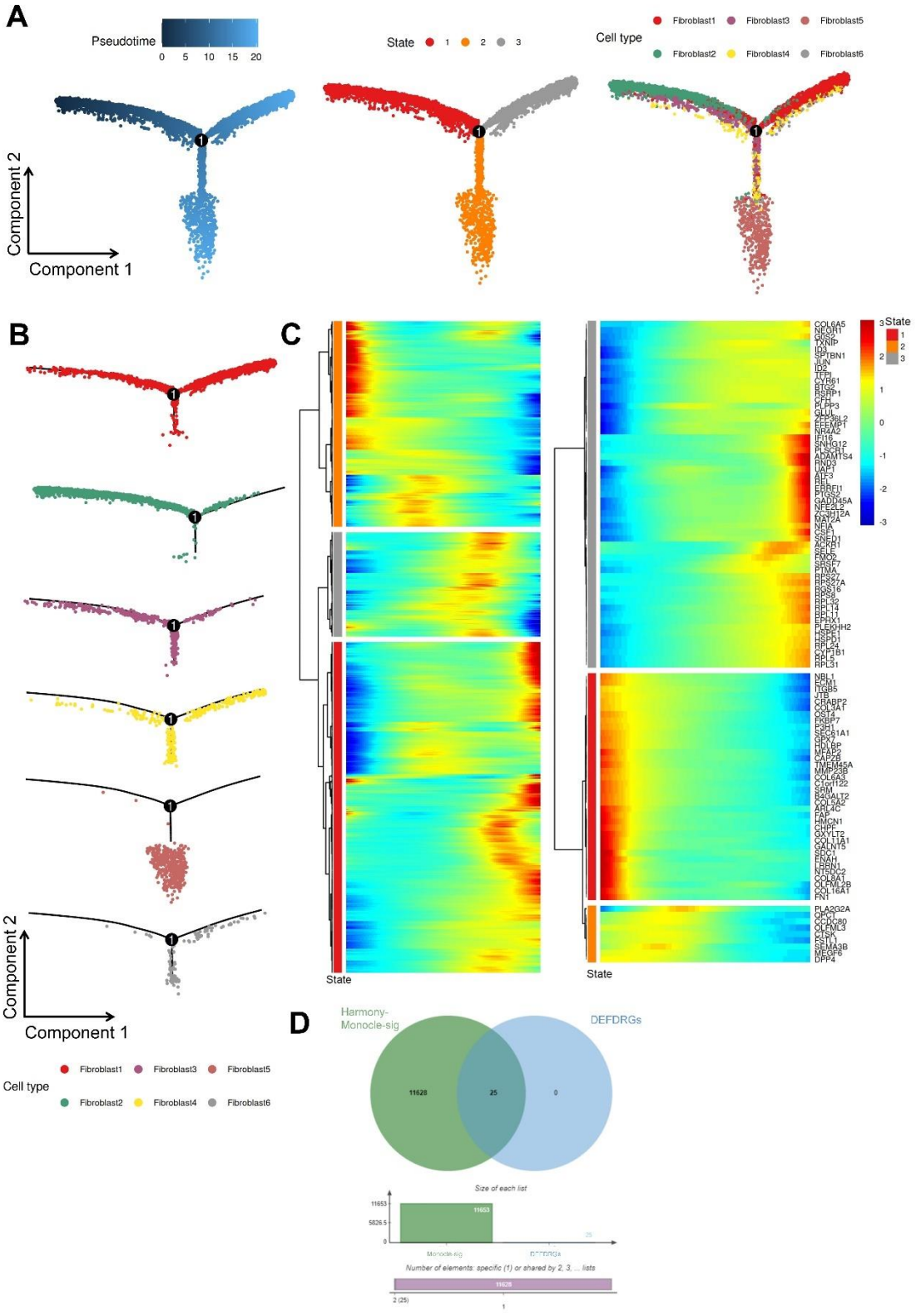


Figure S8 in revised manuscript Pseudotime analysis and identification of DEFDRGs adjusted by harmony. (A) Fibroblast differentiation sequence and detection of three differentiation states. (B) Display of fibroblast subtypes on different differentiation trajectories. (C) The tendency to different differentiation directions of different states and the expression of the top 100 representative fate-related genes differentially expressed across pseudo-time. (D) Newly obtained FDRGs include the 25 DEFDRGs identified, proving the stability of the results.

Comment 5: The manuscript uses two distinct control tissues—normal scars for scRNA-seq and normal skin for bulk RNA-seq—yet intersects these comparisons without justification for such choice.

Response: We concur with the reviewer's suggestion and have made the corresponding revisions. We made such a contrast because of the limitations of public database. There are few relevant single-cell sequencing, and we finally chose the dataset (GSE163973), which is a comparison of keloid and normal scar. Through systematic analysis of this data, we identified a gene set related to fibroblast differentiation. Besides, we systematically retrieved keloid-related bulk sequencing and found two kinds of datasets: one comparing keloid with normal skin and the other one comparing keloid with normal scar. To reveal the molecular mechanism of keloid formation and find the most relevant gene sets, we chose the SRP137071 dataset, which enrolled relatively more samples and made a time series transcriptomic analysis of keloids with different time and normal skins. At the same time, we performed a differential gene expression analysis of the 25 differentiation-related genes on other datasets (GSE7890; GSE145725; GSE188952). Besides, protein sequencing was also conducted to identify the protein expression differences of 25 key genes. The above results are shown in the **Figure S9 C-F and Supplementary table 2** in revised manuscript.

Corresponding revisions in the manuscript are as follows:

Result section (Page 13-14, Line 384-393)

The expression of 25 DEFDRGs in three FDBKC clusters was analyzed by the ssGSEA algorithm. Significant differences in gene expression levels among clusters were found (Figure 4G). Almost all DEFDRGs were highly expressed in cluster 1, consistent with the above results (Figure 4A). **In addition, proteomics sequencing results of clinical keloids, hypertrophic scars, and normal skin identified 16 protein expressions of 25 DEFDRGs, among which 6 proteins showed obvious expression differences (Figure S9C), the baseline information of patients was in Supplementary table 1. To make our results more solid, we additionally selected five datasets for verification, including GSE188952, GSE7890, and GSE44270. Almost all DEFDRGs can be identified in these gene sets, and many genes have statistical differences (Figure S9D-F and Supplementary table 2).**

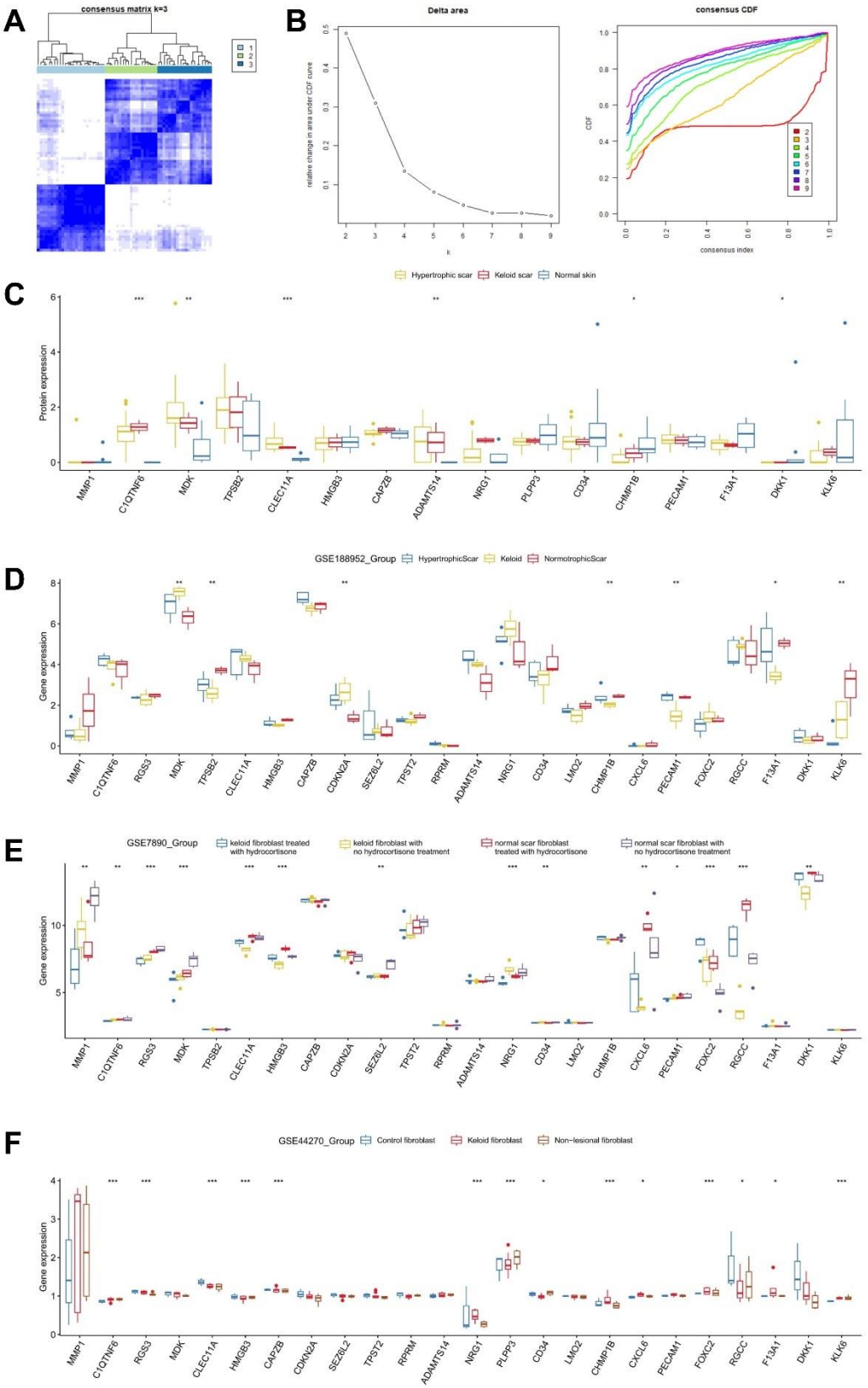


Figure S9 in revised manuscript Consensus clustering and verification in public databases (related to Figure 4). (A)Consensus clustering analysis based on 25 DEFDRGs. (B)The clustering effect is more specific when k = 3. (C)The differential expression of proteins encoded by 25 DEFDRGs in normal skin, hypertrophic scar, and keloid. (D-F) The

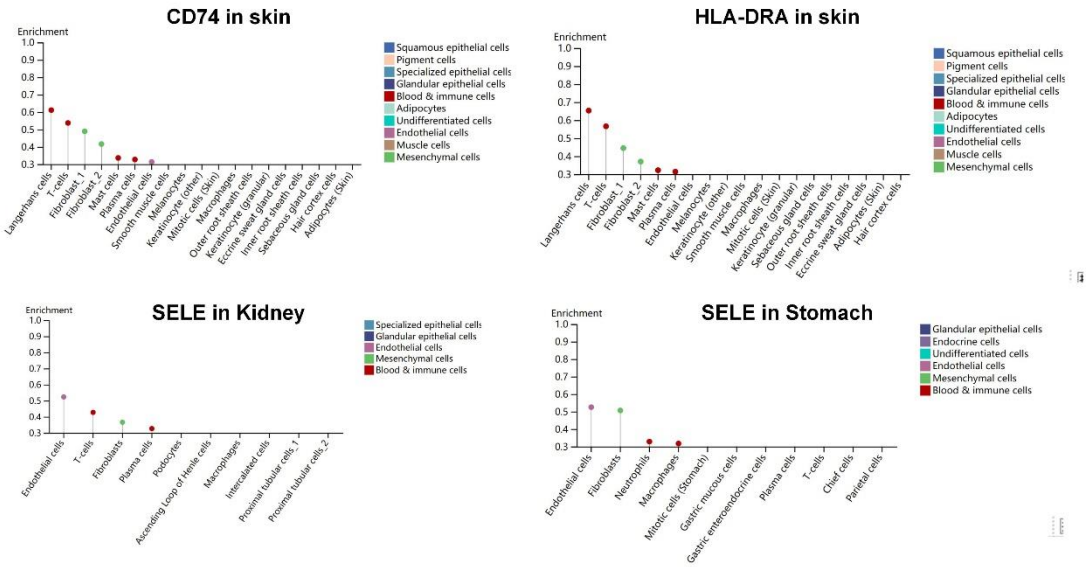
verification of the differential expression of 25 DEFDRGs using sequencing data in public databases. These three datasets found that many genes among these 25 DEFDRGs had statistical differences.

GEO ID	Sequencing type	Sample number	Sample information	Detected gene of 25 DEFDRGs	DEG Number	Proportion of DEG	Reference
GSE188952	RNA sequencing	12	tissues of normotrophic scars, hypertrophic scars, and keloids	24	7	29.2%	/
GSE7890	Microarray	19	fibroblasts of normal scars and keloids	24	14	58.3%	PMID: 17989729
GSE44270	Microarray	32	fibroblasts of keloid scars, adjacent non-keloid skins, and normal skins	24	14	58.3%	PMID: 23815228

Supplementary table 2 in revised manuscript Data from public databases were used to verify the differential expression changes of 25 DEFDRGs in keloids, normal scars and normal skin.

Comment 6: The top 5 DEGs in fibroblast subtype 4 are indicative of endothelial cells (SELE) and immune cells (CD74, HLA-DRA). This prompts an investigation into whether this cluster represents doublets, as no doublet detection or removal was conducted.

Response: Thank you for your valuable advice. As mentioned in the previous response in question 4, we performed canonical correlation analysis (CCA) between different public data according to the analysis method reported in the literature. We also searched the HPA database, and found that CD74 and HLA-DRA could be detected in skin fibroblast, and SELE was expressed in fibroblasts of stomach and kidney tissues **(Response letter Figure 2)**, suggesting that it may be a reasonable result. On the other hand, the harmony algorithm was used for batch effect correction, and the 25 differentiation-related genes were found in final fibroblast differentiation-related gene set **(Figure S7 and S8 in revised manuscript)**, which consistent with the current results.



Response letter figure 2 The expression of CD74, HLA-DRA, and SELE queried on the HPA website. CD74 (https://www.proteinatlas.org/ENSG00000019582-CD74/tissue+cell+type);

HLA-DRA (https://www.proteinatlas.org/ENSG00000204287-HLA-DRA/tissue+cell+type); SELE（https://www.proteinatlas.org/ENSG00000007908-SELE/tissue+cell+type）

Conment 7: Fibroblast 2 expresses genes like POSTN, ASPN, and COMP, consistent with previous scRNA studies identifying this cluster as a major contributor to keloid formation and potentially terminal cells in keloids. However, characterizing fibroblast 2 as the origin of differentiation in the study contradicts prior findings (Liu et al. 2022).

Response: Apologies for any misunderstanding caused by our inappropriate expression. As depicted in Figure 2 of our manuscript, fibroblast 2 constituted a relatively high proportion in keloid samples and showed high expression of the POSTN, ASPN, and COMP genes, suggesting its potential crucial role in the formation of keloids. We have made revisions in the manuscript. In this study, we used monocle 2 algorithm for pseudotime analysis. However, due to the limitations of the algorithm itself, the results were unreliable when there was just algorithmic validation. To get time-series clinical samples in the process of keloid formation is the best option. Only by matching the fitting time with the real time point can the exact time sequence change be obtained. However, it is challenging in clinical practice especially in early stage of keloid formation. Our research group is dedicated to conducting this work to collect clinically relevant samples and corresponding data.

Corresponding revisions in the manuscript are as follows:

Result section (Page 11, Line 305-318):

3.4. Reveal Keloid Fibroblast Differentiation Fate Decisions and Dynamics of Expression with Trajectory Analysis

To simulate fibroblast gene expression patterns in keloids, we used Monocle 2 for trajectory analysis and reconstructed the cell lineage trajectory based on transcriptional characteristics in fibroblast differentiation. Each fibroblast cluster was extracted and trajectory analysis was performed. Figure 3A-C demonstrates the pseudo-time trajectory and possible differentiation order of all fibroblast subtypes. Different expression levels of pivotal genes determined cell differentiation fate and divided fibroblasts into five different states by branch points 1 and 2. Furthermore, the pseudo-time and cell lineage trajectory of each fibroblast cluster were illustrated separately in trajectory plots (Figure 3D). Both fibroblast 2 and fibroblast 4 are in state 1, indicating that the two may be in a relatively similar differentiation state, and fibroblast 4 might affect keloid formation. The differential gene expression profiles of the five states are shown in Figure 3E. State 1, state 3, and state 5 have a similar differentiation trend, while state 2 showed an opposite direction.

Comment 8: While 25 DEFDRGs were identified with correlations to keloids or healthy individuals, these associations lack clinical relevance without considering recurrence—a pivotal factor absent from the analysis.

Response: Thank you for your valuable advice. We regret that we are unable to accommodate this suggestion at this time due to the lack of data related to keloid recurrence in public dataset. Just as answered for question 2, it is difficult for us to reveal the molecular mechanism related to keloid recurrence. Therefore, this study is devoted to understanding the occurrence and development of keloids. Moreover, we enrolled the recurrence indicators of 19 patients in conducting clinical

correlation analysis of molecular subtypes. However, due to the limited cases and the loss of follow-up, the results showed no statistical differences **(Response letter figure 1)**, which also limited our study in keloid recurrence. We also wrote this limitation into the "discussion" section of the manuscript.

Corresponding revisions in the manuscript are as follows:
Discussion section (Page 21, Line 607-616):
However, this study has limitations. MCs and endothelial cells have also been reported to be closely associated with keloid formation and activity[41,42]. However, in this study, considering easier clinical translation, keloid classification is only carried out based on the differentiation of fibroblasts. Besides, due to biases in pure bioinformatics analysis, it cannot fully generalize fibroblast lineagetrajectories and cell differentiation states. Further cellular and animal experiments are needed to validate the correlation between identified DEFDRGs and phenotypes and to verify the clinical relevance of the new keloid classification. **Furthermore, due to the lack of sequencing data related to keloid prognosis, especially recurrence, in public databases, this keloid classification is only for interpretation of keloid development and cannot predict the prognosis, which partially limits the clinical application of this classification.**

Comment 9: The manuscript's claim that "FIBKC was significantly correlated with patient disease diagnosis and diagnostic time" does not substantiate the conclusion that "cluster 1 in FIBKC had a better prognosis, while cluster 3 was the opposite."
Response: We accept the reviewer's suggestion and have revised the manuscript accordingly. We are sorry that we made mistakes in the expression of the article, and we have checked and revised the full manuscript.

Corresponding revisions in the manuscript are as follows:
Result section (Page 12-13, Line 356-372; Page 17, Line 481-489)
3.6. Construction FDBKC by Consistent Clustering Based on DEFDRGs

We constructed FDBKC based on the expression levels of 25 DEFDRGs through a consistent clustering algorithm, which was divided into three clusters (Figure S9A). The delta-area diagram showed changes in the area under the Cumulative Distribution Function (CDF) curve with different k values. The P value of $k = 3$ is the smallest, indicating that $k = 3$ has the best clustering statistic expression (Figure S9B). We illustrated the expression of 25 key DEFDRGs in various dimensions, including consensus cluster, state, sex, age, and diagnosis (Figure 4A). Gene expression levels in three consensus clusters were relevant to clinicopathological characteristics. Strikingly, strong connections between fibroblast state and consensus clusters were observed. The overall expression levels of fibroblast fate determinant genes were higher in cluster 1, and lower in cluster 2 and cluster 3 with different expression patterns, demonstrating obvious heterogeneity in the obtained clusters (Figure 4A).

Furthermore, it was confirmed that FDBKC was significantly correlated with keloid diagnosis and time by chi-square test ($P < 0.001$) (Figure 4B-C). Keloid time represented day 0 and day 42 after injury in keloid-prone patients. The stacking bar chart showed that healthy samples were mainly in cluster 1 and keloid samples (including day 0 and day 42) were mostly in cluster 3, indicating that cluster 3 has a more severe pathological state.

3.11 Pathological manifestations of different clinical subtypes of keloids

According to the clinical manifestations of patients with different subtypes, the average VSS score in the HFDC group was significantly higher than that in the other two groups (Figure 7B). Keloids in the HFDC group were observed to be thicker than those in the other two groups, nearly 83% greater than 2 mm (Figure 7C). Furthermore, the keloids in the HFDC group were also found to be harder, possibly due to the abnormal distribution of fibroblasts and collagen (Figure 7D). The proportions of keloids with pain, itching, and abnormal blood vessels were relatively high in the HFDC group (Figure 7E-G), reflecting the more severe clinical manifestations of keloids in the HFDC group. These findings were consistent with the results of sc-RNA seq and bulk-RNA seq analyses and proved the efficacy of our novel keloid classification.

Several minor corrections:

Comment 10: Line 63 should replace "topical steroid injection" with the correct description.

Line 76 should read "at a single-cell level".

Line 124 should be "single-cell RNA-sequencing was performed on samples".

Response: We accept the reviewer's suggestion and have revised the manuscript accordingly. Besides, we apologize for any typos and grammatical errors found in our manuscript. We have carefully reviewed and corrected them in our revised version. As non-native English authors, we fully acknowledge that our writing may contain grammar, spelling, and word choice issues, and we appreciate your understanding. To enhance the overall quality of our paper and better convey our points, we first utilized Grammarly (https://www.grammarly.com) and 1Checker (https://www.1checker.com) to detect and correct errors in spelling, grammar, punctuation, and other aspects. Additionally, we used StyleWriter (http://www.stylewriter-usa.com) to refine the style of our writing, in order to improve the clarity, fluency, and readability of the article. Then we thoroughly revised our paper from start to finish using WhiteSmoke (https://www.whitesmoke.com) and QuillBot (https://quillbot.com). Finally, we sought assistance from a native speaker to polish our paper. All amendments are highlighted in red in the revised manuscript. If you believe that further refinement is necessary, please feel free to recommend.

Corresponding revisions in the manuscript are as follows:

Introduction section (Page 1, Line 27-29; Page 2, Line 44-47):

Currently, various treatment therapies on keloids have been developed, including intralesional corticosteroid injection, cryotherapy, surgical resection, radiotherapy, and laser therapy[4].

With the development of single-cell RNA sequencing (sc-RNA seq) , analyzing a large number of transcriptome landscape, heterogeneity, and regulatory networks of fibroblasts under both homeostasis and pathological conditions at a single-cell level is feasible[9].

Materials and Methods section (Page 3, Line 67-69):

Keloid and normal scar tissues were obtained from patients who were confirmed to have definite clinical evidence of keloids or normal scars, and sc-RNA seq was performed on these samples using the 10x genomics platform (accession code: GSE163973).

Comment 11: The term "muscle fiber cell" is unconventional in skin scRNA-seq studies; "smooth muscle cells" or "mural cells" would be more common.
Response: Thank you for your valuable and constructive feedback. We have checked the manuscript and figure, and replace “muscle fiber cell” with “smooth muscle cell”.

Comment 12: Consistency in color representation of different fibroblast subtypes and different cell states across different panels (Figure 2B, Figure 3E, and Figure S5) should be maintained to avoid confusion.
Response: Thank you for your constructive suggestions. We have performed color unification for different fibroblast subtypes and different cell states in Figure 2B, Figure 3E, and Figure S5. In addition, we also checked all the figures and unified the colors of the same indicators to avoid confusion.

Comment 13: The manuscript fails to specify which branches cell fates 1 or 2 represent in Figure 3.
Response: Thank you for your praise and the constructive suggestions. Because of the limitations of pseudotemporal analysis in interpreting the true trajectory of fibroblast differentiation, the purpose of our pseudotemporal analysis was mainly to find the genes related to fibroblast differentiation. Thus, cell fate 1 and 2 represent only two directions in the trajectory analysis showing the different processes in which genes may be involved, but with no substantial implications.

Reviewer #2:

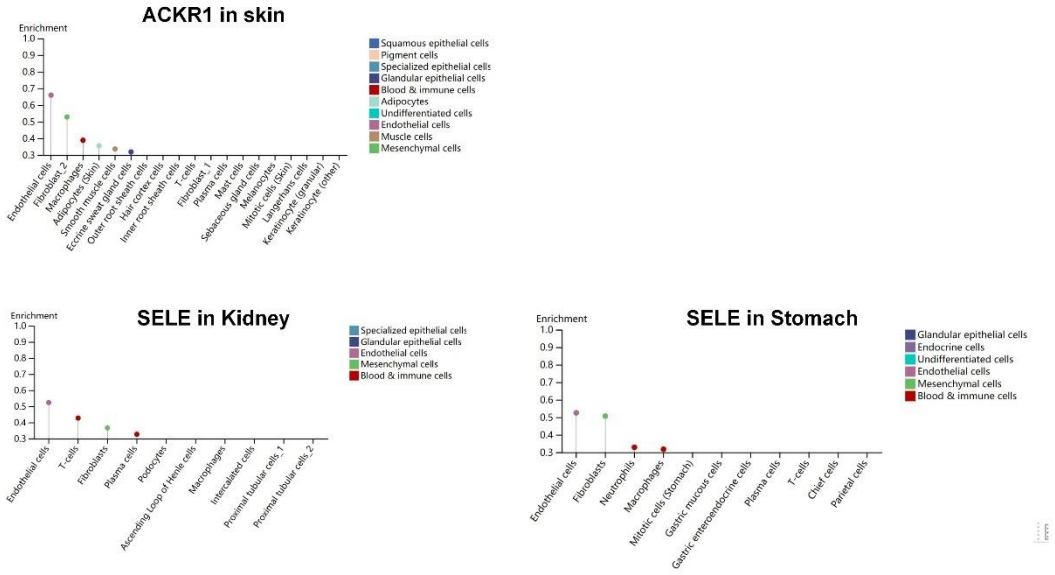
Comment 1: The authors utilize scRNA-seq data published by Deng et al. in Nature Communications, where the keloid fibroblast was divided into four subpopulations (Mesenchymal, Papillary, Reticular, and Inflammatory fibroblasts) after removing the epidermis and assessing the dermal component of keloids.

Given the use of the same data, it raises questions about the significance of introducing a new classification system.

Additionally, the potential presence of doublet cells in Fibroblast 4 should be addressed, especially considering the expression of endothelial cell markers like SELE and ACKR1.

There is a question regarding whether Fibroblast6 soley comprised Keloid1.

Response: We concur with the reviewer's suggestion and have made the corresponding revisions. We conducted an analysis using single-cell sequencing from public databases, which was published by Deng et al. in Nature Communications. The single-cell sequencing data reported the heterogeneity of fibroblasts. Additionally, it was discovered that mesenchymal fibroblast cells play an important role in keloid development. This indicates that identifying the heterogeneity and characteristics of fibroblasts holds great guiding significance for uncovering the occurrence and development of keloids. In our research, we further identified fibroblast differentiation-related genes by utilizing the high-quality single-cell sequencing data of this article and combined it with bulk data with time series. As a result, we identified a gene set related to both the clinical event of keloids and fibroblast differentiation, named DEFDRGs. Moreover, keloid typing was constructed based on this gene set, providing a reference value for further interpreting the development of keloids and evaluating their severity. Furthermore, with regard to the issue of fibroblast 4 expressing SELE and ACKR1, we searched the HPA database and found that ACKR1 was expressed in skin fibroblasts (Response letter Figure 1). However, SELE was not detected in skin fibroblasts but was expressed in fibroblasts of stomach and kidney tissues, suggesting that it may be a reasonable result (Response letter Figure 1). On the other hand, the harmony algorithm was used for batch effect correction, and the 25 differentiation-related genes were found in the final fibroblast differentiation-related gene set (Figure S7-S8 in revised manuscript), which is consistent with the current results.



Response letter Figure 1 The expression of ACKR1 and SELE queried on the HPA website. ACKR1 (https://www.proteinatlas.org/ENSG00000213088-ACKR1/tissue+cell+type); SELE（https://www.proteinatlas.org/ENSG00000007908-SELE/tissue+cell+type）

Corresponding revisions in the manuscript are as follows：

Result section (Page 12, Line 351-355):

Moreover, the harmony algorithm was employed in this study to mitigate the batch effect. Subsequently, fibroblast subtype analysis and trajectory analysis were conducted to identify the DEG set related to fibroblast differentiation, named the Harmony-Monocle-sig set. This set contains the 25 DEFDRGs determined previously, further validating the reliability of our analysis (Figure S7-S8).

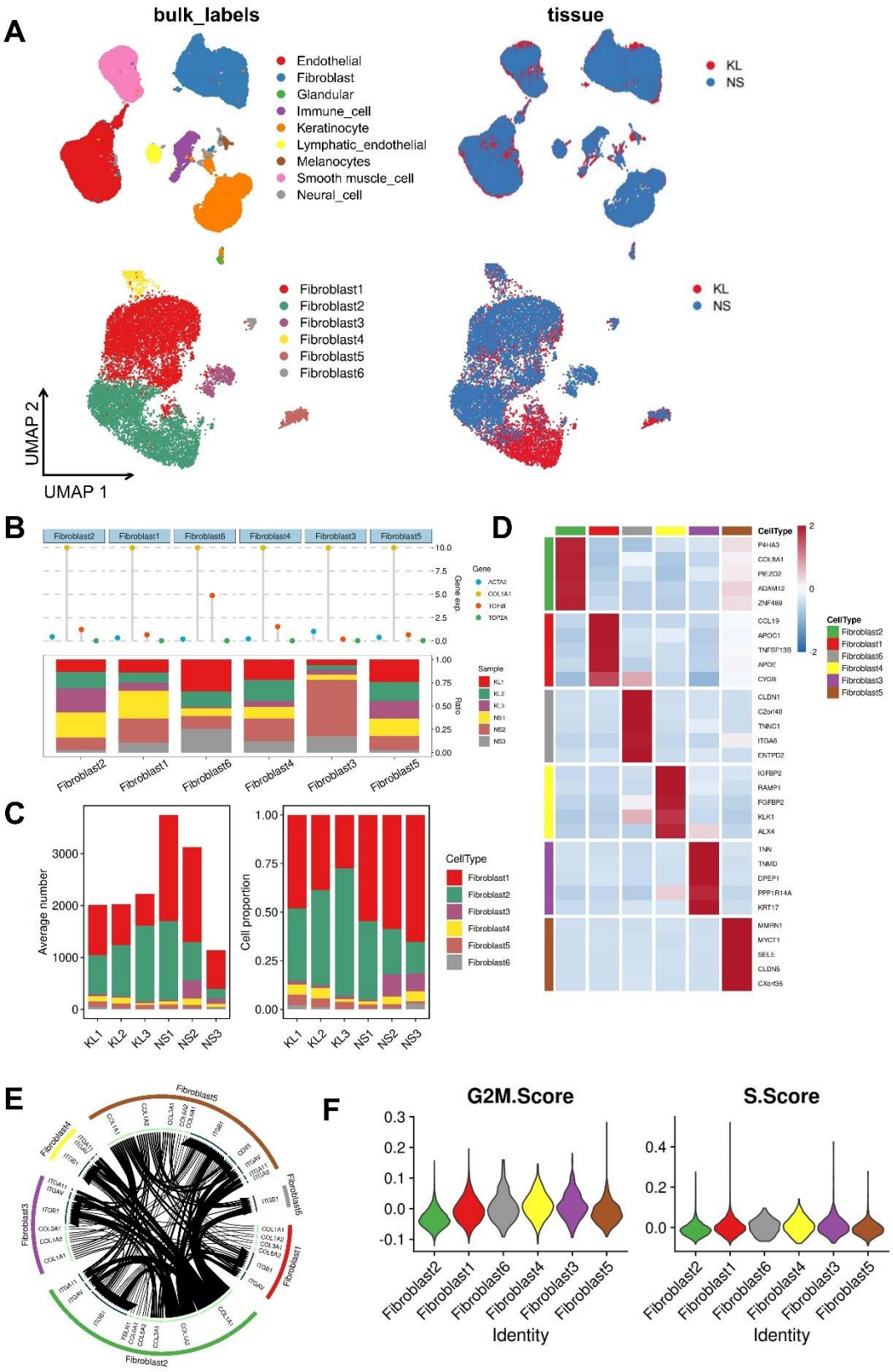


Figure S7 in revised manuscript Sc-RNA seq and fibroblast subtype analysis adjusted by harmony. (A) Nine cell types were identified. Among them, fibroblasts can be divided into six subtypes. (B) The Cleveland dot plot shows 4 classical markers' expression (ACTA2, COL1A1, TGFBI, and TOP2A) in fibroblasts. The bar plot displays the proportion of each

fibroblast cluster in 6 samples. (C) The histogram shows the average number and proportion of fibroblast subtypes in different samples. (D) Top 5 marker genes of each fibroblast subtype. (E) Cell communication between fibroblast subtypes. (F) The cell cycle status of the fibroblast subtypes.

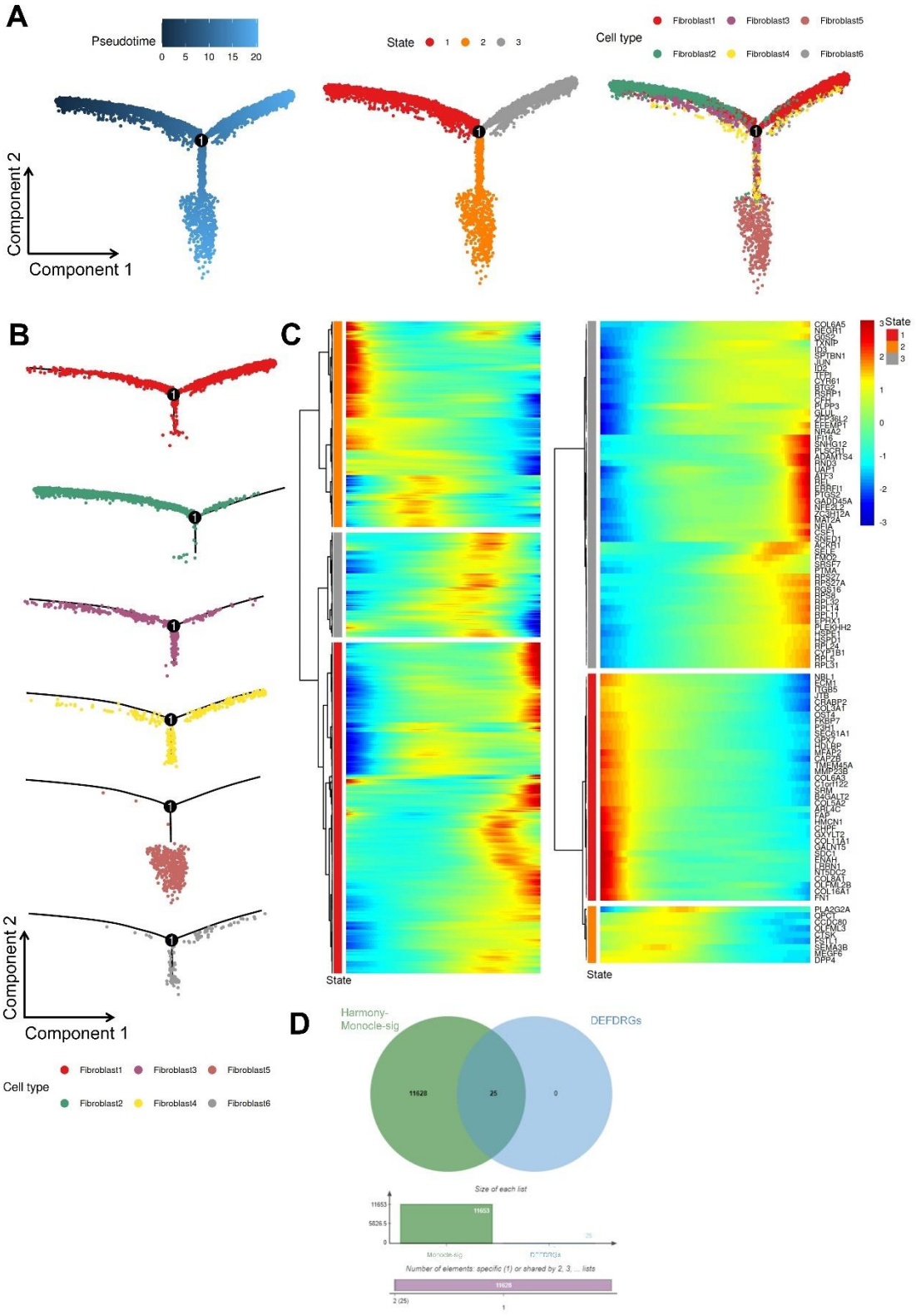


Figure S8 in revised manuscript Pseudotime analysis and identification of DEFDRGs adjusted by harmony. (A) Fibroblast differentiation sequence and detection of three differentiation states. (B) Display of fibroblast subtypes on different differentiation trajectories. (C) The tendency to different differentiation directions of different states and the

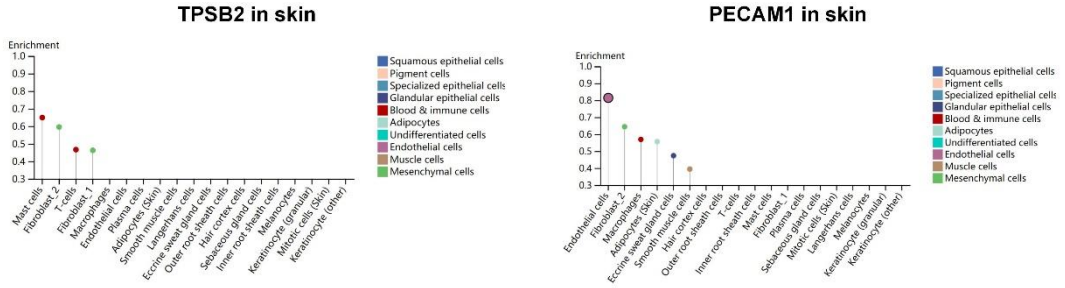
expression of the top 100 representative fate-related genes differentially expressed across pseudo-time. (D) Newly obtained FDRGs include the 25 DEFDRGs identified, proving the stability of the results.

Comment 2: In recent analyses, there have been findings suggesting that interactions and diversity within immune cells and endothelial cells have an impact on prognosis rather than solely relying on the characteristics of fibroblasts. Is it feasible to distinguish risk based solely on the features of fibroblasts? However, the identification of markers like TPSB2 or PECAM1, seemingly unrelated to fibroblasts, warrants an explanation for this discrepancy.
Response: The reviewer's comment is well taken, and we have revised the section as suggested. Multiple cells play important roles in the progression of keloid, including immune cells and endothelial cells [1,2]. The abnormal proliferation of fibroblasts and the resulting excessive collagen deposition, without question, promote the development of keloids. Recently, many high-quality studies have confirmed the key role of fibroblast heterogeneity and increased mesenchymal in keloid [3,4]. All the studies showed the importance of fibroblasts. At the same time, this study aimed to construct a diagnostic classification of keloid and find reliable marker genes, which would be transformed into immunohistochemical kits easily. It is undeniable that cell interaction is important in the development of keloid, which is also the limitation of our article, and we described this limitation in discussion.
Besides, in this study, by combining the differential genes related to fibroblast differentiation, differential genes related to clinical keloid diagnosis, and differential genes in the sequential formation of keloid, 25 key DEFDRGs were found, among which TPSB2 and PECAM1 were involved. Generally, TPSB2 and PECAM1 are the classical markers of mast cells and endothelial cells respectively. We further searched the HPA database and found that TPSB2 and PECAM1 were both expressed in the skin next to mast cells and endothelial cells (Response letter Figure 2). In addition, we selected bulk sequencing data (GSE7890; GSE145725) from keloid, normal skin, or normal scar fibroblasts for differential expression analysis, and found that TPSB2 and PECAM1 expression could be found. The results showed that the two genes may play important roles in the process of fibroblast differentiation.

References

[1] Yeo E, Shim J, Oh SJ, Choi Y, Noh H, Kim H, Park JH, Lee KT, Kim SH, Lee D, Lee JH. Revisiting roles of mast cells and neural cells in keloid: exploring their connection to disease activity. Front Immunol. 2024 Mar 8;15:1339336. doi: 10.3389/fimmu.2024.1339336.
[2] Liu X, Chen W, Zeng Q, Ma B, Li Z, Meng T, Chen J, Yu N, Zhou Z, Long X. Single-Cell RNA-Sequencing Reveals Lineage-Specific Regulatory Changes of Fibroblasts and Vascular Endothelial Cells in Keloids. J Invest Dermatol. 2022 Jan;142(1):124-135.e11. doi: 10.1016/j.jid.2021.06.010. Epub 2021 Jul 7.
[3] Deng CC, Hu YF, Zhu DH, Cheng Q, Gu JJ, Feng QL, Zhang LX, Xu YP, Wang D, Rong Z, Yang B. Single-cell RNA-seq reveals fibroblast heterogeneity and increased mesenchymal fibroblasts in human fibrotic skin diseases. Nat Commun. 2021 Jun 17;12(1):3709. doi: 10.1038/s41467-021-24110-y.
[4] Shim J, Oh SJ, Yeo E, Park JH, Bae JH, Kim SH, Lee D, Lee JH. Integrated Analysis of Single-Cell and Spatial Transcriptomics in Keloids: Highlights on Fibrovascular Interactions in Keloid Pathogenesis. J Invest Dermatol. 2022 Aug;142(8):2128-2139.e11. doi:

10.1016/j.jid.2022.01.017.



Response letter Figure 2 The expression of ACKR1 and SELE queried on the HPA website. TPSB2(https://www.proteinatlas.org/ENSG00000197253-TPSB2/tissue+cell+type); PECAM1(https://www.proteinatlas.org/ENSG00000261371-PECAM1/tissue+cell+type).

Corresponding revisions in the manuscript are as follows:

Discussion section (Page 21, Line 607-616):

However, this study has limitations. MCs and endothelial cells have also been reported to be closely associated with keloid formation and activity[41,42]. However, in this study, considering easier clinical translation, keloid classification is only carried out based on the differentiation of fibroblasts. Besides, due to biases in pure bioinformatics analysis, it cannot fully generalize fibroblast lineagetrajectories and cell differentiation states. Further cellular and animal experiments are needed to validate the correlation between identified DEFDRGs and phenotypes and to verify the clinical relevance of the new keloid classification. Furthermore, due to the lack of sequencing data related to keloid prognosis, especially recurrence, in public databases, this keloid classification is only for interpretation of keloid development and cannot predict the prognosis, which partially limits the clinical application of this classification.

Comment 3: In the immunohistochemistry (IHC) results (Figure 7A), TPSB2, a mast cell marker, shows staining in mast cells rather than fibroblasts.

It is worth noting a recent scRNA-Seq study on keloids published in Frontiers in Immunology by Yeo et al., which investigates the roles of mast cells in keloid disease activity. The authors should also address findings from this study, titled "Revisiting roles of mast cells and neural cells in keloid: exploring their connection to disease activity" (Yeo et al., Front Immunol. 2024).

Response: We acknowledge the reviewer's point and have revised the manuscript to address the concern. TPSB2 serves as a marker for mast cells. Inevitably, there will be positive expression of mast cells in the immunohistochemical staining of keloids. Therefore, our results could only show the overall expression of TPSB2 in keloid, but were unable to show the expression of TPSB2 in fibroblasts specifically. We have revised the expression in the article to make it more accurate and avoid misunderstanding. At the same time, we have studied the article you recommended. In this study, single cell sequencing of active keloid, quiescent keloid and mature scar samples was performed, and mast cells, lymphocytes and macrophages were significantly found aggregated in active keloid. Genes related to mast cell activation and degranulation, including FCER1G, BTK and GATA2, were found to be significantly associated with mast cell activation and degranulation, which revealed the role

of mast cell heterogeneity in keloid disease activity status. However, in our analysis, the differences in immune cells among keloids of different risks are relatively small (Figure S10 in manuscript). This might be caused by the small sample size or individual differences among the population.

Corresponding revisions in the manuscript are as follows:

Result section (Page 14, Line 395-403: Page 16-17, Line 468-480)

The correlation between immune cells or function and PCA FS was demonstrated in Figure S10A. Additionally, we investigated the immune infiltration characteristics of 23 immune cells in all samples from the high- and low-risk groups. The relative percent of every immune cell subtype in each sample is shown in Figure S10B. The difference in infiltration and function of immune cells between high- and low-risk groups is also displayed in Figure S10C-D. The above results suggest that there was little difference in the immune infiltration pattern between the high- and low-risk groups. The result was not consistent with previous studies, which found that mast cell (MC) heterogeneity was related to keloid activity status[42]. The reason may be limited sample size for analysis and individual differences.

3.11 Pathological manifestations of different clinical subtypes of keloids

Informed consent from 19 keloid patients was obtained before collecting abandoned samples and related disease information. A typical marker from each type of FIBKC was selected for immunohistochemistry (IHC). As HFDC may be associated with worse pathological manifestations and its identification is more instructive for clinical treatment, two markers of HFDC were chosen. **Then, IHC of MMP1, FOXC2, KLK6, and TPSB2 was performed on keloid skin tissues, and professional pathologists were invited to rate the expression of related marker genes in keloid to determine patient clinical subtypes.** The baseline information, Vancouver Scar Scale (VSS) scores, and IHC scores are displayed in Table 1. Among the 19 patients, 12 were classified as HFDC, 3 as MFDC, and the remaining were as LFDC. The IHC results were plotted in Figure 7A. The LFDC group had the highest expression of MMP1. The MFDC group showed the highest expression of FOXC2. The HFDC group exhibited the highest expressions of KLK6 and TPSB2.

Discussion section (Page 21, Line 607-616;):

However, this study has limitations. **MCs and endothelial cells have also been reported to be closely associated with keloid formation and activity[41,42]. However, in this study, considering easier clinical translation, keloid classification is only carried out based on the differentiation of fibroblasts.** Besides, due to biases in pure bioinformatics analysis, it cannot fully generalize fibroblast lineagetrajectories and cell differentiation states. Further cellular and animal experiments are needed to validate the correlation between identified DEFDRGs and phenotypes and to verify the clinical relevance of the new keloid classification. Furthermore, due to the lack of sequencing data related to keloid prognosis, especially recurrence, in public databases, this keloid classification is only for interpretation of keloid development and cannot predict the prognosis, which partially limits the clinical application of this classification.

Comment 4: The authors state: "Furthermore, it was confirmed that FIBKC was significantly correlated with patient disease diagnosis and diagnostic time by chi-square test ($P < 0.001$)

(Figure 4D-E)." The significance of "diagnostic time" and the meaning of Day 0 and Day 42 require clarification in Figure 4D-E.

Response: We acknowledge the reviewer's point and have revised the manuscript to address the concern. The main objective of our study was to identify transcriptomic changes during keloid formation, and the dataset (SRA ID: SRP13707) , which included temporal data of keloid, was selected. Biopsies from keloid-prone, unaffected family members from families with a history of autosomal dominant keloid formation and from healthy cases without any predisposition to keloid formation were taken in day0 and day42. Forty-two days later, the original sampling site was continued (day 42). The integrated analysis of the dataset can help us to identify the key genes involved in the formation of keloids, which is more meaningful for the understanding of the biological process of keloids. We are sorry that we did not elaborate on the "diagnostic time", This mainly represents the sampling time point after injury for keloid-prone patients. To avoid confusion, we changed it to “keloid time” in revised manuscript. Combining with your suggestion, we provided further explanations and descriptions of Figure 4D-E in previous manuscript (Figure 4B-C in revised manuscript) and expounded on their significance.

Corresponding revisions in the manuscript are as follows:

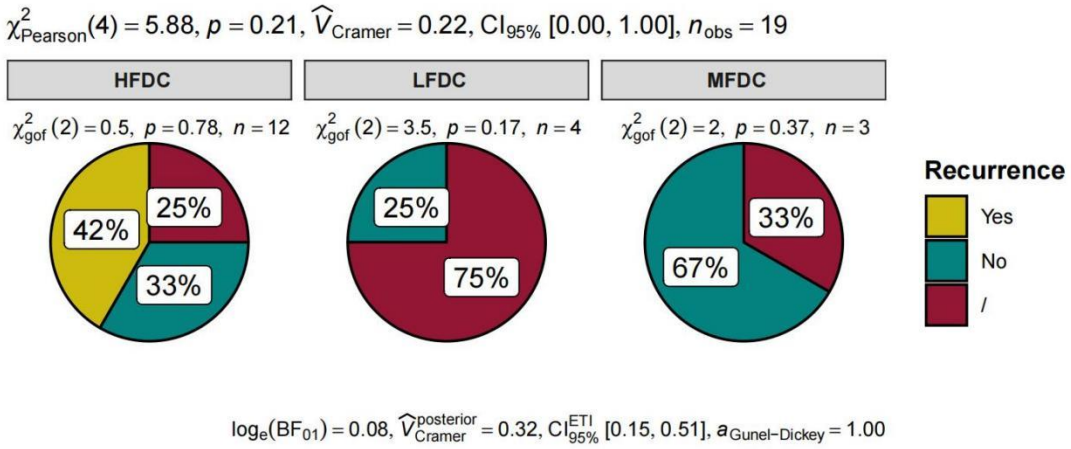
Result section (Page 13, Line 368-372)

Furthermore, it was confirmed that FDBKC was significantly correlated with keloid diagnosis and time by chi-square test ($P < 0.001$) (Figure 4B-C). Keloid time represented day 0 and day 42 after injury in keloid-prone patients. The stacking bar chart showed that healthy samples were mainly in cluster 1 and keloid samples (including day 0 and day 42) were mostly in cluster 3, indicating that cluster 3 has a more severe pathological state.

Comment 5: While fibroblast-related genes may correlate with keloid characteristics, predicting treatment responses is crucial. Without indicators of treatment response or post-surgical recurrence, discussing high risk/low risk is challenging, making it difficult to assert that HFIC correlates with a worse prognosis.

Response: Thank you for your valuable suggestions. Currently, the lack of sequencing data related to keloid recurrence in public databases restricted our analysis. As a result, our study has limitations in predicting the treatment effect and recurrence prognosis of keloids. Therefore, this study primarily focuses on elaborating on the occurrence and development of keloids to provide insights into the formation process of keloids. In addition, during the clinical relevance verification of the keloid classification, we attempted to analyze the recurrence data. However, despite intense efforts, it’s hard for all patients to return visit during the pandemic. For the limited enrolled patients (n=19) and high percentage of patients withdrawn, the results of the clinical relevance of keloid recurrence and keloid classification had no statistical difference **(Response letter Figure 3)**. With your suggestion, we have described this limitation in the discussion. In addition, to better reflect the close correlation between the novel keloid classification and the fibroblast differentiation state, we have renamed the classification as FDBKC (Fibroblast Differentiation-based Keloid Classification), which is divided into LFDC, MFDC, and HFDC to correspond to LFIC, MFIC, and HFIC. Modifications and replacements are made throughout the text. At the same time, to further

verify the difference and importance of our 25 key genes, we performed a differential gene expression analysis of the 25 differentiation-related genes on other datasets (GSE7890; GSE145725; GSE188952). Besides, protein sequencing was also conducted to identify the protein expression differences of 25 key genes. The above results are shown in the **Figure S9 C-F and Supplementary table 2** in revised manuscript.



Response letter Figure 3 The pie charts with Chi-square test revealed no significant differences in the recurrence among HFDC, LFDC, and MFDC, three cluster in the novel keloid classification. The results may be caused by limited enrolled patients (n=19) and high percentage of patients withdrawn.

Corresponding revisions in the manuscript are as follows:

Abstract section (Page 1, Line 3-15)

Keloids are dermal fibroproliferative skin disorders caused by abnormal wound healing, resulting in impaired skin function and aesthetic defects. Abnormal fibroblast proliferation and excessive collagen deposition are involved in keloid formation. This study investigated the role of fibroblast differentiation in the development of keloid. Single-cell and bulk RNA sequencing data of keloids were comprehensively analyzed and 25 clinically relevant fibroblast-differentiation-related genes were finally identified. A novel keloid diagnostic classification, including LFDC, MFDC, and HFDC, was constructed, indicating that this set of genes could serve as new therapeutic targets for keloids. Besides, multiple microarray data, protein sequencing data and immunohistochemistry of key markers in clinical keloid samples were used for further verification. **In conclusion, this study constructed keloid molecular classification based on fibroblast differentiation for the first time, contributing to the further understanding of the keloid pathogenesis, which might provide new insights for keloid diagnosis and treatment.**

Result section (Page 12-13, Line 356-372; Page 17, Line 481-489)

3.6. Construction FDBKC by Consistent Clustering Based on DEFDRGs

We constructed FDBKC based on the expression levels of 25 DEFDRGs through a consistent clustering algorithm, which was divided into three clusters (Figure S9A). The delta-area diagram showed changes in the area under the Cumulative Distribution Function (CDF) curve with different *k* values. The *P* value of *k* = 3 is the smallest, indicating that *k* = 3 has the best clustering

statistic expression (Figure S9B). We illustrated the expression of 25 key DEFDRGs in various dimensions, including consensus cluster, state, sex, age, and diagnosis (Figure 4A). Gene expression levels in three consensus clusters were relevant to clinicopathological characteristics. Strikingly, strong connections between fibroblast state and consensus clusters were observed. The overall expression levels of fibroblast fate determinant genes were higher in cluster 1, and lower in cluster 2 and cluster 3 with different expression patterns, demonstrating obvious heterogeneity in the obtained clusters (Figure 4A).

Furthermore, it was confirmed that FDBKC was significantly correlated with keloid diagnosis and time by chi-square test ($P < 0.001$) (Figure 4B-C). Keloid time represented day 0 and day 42 after injury in keloid-prone patients. The stacking bar chart showed that healthy samples were mainly in cluster 1 and keloid samples (including day 0 and day 42) were mostly in cluster 3, indicating that cluster 3 has a more severe pathological state.

3.11 Pathological manifestations of different clinical subtypes of keloids

According to the clinical manifestations of patients with different subtypes, the average VSS score in the HFDC group was significantly higher than that in the other two groups (Figure 7B). Keloids in the HFDC group were observed to be thicker than those in the other two groups, nearly 83% greater than 2 mm (Figure 7C). Furthermore, the keloids in the HFDC group were also found to be harder, possibly due to the abnormal distribution of fibroblasts and collagen (Figure 7D). The proportions of keloids with pain, itching, and abnormal blood vessels were relatively high in the HFDC group (Figure 7E-G), reflecting the more severe clinical manifestations of keloids in the HFDC group. These findings were consistent with the results of sc-RNA seq and bulk-RNA seq analyses and proved the efficacy of our novel keloid classification.

Discussion section (Page 21, Line 607-616):

However, this study has limitations. MCs and endothelial cells have also been reported to be closely associated with keloid formation and activity[41,42]. However, in this study, considering easier clinical translation, keloid classification is only carried out based on the differentiation of fibroblasts. Besides, due to biases in pure bioinformatics analysis, it cannot fully generalize fibroblast lineagetrajectories and cell differentiation states. Further cellular and animal experiments are needed to validate the correlation between identified DEFDRGs and phenotypes and to verify the clinical relevance of the new keloid classification. **Furthermore, due to the lack of sequencing data related to keloid prognosis, especially recurrence, in public databases, this keloid classification is only for interpretation of keloid development and cannot predict the prognosis, which partially limits the clinical application of this classification.**

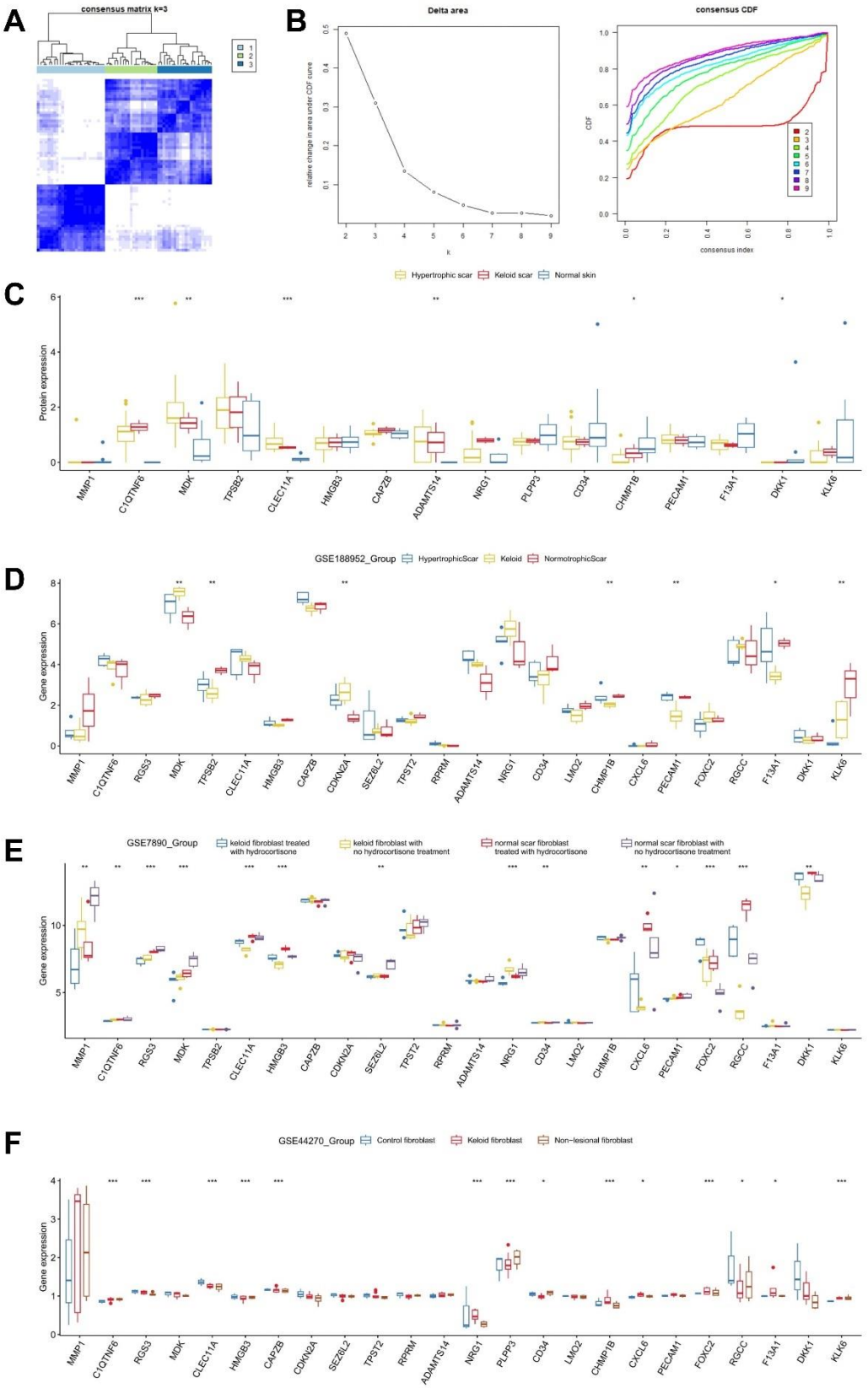


Figure S9 in revised manuscript Consensus clustering and verification in public databases (related to Figure 4). (A)Consensus clustering analysis based on 25 DEFDRGs. (B)The clustering effect is more specific when k = 3. (C)The differential expression of proteins encoded by 25 DEFDRGs in normal skin, hypertrophic scar, and keloid. (D-F) The

verification of the differential expression of 25 DEFDRGs using sequencing data in public databases. These three datasets found that many genes among these 25 DEFDRGs had statistical differences.

GEO ID	Sequencing type	Sample number	Sample information	Detected gene of 25 DEFDRGs	DEG Number	Proportion of DEG	Reference
GSE188952	RNA sequencing	12	tissues of normotrophic scars, hypertrophic scars, and keloids	24	7	29.2%	/
GSE7890	Microarray	19	fibroblasts of normal scars and keloids	24	14	58.3%	PMID: 17989729
GSE44270	Microarray	32	fibroblasts of keloid scars, adjacent non-keloid skins, and normal skins	24	14	58.3%	PMID: 23815228

Supplementary table 2 in manuscript Data from public databases were used to verify the differential expression changes of 25 DEFDRGs in keloids, normal scars and normal skin.

Comment 6: Claiming Ouabain as a potential inhibitor without supporting cell experiments lacks substantiation and should be approached with caution.

Response: Apologies for any misunderstanding caused by our inappropriate expression in the article. We have revised the presentation in the article. We obtained possible targeted drugs only by the prediction of Camp database, which was a prediction to be tested.

Corresponding revisions in the manuscript are as follows:

Abstract section (Page 1, Line 3-15)

Keloids are dermal fibroproliferative skin disorders caused by abnormal wound healing, resulting in impaired skin function and aesthetic defects. Abnormal fibroblast proliferation and excessive collagen deposition are involved in keloid formation. This study investigated the role of fibroblast differentiation in the development of keloid. Single-cell and bulk RNA sequencing data of keloids were comprehensively analyzed and 25 clinically relevant fibroblast-differentiation-related genes were finally identified. A novel keloid diagnostic classification, including LFDC, MFDC, and HFDC, was constructed, indicating that this set of genes could serve as new therapeutic targets for keloids. Besides, multiple microarray data, protein sequencing data and immunohistochemistry of key markers in clinical keloid samples were used for further verification. In conclusion, this study constructed keloid molecular classification based on fibroblast differentiation for the first time, contributing to the further understanding of the keloid pathogenesis, which might provide new insights for keloid diagnosis and treatment.

Discussion section (Page 20, Line 584-596)

So far, there is a lack of precise treatment targeting keloid pathophysiology molecular mechanisms. Computational in silico analyses were widely used in developing and evaluating pharmacological hypotheses and are suitable for identifying and optimizing novel small molecules with unique affinity for specific targets[84]. Based on comprehensive Camp analysis, we identified the specific inhibitor for FIBKC, especially for HFDC, which may reverse related gene expression alterations. Among the predicted top 10 inhibitors of HFDC, Ouabain may be a drug worthy of future research. As a selective inhibitor of Na, K-ATPase, Ouabain is mainly used for treating heart failure clinically. Studies verified Ouabain also showed anti-fibrotic effects. Ouabain can ameliorate pulmonary fibrosis by inhibiting TGF-β-induced fibrotic signaling, myofibroblast

differentiation, and smooth muscle α-actin (α-SMA), collagen-1, and fibronectin expression[85]. Besides, Ouabain could also improve bleomycin-induced pulmonary fibrosis by directly inhibiting the proliferation and promoting the apoptosis of lung fibroblasts[86]. Therefore, assessing the efficacy and safety of Ouabain would be a future direction of this research.

Minor

Comment 7: In Figure 5D, it's challenging to distinguish between male and female colors. Additionally, it's advisable to use "sex" instead of "gender."

Response: Thank you for your comments. We have changed "gender" to "sex". At the same time, we have unified and modified the “male” and “female” colors in the manuscript and figure, cyan for male and purple for female.

Comment 8: The transition from Cluster 1/Cluster 2/Cluster 3 to LFIC/MFIC/HFIC in the figure is confusing. Consider adding parentheses next to LFIC and labeling it as Cluster 1 in the figure.

Response: We accept the reviewer's suggestion and have revised the manuscript accordingly. In addition, to better reflect the close correlation between the novel keloid classification and the fibroblast differentiation state, we have renamed the classification as FDBKC (Fibroblast Differentiation-based Keloid Classification), which is divided into LFDC, MFDC, and HFDC to correspond to LFIC, MFIC, and HFIC. Besides, modifications and replacements are made throughout the text. According to your suggestion, we have added (Cluster 1)/(Cluster 2)/(Cluster 3) next to LFDC/MFDC/HFDC in the figure (Figure 4, Figure 6, and Figure 7), respectively.

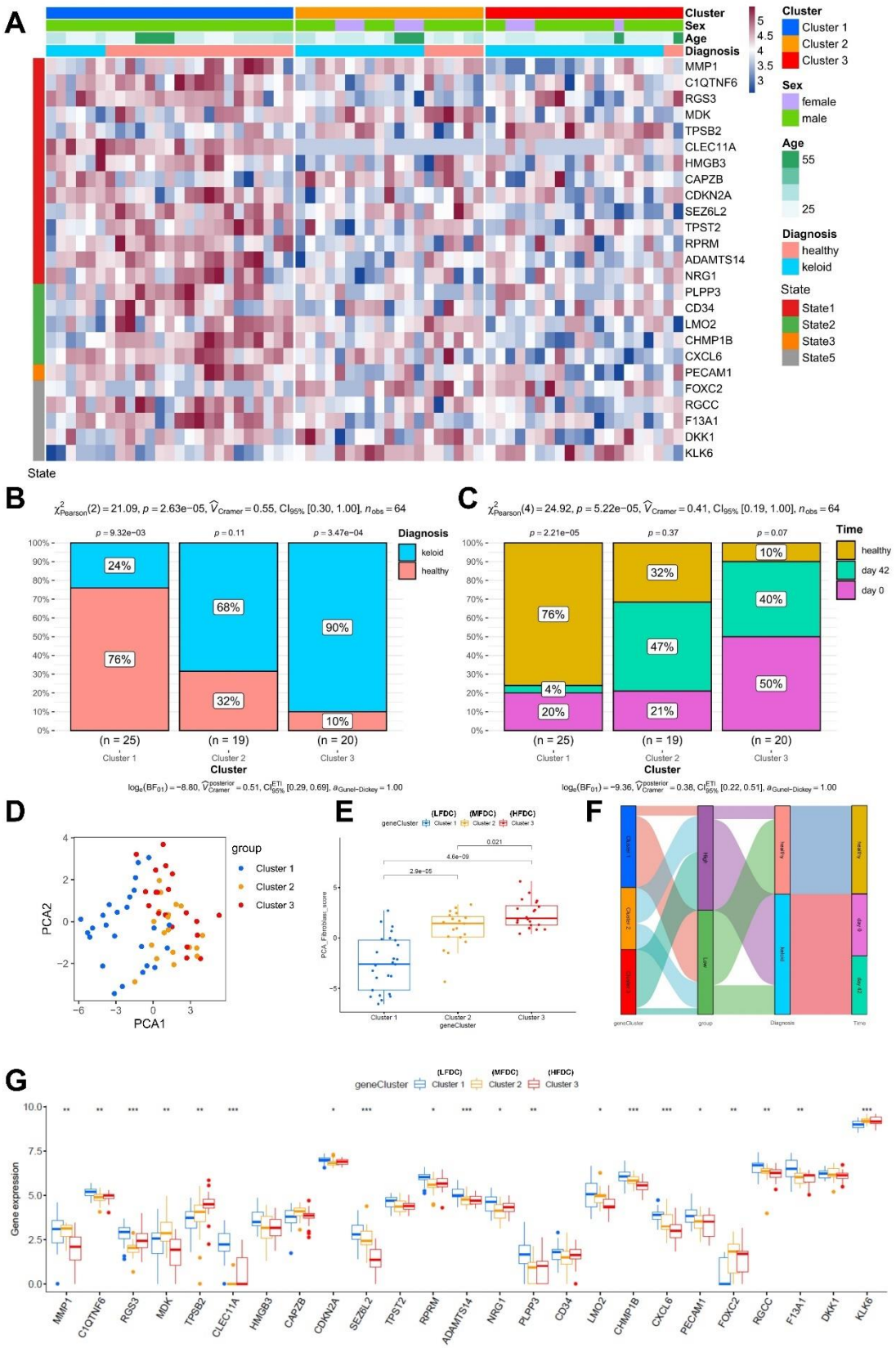


Figure 4 in revised manuscript Consensus clustering and clinical correlation analysis. (A) The heat map illustrated the expression of 25 DEFDRGs in each sample of three clusters. (B) The proportions of keloid and health diagnoses varied among 3 clusters. (C) The proportions of healthy samples and different keloid times varied among 3 clusters. (D) PCA score plot

based on DEFDRG expression. (E) Different PCA fibroblast score of 3 clusters. The score of Cluster 1 was the lowest, and the score of Cluster 3 was the highest. (F) Proportions and relationships of high and low fibroblast score with consensus cluster and keloid diagnosis and time. (G) The differential expression of 25 DEFDRGs in 3 clusters. Many genes were highly expressed in cluster 1, while TPSB2 and KLK6 were highly expressed in cluster 3.

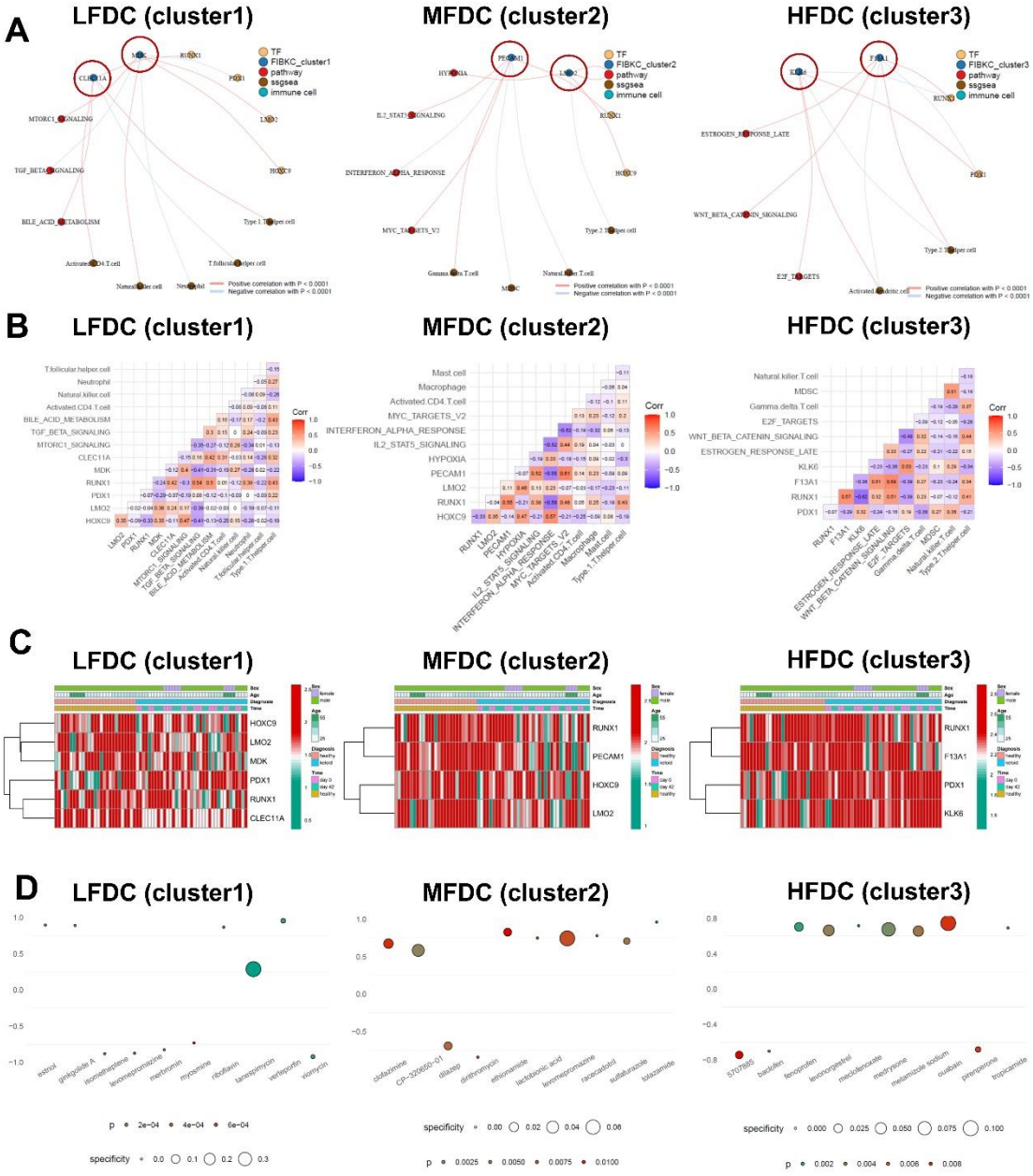


Figure 6 in revised manuscript Construction of regulatory networks of different keloid classifications and identification of the potential drugs. (A) The subtype-specific and DEFDRG-centered regulatory network containing upstream DETFs, downstream signaling pathways, and potential regulation abilities of immune cells/immune functions in LFDC, MFDC, and HFDC. (B) The co-expression patterns and correlations among all components within the regulation network in LFDC, MFDC, and HFDC. (C) Expression levels of key DEFDRGs and DETFs in the regulation network of LFDC, MFDC, and HFDC. (D) Inhibitors

identified by Cmap analysis that may target key DEFDRGs and DETFs in LFDC, MFDC, and HFDC.

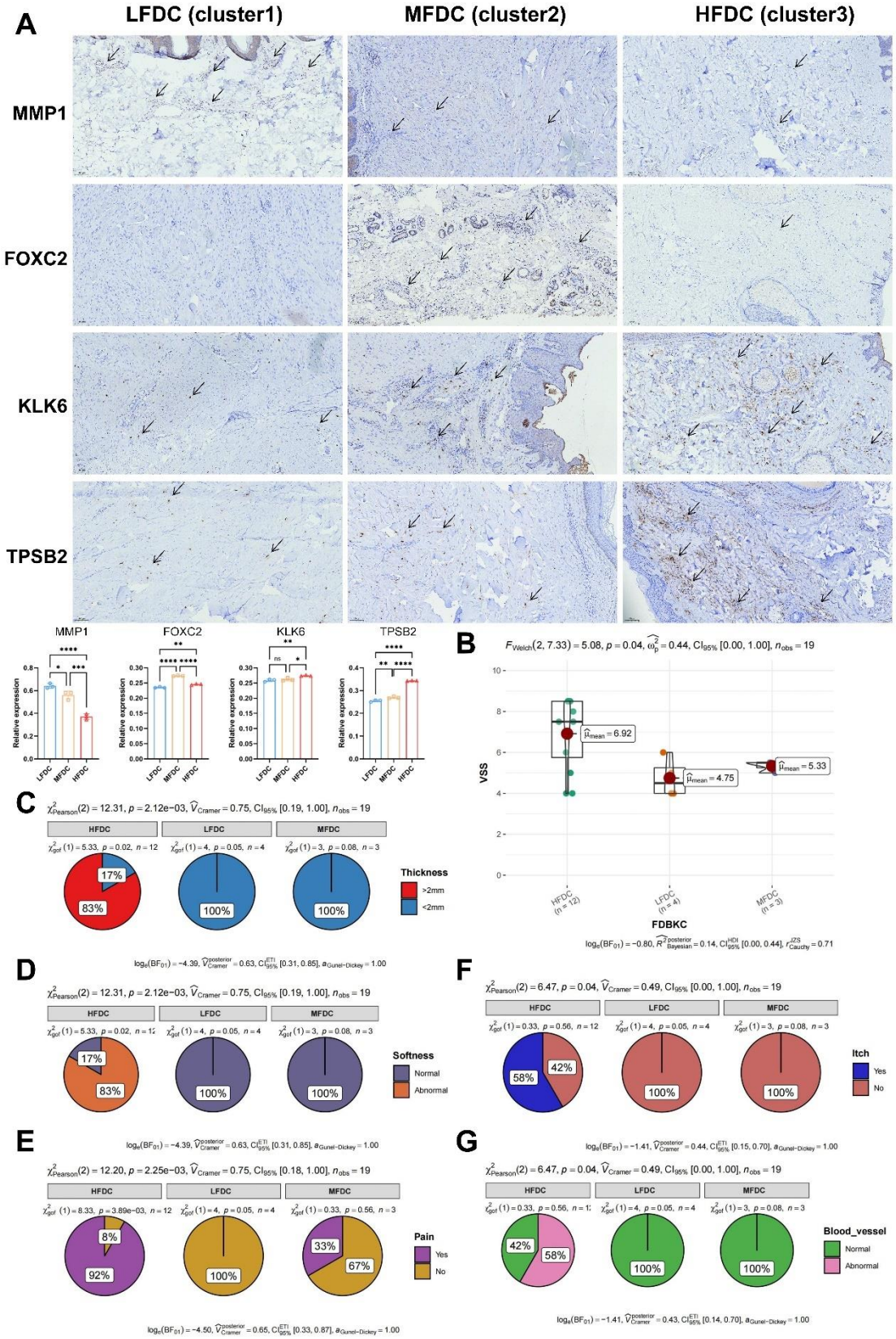


Figure 7 in revised manuscript IHC results of clinical samples and clinical correlation analysis. (A) The IHC image and score of keloid classification specific marker in clinical

keloid samples (n = 3). (B) The correlation of novel keloid classification and VSS score. (C) The correlation of novel keloid classification and keloid thickness. (D) The correlation of novel keloid classification and keloid softness. (E) The correlation of novel keloid classification and pain. (F) The correlation of novel keloid classification and itch. (G) The correlation of novel keloid classification and blood vessels.

Comment 9: Considering the flow of the paper, it might be beneficial to promote Figure S6A to the main figure to enhance clarity and follow the paper's progression.
Response: Thank you for your comments. They are of great help to the logic of our article. We adjusted Figure S6A in previous manuscript to the main figure (Figure 3F in revised manuscript). In addition, we have added new verification data to verify the reliability of our classification, which is in the appendix of the article (Figure S7-9).

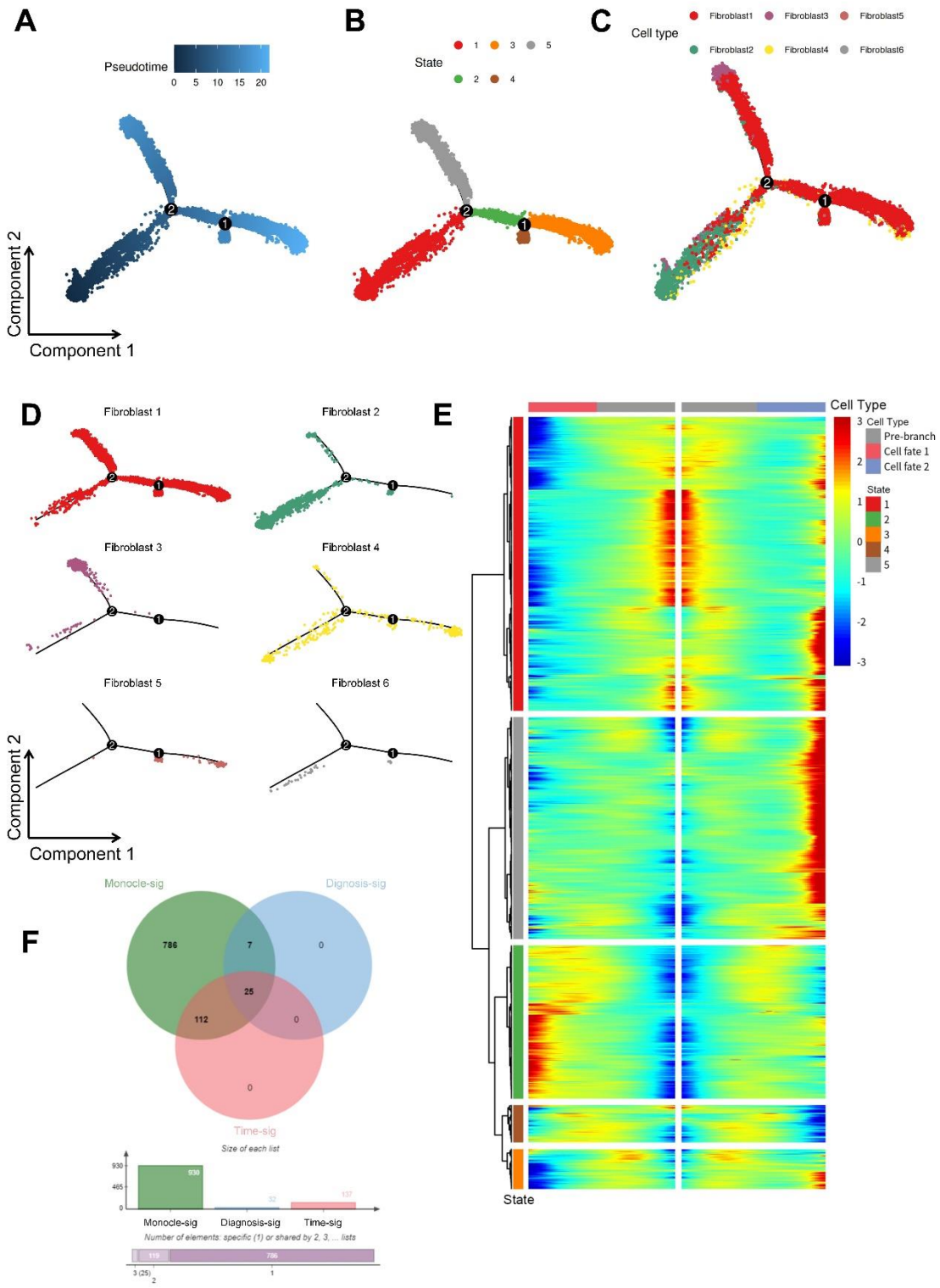


Figure 3 in revised manuscript Pseudotime analysis and identification of DEFDRGs. (A) The pseudotime plot showed the fibroblast differentiation sequence. (B) Detection of five differentiation states. (C-D) Display of fibroblast subtypes on different differentiation trajectories. (E) The tendency to different differentiation directions of different states. (F) Identification of DEFDRGs. DEFDRGs were obtained by taking the intersection of the

fibroblast differentiation-related DEGs and the DEGs related to clinical keloid diagnosis and time.

No.: COMMSBIO-24-1782A

Manuscript title: Uncovering a Novel Fibroblast Differentiation-Based Keloid Classification by Integration of Single-cell and Bulk RNA Sequencing

Communications Biology

21st February 2025

Dear reviewers

The interactive review started on: 21st November 2024

We are profoundly grateful for the continued guidance from the reviewers. The newly raised suggestions demonstrate remarkable academic consensus, with particularly insightful advice regarding our scRNA-seq analysis. In response to these expert recommendations, we have rigorously re-analyzed the scRNA-seq data using the Harmony algorithm. Corresponding content has been updated in figures and manuscript.

After careful revision, we resubmit the manuscript for your consideration. All revised parts are highlighted in red in the manuscript. Additionally, the point-by-point responses are appended to this letter.

We hope that the revision is acceptable. We look forward to hearing from you soon.

Sincerely yours,

Runzhi Huang, Zhaofan Xia, Shizhao Ji

21st February 2025

Reviewer #2:

Comment 1: The HPA database identifies PECAM1 as an endothelial cell marker, and its apparent expression in fibroblasts may be due to doublets or ambient RNA contamination. Figure S2 suggests doublets in the data, likely due to single-cell analysis limitations. The claim that PECAM1 is expressed in skin fibroblasts based solely on database results lacks strong support. Co-staining PECAM1 with fibroblast-specific markers would be essential to confirm this expression.

Response:

We sincerely appreciate the reviewer's examination and highly valuable feedback. Indeed, as you pointed out, PECAM1 is a classic marker of endothelial cells, and merely verifying its expression in fibroblasts through the HPA database is insufficient. In accordance with your suggestion, we have revised the analysis method for the scRNA-seq data. We re-analyzed the scRNA-seq data using the Harmony algorithm and also employed DoubletFinder to identify doublets during the analysis process **(Figure 1 of the response letter)**. The results showed that the proportion of doublets in the data was very low, reflecting the high data quality. We have updated corresponding figures and text in the manuscript based on the results of the Harmony analysis **(Figure 1, 2, and 3; Figure S2 and S3 in the revised manuscript)**. In the new analysis results, there is no doublet contamination in either the analysis of major cell types or the analysis of fibroblast subsets. Furthermore, based on the new scRNA-seq analysis, our study obtained 25 DEFDRGs, including PECAM1, which is consistent with the previous data. Meanwhile, we also performed immunofluorescence co-staining of PECAM1, along with the classic fibroblast markers COL1 and VIM, on human primary dermal fibroblasts to clarify whether PECAM1 is expressed in fibroblasts. The immunofluorescence results confirmed that PECAM1 is expressed in some fibroblasts, albeit in a relatively small number **(Figure 2 of the response letter)**. The antibodies used were VIM (#5741, CST), COL1 (ab138492, abcam), and PECAM1 (ab182981, abcam).

Upon further literature review, we discovered that fibroblasts and endothelial cells are closely related. Dermal fibroblasts can be directly reprogrammed into endothelial - like

cells under the influence of microRNAs [1]. In this study, vasculogenic subset with a distinct fibroblast transcriptome was identified through scRNA-seq and its angiogenic capacity and the expression of the endothelial cell marker PECAM1 were both demonstrated *in vitro* and *in vivo*. Additionally, another study identified a subset of fibroblasts expressing endothelial cell genes in a hind-limb ischemia model, which contributed to angiogenesis [2]. All the abovementioned evidence suggests that it may be possible for fibroblasts to express endothelial markers.

[1] Pal D, Ghatak S, Singh K, et al. Identification of a physiologic vasculogenic fibroblast state to achieve tissue repair. Nat Commun. 2023 Feb 28;14(1):1129.

[2] Meng S, Lv J, Chanda PK, et al. Reservoir of Fibroblasts Promotes Recovery From Limb Ischemia. Circulation. 2020 Oct 27;142(17):1647-1662.

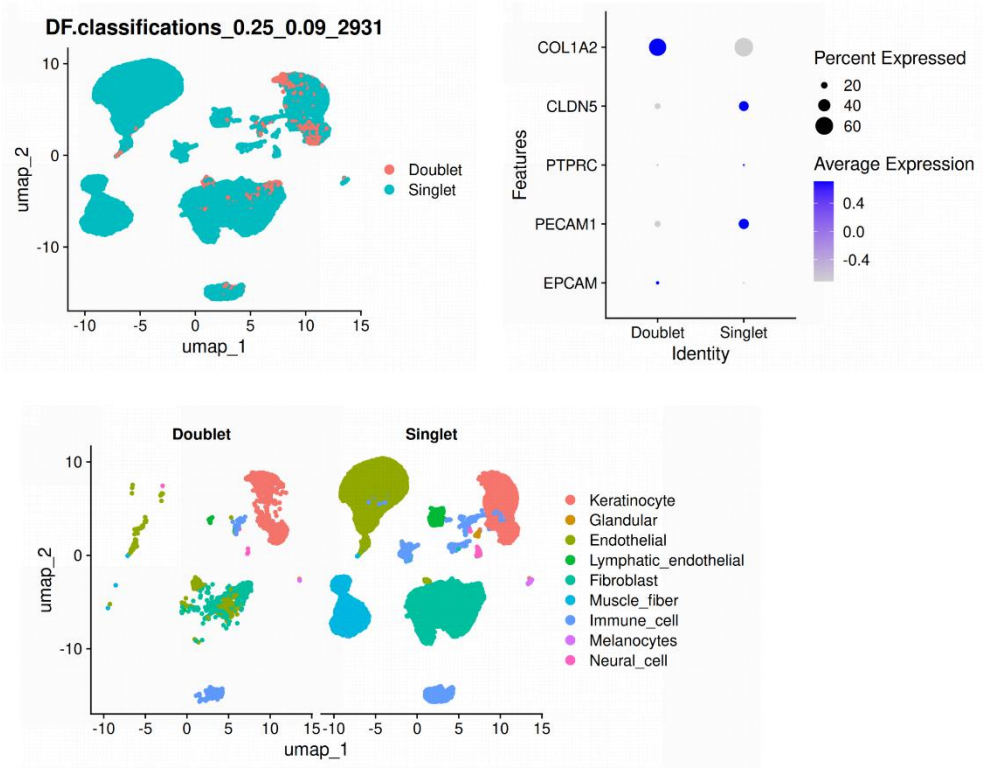


Figure 1 of the response letter. Doublets in scRNA-seq of GSE163973 were identified. The proportion of doublets was very low, reflecting the high quality of the data.

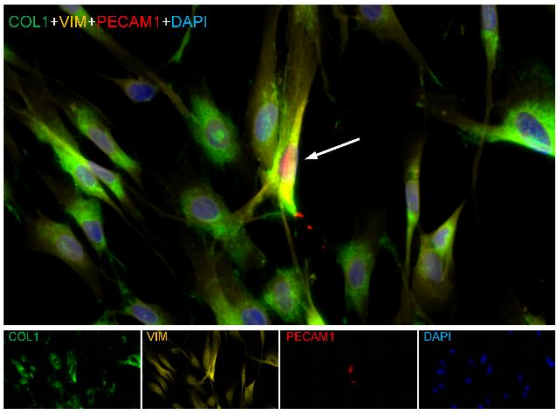


Figure 2 of the response letter. Immunofluorescence co-staining of COL1, VIM, PECAM1, and DAPI in human primary dermal fibroblasts.

Corresponding revisions in the manuscript are as follows：

Method section (Page 3-4, Line 88-99):

2.2 ScRNA-seq Data Processing

Then, the quantified gene expression matrices were further analyzed with the Seurat pipeline (Version: 3.2.2)[19]. Only cells with transcript expression greater than 100,000, mitochondrial gene proportion less than 10%, and genes expressed in at least three cells could be included for subsequent analyses. Following quality control (QC) process, keloid and normal scar samples were merged into a single Seurat object. And then, the Seurat object was scaled and standardized by the "ScaleData" function. Besides, we determined the top 2,000 most variable genes using the "vst" method, which were defined as highly variable genes (HVG). Principal component analysis (PCA) was performed on these HVGs to reduce dimensionality. To mitigate batch effects between samples, harmonization was conducted using the Harmony algorithm with default parameters, integrating the data across samples while preserving biological variability[19]. DoubletFinder was utilized to detect doublets during the analytical process. The first twenty harmonized principal components (PCs) were subsequently utilized for Uniform Manifold Approximation and Projection (UMAP) analysis.

Result section (Page 9-11, Line 256-315):

3.1 Cell Type Annotation

After stringent QC filtering, a total of 44,463 single cells (keloids: 23,638; normal scars: 20,825) from scRNA-seq data met the criteria and were included for downstream analyses. The experimental workflow is detailed in Figure 1A and Figure S1. UMAP-based clustering identified 18 distinct cell

clusters, which were annotated into 9 major cell types: endothelial cells, fibroblasts, glandular cells, immune cells, keratinocytes, lymphatic endothelial cells, melanocytes, smooth muscle cells, and neural cells (Figure 1B). Spatial distribution patterns of these cell populations between keloid and normal scar tissues are visualized in Figure 1B. Feature plots demonstrate the expression levels and cellular distribution of canonical biomarkers retrieved from the CellMarker database across all annotated cell types (Figure S2A). Figure S2B further validates the cellular annotation accuracy through marker gene expression patterns. The Cleveland dot plot was used to describe expression levels of typical marker genes (*DCN*, *KRT1*, *PTPRC*, and *VWF*) and the proportions of the 9 cell types in different samples (Figure S2C). Figure 1C illustrates the average number and proportion of each cell type. The results indicate that endothelial cells and fibroblasts are predominant across all samples, which is in line with previous research[40].

3.2 Cell Communication Analysis

Cell communication analysis demonstrated fibroblasts as the predominant cellular interactors, exhibiting the highest communication frequency across all cell types (Figures 1D-1E). Quantitative evaluation revealed significantly enhanced global communication activity in keloids compared to normal scar tissues (Figure 1F). Notably, fibroblasts exhibit the strongest outgoing interaction strength in both keloids and normal scars (Figure S2D).

3.3 Fibroblast Subtype Analysis Revealed Cell Heterogeneity

A total of 13,925 fibroblasts were extracted only for further analysis. Through unsupervised clustering, 6 distinct clusters were identified and annotated as fibroblast 1-6 (Figure 2A). Among them, fibroblast 1 and 2 predominated in normal scars, while fibroblast 3 occupied most of the keloids (Figure 2B). The expression levels of typical marker genes (*ACTA2*, *COL1A1*, *TGFBI*, and *TOP2A*) were used to prove the accuracy of the fibroblast screening, and the proportion of each fibroblast cluster in keloid and normal scars was visualized by the bar plot, also proving that fibroblast 3 is mainly found in keloids (Figure 2C). Fibroblast 3 highly expresses *POSTN*, *ASPN*, and *COMP*, which have been reported to be involved in keloid formation (Figure 2D)[14]. Extensive cell communication existed among different fibroblast subtypes. Particularly, fibroblast 3 interacted mostly with others, and COL1A1 and COL1A2 were key intercellular association genes (Figure 2E-F). Enriched hallmark signaling pathways in fibroblast 3 were collagen formation and ECM receptor

interaction (Figure S3A). Figure S3B-C demonstrate distinct cell cycle phases among fibroblast subpopulations.

3.4. Reveal Differentiation Fate Decisions and Expression Dynamics of Keloid Fibroblasts via Trajectory Analysis

To delineate the differentiation hierarchy of keloid fibroblasts, we employed Monocle 2 for single-cell pseudotemporal ordering analysis. This computational framework reconstructed lineage trajectories through dimensionality reduction (DDRTree algorithm) and RNA velocity estimation, revealing 7 distinct cellular states along fibroblast maturation paths (Figure 3A-C). Furthermore, the pseudotime and cell lineage trajectory of each fibroblast cluster were illustrated separately in Figure 3D. Notably, fibroblast 3 predominantly occupied intermediate-late pseudotime positions (States 3-6). Through trajectory analysis, DEGs associated with fibroblast differentiation were identified and named as FDRGs. Top 100 FDRGs were shown in the heatmap (Figure 3E).

3.5. Identification of Keloid DEFDRGs

Through the aforementioned analyses, we mapped the transcriptomes of all fibroblast states and used Monocle 2 to identify 14,323 FDRGs (Monocle-sig, $P < 0.05$), which are the genes determining cell fate. Subsequent integration of RNA-seq profiles with clinical metadata from the SAR database through univariate Cox regression revealed 32 diagnosis-associated (Diagnosis-sig, $P < 0.05$) and 137 time-significant (Time-sig, $P < 0.05$) DEGs. Intersectional analysis of these three gene sets identified 25 core FDRGs, named DEFDRGs, exhibiting both differentiation-dependent expression patterns and clinical relevance (Figure 3F and S4A). To elucidate clinical correlations, a co-expressive network of DEFDRGs was constructed using nonparametric tests and Pearson correlation (Figure S4B). Labels of various colors meant genes for different monocle states. Curves in different colors represented positive or negative correlations. The DEFDRGs had tight associations, indicating underlying regulation mechanisms of clinical outcomes and fibroblast fates. Most genes were positively correlated, while *Kallikrein Related Peptidase 6 (KLK6)* and *Tryptase Beta 2 (TPSB2)* were mostly negatively correlated with the others. Boxplot validation confirmed differential expression patterns across diagnosis and time (Figure S4C).

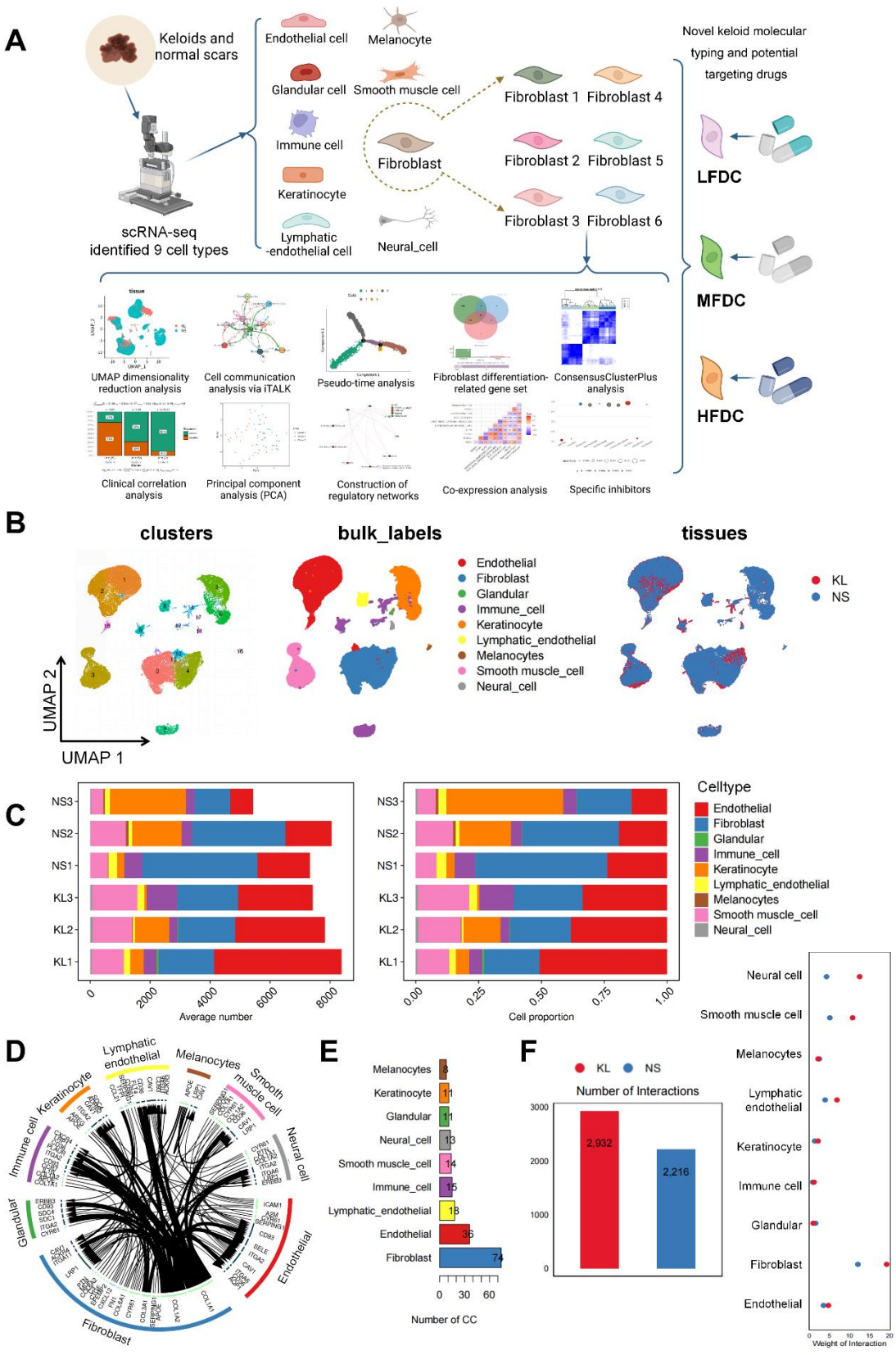


Figure 1 in revised manuscript: Comprehensive analysis of scRNA-seq revealed the heterogeneity of keloids and normal scars. (A) Schematic diagram of workflow in study design and analysis. (B) UMAP dimensionality reduction analysis clustering

44,463 cells into nine cell types and showing the distribution of the above cells in keloids and normal scars. (C) The average number and proportion of nine cell types in each sample. Fibroblasts and endothelial cells account for a relatively high proportion. (D-E) Cell communications of each cell type. Fibroblasts are involved in the largest number of cell communications. (F) The left figure shows the number of interactions in keloids and normal scars, and the right figure shows the weight of interactions of each cell type in the two groups.

Abbreviations: KL: Keloid; NS: Normal skin; CC: Cell Communication

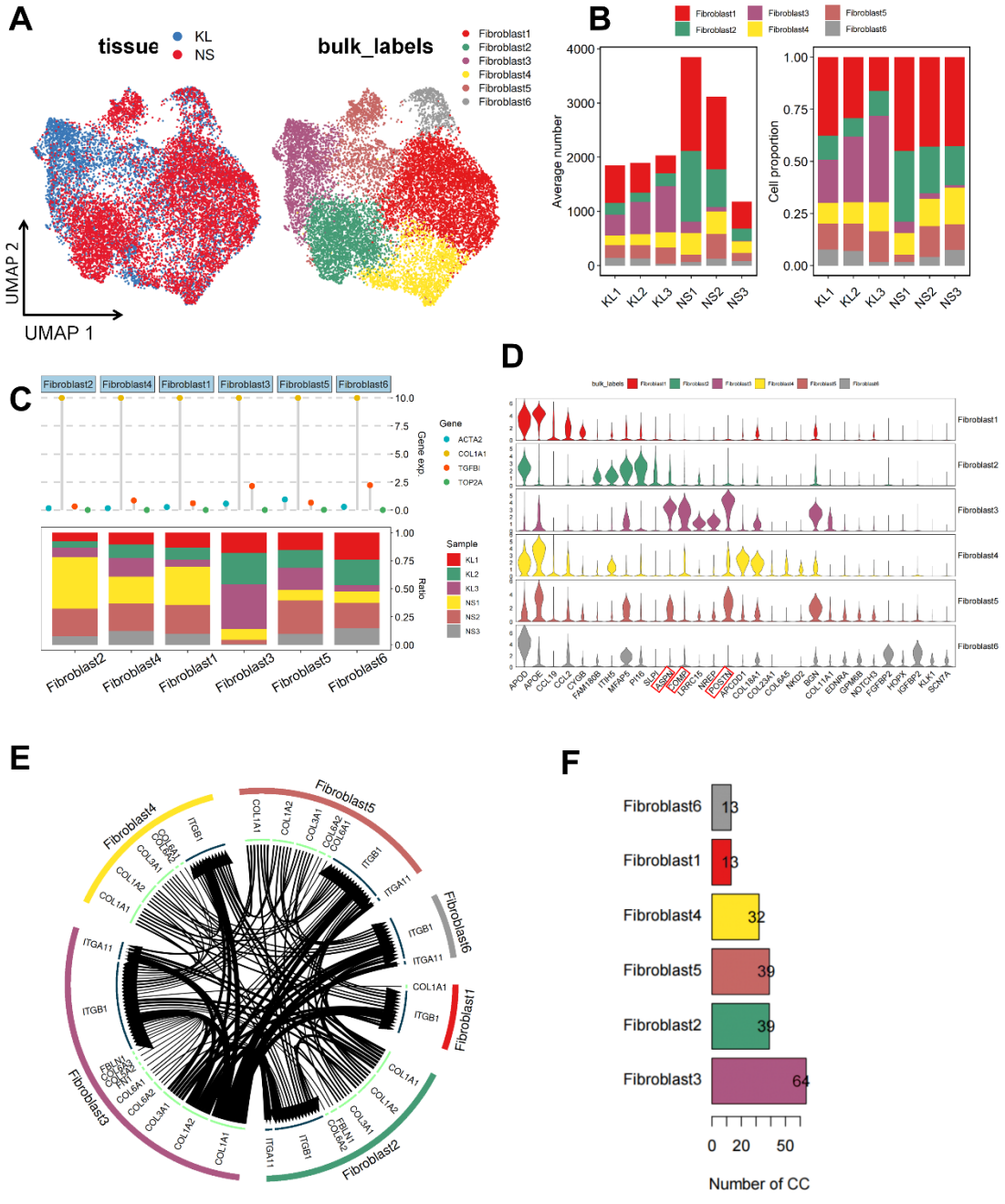


Figure 2 in revised manuscript: Identification of fibroblast subtypes and cell communication. (A) Six fibroblast subtypes were identified by UMAP dimensionality reduction analysis. (B) The histogram shows the average number and proportion of fibroblast subtypes in different samples. Fibroblast 3 predominates in keloids, whereas fibroblast 2 dominates normal scars. (C) The Cleveland dot plot shows 4 classical markers' expression (ACTA2, COL1A1, TGFBI, and TOP2A) in fibroblasts. The bar plot displays the proportion of each fibroblast cluster in 6 samples (3 keloids and 3 normal scars). (D) The expression of marker genes in each fibroblast subtype. ASPN, COMP, and POSTN are highly expressed in fibroblast 3. (E-F) Cell communication between fibroblast subtypes. Fibroblast 3 takes part in a greater number of cell communications.

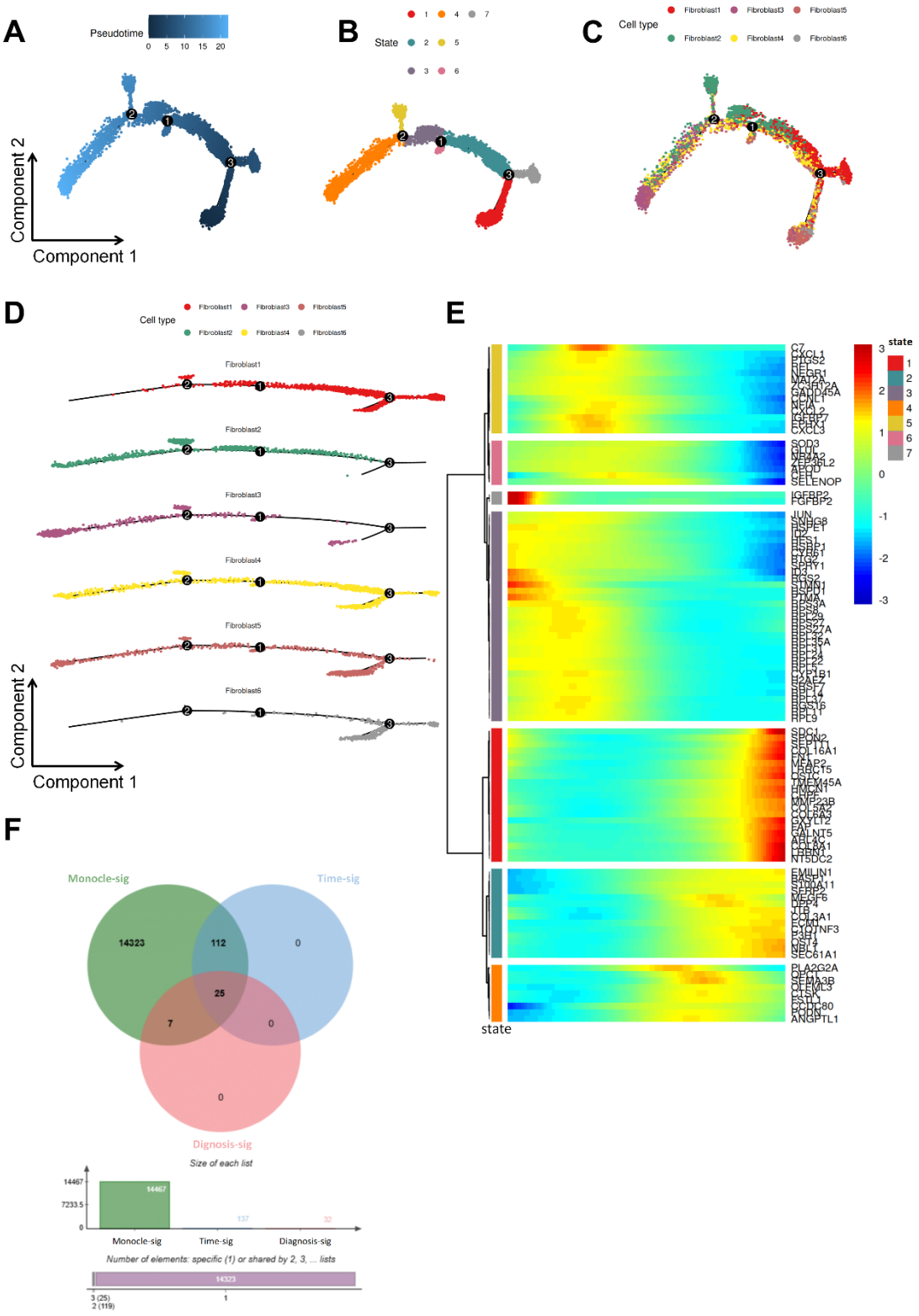


Figure 3 in revised manuscript: Pseudotime analysis and identification of DEFDRGs. (A) The pseudotime plot showed the potential fibroblast differentiation sequence. (B) Detection of 7 differentiation states. (C-D) Display of fibroblast subtypes on different differentiation trajectories. (E) The expression of the top 100 representative

fate-related genes differentially expressed across. (F) Identification of DEFDRGs. DEFDRGs were obtained by taking the intersection of the fibroblast differentiation-related DEGs and the DEGs related to clinical diagnosis and time.

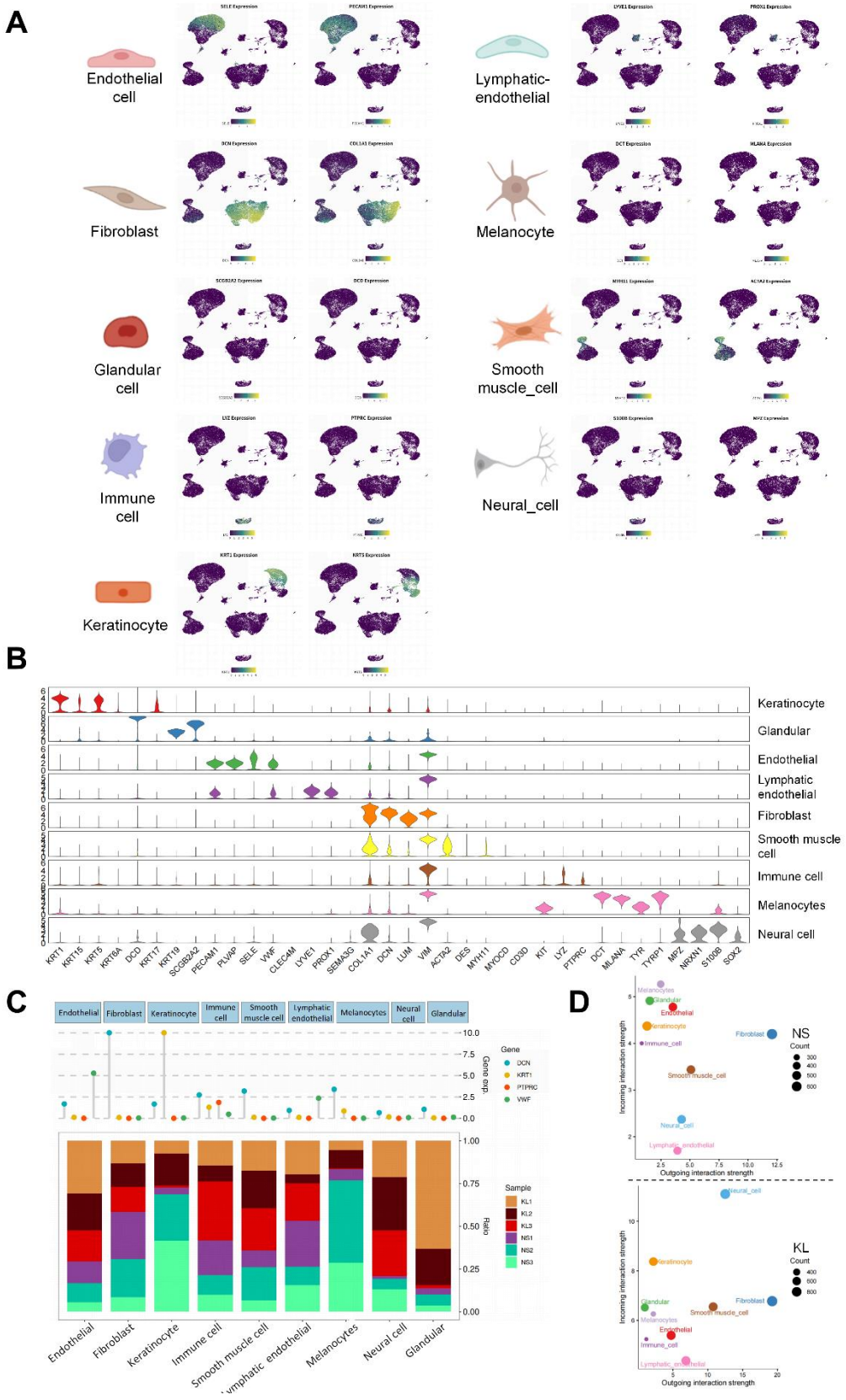


Figure S2 in revised manuscript: Integrated analysis of differential expression gene (DEG) and cell communication (related to Figure 1). (A) Feature plots of expression distribution of selected marker genes reported in the CellMarker database. (B) Violin plots of the marker genes for each cell type are presented. (C) The Cleveland dot plot shows the expression of typical genes (DCN, KRT1, PTPRC, and VWF) in different cell types; the histogram shows the proportion of each cell type in different samples. (D) The incoming and outgoing interaction strength of different cells.

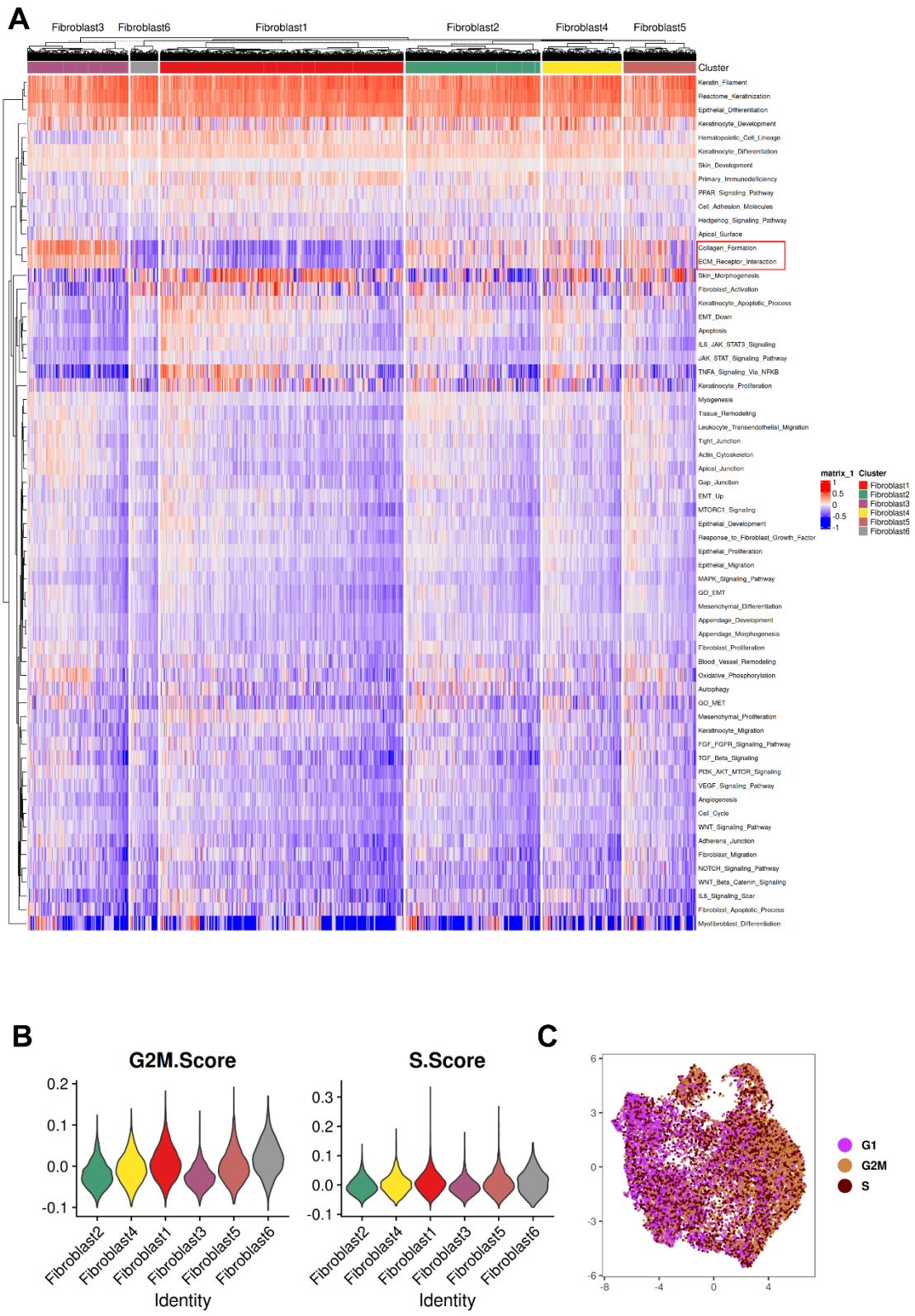


Figure S3 in revised manuscript: Analysis of signaling pathways enriched and cell cycle in fibroblast subtypes. (A) Signaling pathways in fibroblast clusters. Enriched

hallmark signaling pathways in fibroblast 3 were collagen formation and ECM receptor interaction. (B-C) The cell cycle phases among fibroblast subpopulations.

Comment 2: The expression of SELE, ACKR1, CD74, and HLA-DRA in Fibroblast cluster 4 raises the possibility that this cluster may consist of doublets. Moreover, the GSE163973 study by Deng et al. does not report a similar cluster, which calls for clarification regarding this discrepancy.

Response:

We sincerely appreciate your constructive suggestions. SELE and ACKR1 are classic markers of endothelial cells, and CD74 and HLA-DRA have been reported as classic markers of immune cells. In our study, the presence of the abovementioned markers in Fibroblast subset 4 is indeed suspicious, and your suspicion of doublet cell contamination is reasonable. Therefore, we switched to the harmony algorithm and conducted doublets detection. The results showed that the proportion of doublet cells was very low. Among 13,925 fibroblasts, there were only 543 doublets. Taking your comments into consideration, we have updated the scRNA-seq data throughout the manuscript **(Figure 1, 2, and 3; Figure S2 and S3 in the revised manuscript)**. In the new analysis results, no similar doublet cell contamination was found in either the analysis of major cell types or the analysis of fibroblast subsets. Meanwhile, the proportion of fibroblast subset 3 is high in keloid samples. Genes highly expressed in this subset, such as ASPN, COMP, and POSTN, are consistent with the results reported in the article by Deng et al. [1].

[1] Deng CC, Hu YF, Zhu DH, et al. Single-cell RNA-seq reveals fibroblast heterogeneity and increased mesenchymal fibroblasts in human fibrotic skin diseases. Nat Commun. 2021 Jun 17;12(1):3709.

The corresponding revisions in the manuscript are the same as the response to Comment 1.

Comment 3: Terms like LFDC, MFDC, and HFDC are used in the abstract without explanation, which may confuse readers unfamiliar with these abbreviations. Providing definitions in the abstract would improve clarity.

Response:

Thank you for your valuable advice. We sincerely apologize for not explaining the abbreviations LFDC, MFDC, and HFDC in the abstract. The main reason is that the journal has a word-limit for the abstract. In an attempt to reduce the word count, we only used the abbreviations, which unfortunately caused inconvenience to your understanding of the article. In response to your constructive suggestions, we have re-written and streamlined the abstract. We have summarized the three subtypes of LFDC, MFDC, and HFDC as "a novel keloid diagnostic classification system comprising three subtypes" to avoid any ambiguity. Meanwhile, for all the abbreviations in the abstract, we have indicated their full-length forms. For example, we have presented " differentially expressed fibroblast-differentiation-related genes (DEFDRGs)".

Corresponding revisions in the manuscript are as follows：

Abstract section (Page 1, Line 4-16):

Abstract

Keloids are dermal fibroproliferative skin disorders caused by abnormal wound healing, resulting in impaired skin function and aesthetic defects. Abnormal fibroblast proliferation and excessive collagen deposition are involved in keloid formation. This study investigated the role of fibroblast differentiation in keloid development. Single-cell and bulk RNA sequencing data of keloids were comprehensively analyzed, and 25 clinically relevant differentially expressed fibroblast-differentiation-related genes (DEFDRGs) were identified. Based on DEFDRGs, a novel keloid diagnostic classification system comprising three subtypes was constructed, indicating that DEFDRGs could serve as new therapeutic targets. Additionally, multiple microarray data, protein sequencing data, and immunohistochemistry of key markers in clinical keloid samples were used for further verification. In conclusion, this study established a molecular classification of keloids based on fibroblast differentiation for the first time, contributing to the further understanding of keloid pathogenesis and providing new insights for diagnosis and treatment.

Reviewer #1 (Reviewer #2 to the comments from reviewer #1):

Comment 1: In Figure 2A and 2C, Fibroblast 4 is identified as a fibroblast cluster but includes marker genes such as CD74, HLA-DRA, ACKR1, and SELE, typically associated with immune and endothelial cells. Additionally, this cluster is spatially distant from other fibroblast clusters in the UMAP plot, raising concerns that it may consist of doublets. Did the authors perform doublet removal during data processing? If so, what is the doublet score for this cluster? This information is critical to addressing the possibility of contamination by doublets. The dataset used (Deng et al., Nat. Commun.) does not report such a cluster. Furthermore, the authors rely on HPA database evidence for their claims, but there are no existing keloid single-cell studies demonstrating a fibroblast cluster expressing immune or endothelial markers. If such evidence exists, it should be provided to substantiate the authors' claim.

Response:

Thank you for your invaluable suggestions. As you mentioned, the presence of markers of other cell types, such as CD74, HLA-DRA, ACKR1, and SELE, in fibroblasts may indicate the possibility of doublet cell contamination. Although there are data in the HPA database regarding the expression of the abovementioned markers in fibroblasts, the data in the HPA may be affected by doublets or ambient RNA contamination. During the analysis of the single-cell data using the Canonical Correlation Analysis (CCA), deficiency of batch effect correction and doublet detection might have led to the occurrence of the situation.

We re-analyzed the single-cell data using the Harmony algorithm and employed DoubletFinder to identify doublets **(Figure 1 of the response letter)**. The results showed that the number of doublets was very small. In the new analysis results, no fibroblast subtypes expressing immune or endothelial markers were observed. Therefore, taking the reviewer's suggestions comprehensively into account, we re-analyzed the single-cell data in this paper using the Harmony algorithm and accordingly revised the figures and the manuscript **(Figure 1, 2, and 3; Figure S2 and S3 in the revised manuscript)**.

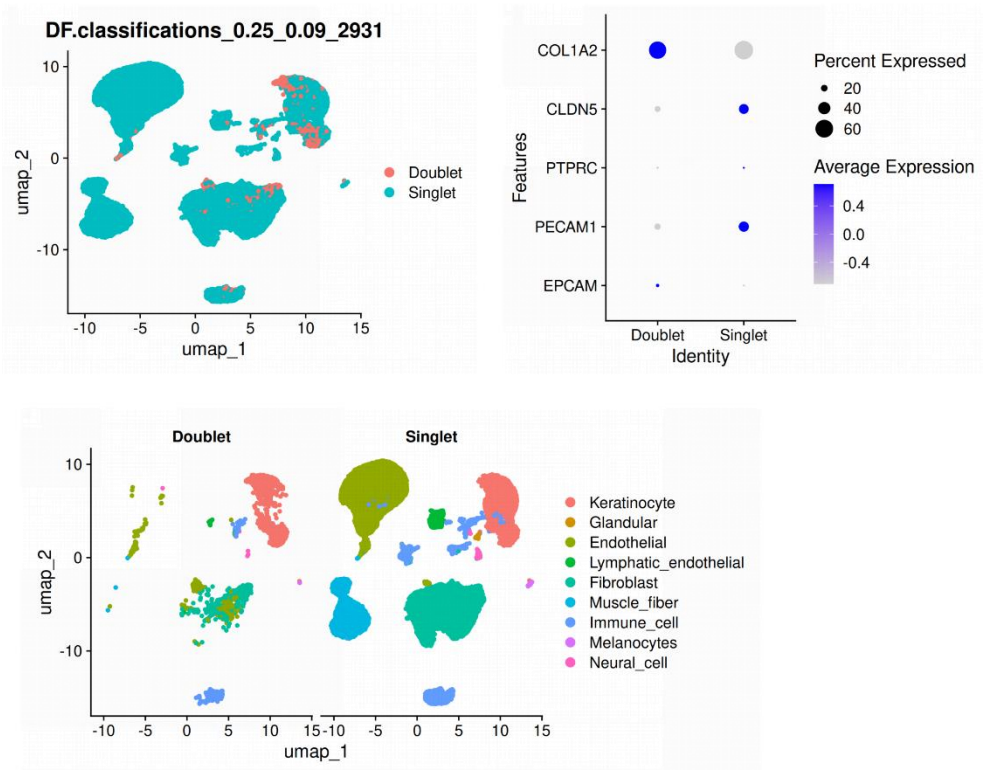


Figure 1 of the response letter. Doublets in scRNA-seq of GSE163973 were identified. The proportion of doublets was very low, reflecting the high quality of the data.

Corresponding revisions in the manuscript are as follows：

Method section (Page 3-4, Line 88-99):

2.2 ScRNA-seq Data Processing

Then, the quantified gene expression matrices were further analyzed with the Seurat pipeline (Version: 3.2.2)[19]. Only cells with transcript expression greater than 100,000, mitochondrial gene proportion less than 10%, and genes expressed in at least three cells could be included for subsequent analyses. Following quality control (QC) process, keloid and normal scar samples were merged into a single Seurat object. And then, the Seurat object was scaled and standardized by the "ScaleData" function. Besides, we determined the top 2,000 most variable genes using the "vst" method, which were defined as highly variable genes (HVG). Principal component analysis (PCA) was performed on these HVGs to reduce dimensionality. To mitigate batch effects between samples, harmonization was conducted using the Harmony algorithm with default parameters, integrating the data across samples while preserving biological variability[19]. DoubletFinder was utilized to detect doublets during the analytical process. The first twenty harmonized principal components (PCs) were subsequently utilized for Uniform Manifold Approximation and Projection (UMAP) analysis.

Result section (Page 9-11, Line 256-315):

3.1 Cell Type Annotation

After stringent QC filtering, a total of 44,463 single cells (keloids: 23,638; normal scars: 20,825) from scRNA-seq data met the criteria and were included for downstream analyses. The experimental workflow is detailed in Figure 1A and Figure S1. UMAP-based clustering identified 18 distinct cell clusters, which were annotated into 9 major cell types: endothelial cells, fibroblasts, glandular cells, immune cells, keratinocytes, lymphatic endothelial cells, melanocytes, smooth muscle cells, and neural cells (Figure 1B). Spatial distribution patterns of these cell populations between keloid and normal scar tissues are visualized in Figure 1B. Feature plots demonstrate the expression levels and cellular distribution of canonical biomarkers retrieved from the CellMarker database across all annotated cell types (Figure S2A). Figure S2B further validates the cellular annotation accuracy through marker gene expression patterns. The Cleveland dot plot was used to describe expression levels of typical marker genes (*DCN*, *KRT1*, *PTPRC*, and *VWF*) and the proportions of the 9 cell types in different samples (Figure S2C). Figure 1C illustrates the average number and proportion of each cell type. The results indicate that endothelial cells and fibroblasts are predominant across all samples, which is in line with previous research[40].

3.2 Cell Communication Analysis

Cell communication analysis demonstrated fibroblasts as the predominant cellular interactors, exhibiting the highest communication frequency across all cell types (Figures 1D-1E). Quantitative evaluation revealed significantly enhanced global communication activity in keloids compared to normal scar tissues (Figure 1F). Notably, fibroblasts exhibit the strongest outgoing interaction strength in both keloids and normal scars (Figure S2D).

3.3 Fibroblast Subtype Analysis Revealed Cell Heterogeneity

A total of 13,925 fibroblasts were extracted only for further analysis. Through unsupervised clustering, 6 distinct clusters were identified and annotated as fibroblast 1-6 (Figure 2A). Among them, fibroblast 1 and 2 predominated in normal scars, while fibroblast 3 occupied most of the keloids (Figure 2B). The expression levels of typical marker genes (*ACTA2*, *COL1A1*, *TGFBI*, and *TOP2A*) were used to prove the accuracy of the fibroblast screening, and the proportion of each fibroblast cluster in keloid and normal scars was visualized by the bar plot, also proving that fibroblast 3 is mainly found in keloids (Figure 2C). Fibroblast 3 highly expresses *POSTN*, *ASPN*,

and *COMP*, which have been reported to be involved in keloid formation (Figure 2D)[14]. Extensive cell communication existed among different fibroblast subtypes. Particularly, fibroblast 3 interacted mostly with others, and COL1A1 and COL1A2 were key intercellular association genes (Figure 2E-F). Enriched hallmark signaling pathways in fibroblast 3 were collagen formation and ECM receptor interaction (Figure S3A). Figure S3B-C demonstrate distinct cell cycle phases among fibroblast subpopulations.

3.4. Reveal Differentiation Fate Decisions and Expression Dynamics of Keloid Fibroblasts via Trajectory Analysis

To delineate the differentiation hierarchy of keloid fibroblasts, we employed Monocle 2 for single-cell pseudotemporal ordering analysis. This computational framework reconstructed lineage trajectories through dimensionality reduction (DDRTree algorithm) and RNA velocity estimation, revealing 7 distinct cellular states along fibroblast maturation paths (Figure 3A-C). Furthermore, the pseudotime and cell lineage trajectory of each fibroblast cluster were illustrated separately in Figure 3D. Notably, fibroblast 3 predominantly occupied intermediate-late pseudotime positions (States 3-6). Through trajectory analysis, DEGs associated with fibroblast differentiation were identified and named as FDRGs. Top 100 FDRGs were shown in the heatmap (Figure 3E).

3.5. Identification of Keloid DEFDRGs

Through the aforementioned analyses, we mapped the transcriptomes of all fibroblast states and used Monocle 2 to identify 14,323 FDRGs (Monocle-sig, $P < 0.05$), which are the genes determining cell fate. Subsequent integration of RNA-seq profiles with clinical metadata from the SAR database through univariate Cox regression revealed 32 diagnosis-associated (Diagnosis-sig, $P < 0.05$) and 137 time-significant (Time-sig, $P < 0.05$) DEGs. Intersectional analysis of these three gene sets identified 25 core FDRGs, named DEFDRGs, exhibiting both differentiation-dependent expression patterns and clinical relevance (Figure 3F and S4A). To elucidate clinical correlations, a co-expressive network of DEFDRGs was constructed using nonparametric tests and Pearson correlation (Figure S4B). Labels of various colors meant genes for different monocle states. Curves in different colors represented positive or negative correlations. The DEFDRGs had tight associations, indicating underlying regulation mechanisms of clinical outcomes and fibroblast fates. Most genes were positively correlated, while *Kallikrein Related Peptidase 6 (KLK6)* and *Tryptase*

Beta 2 (TPSB2) were mostly negatively correlated with the others. Boxplot validation confirmed differential expression patterns across diagnosis and time (Figure S4C).

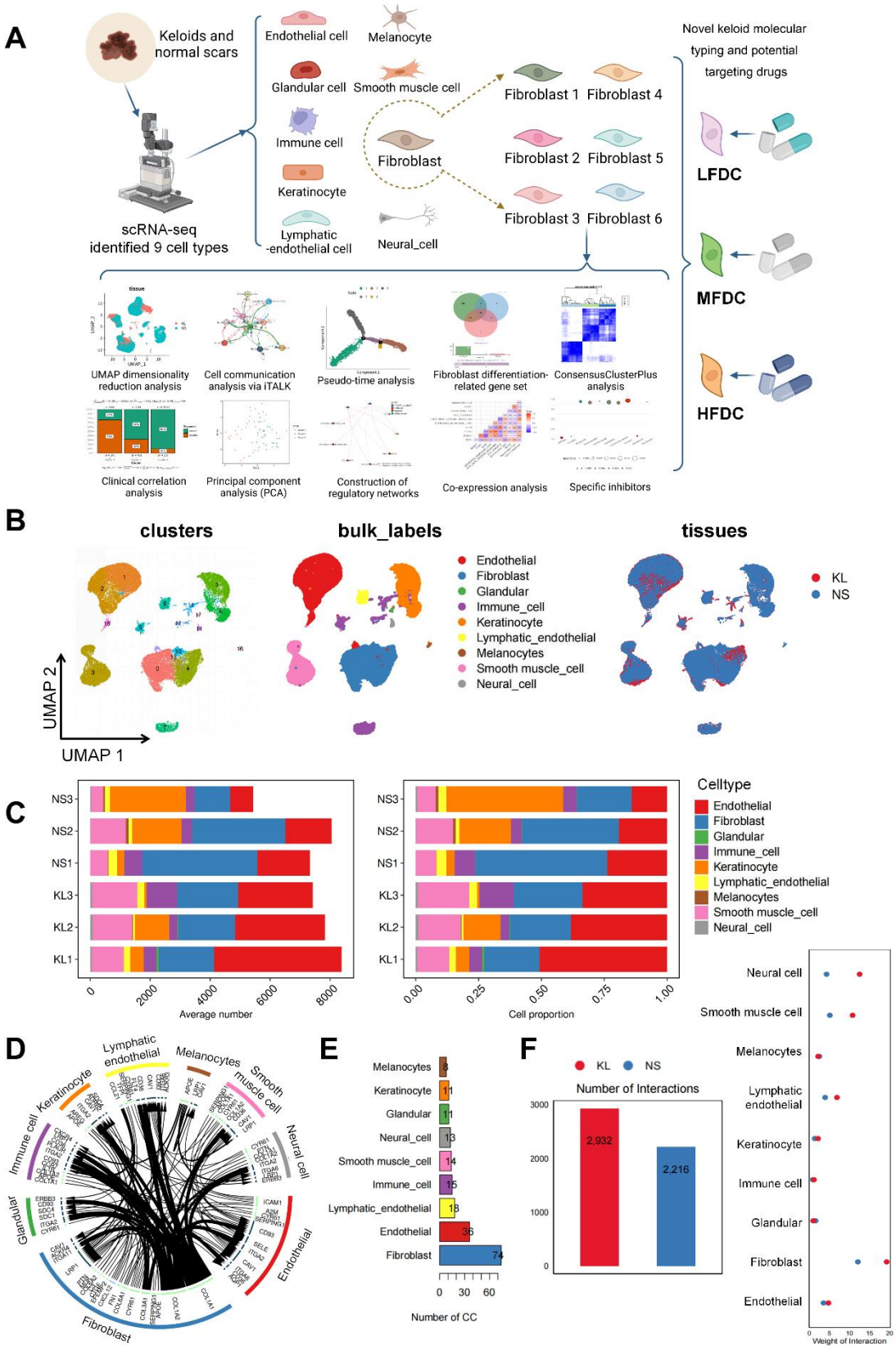


Figure 1 in revised manuscript: Comprehensive analysis of scRNA-seq revealed the heterogeneity of keloids and normal scars. (A) Schematic diagram of workflow in study design and analysis. (B) UMAP dimensionality reduction analysis clustering 44,463 cells into nine cell types and showing the distribution of the above cells in keloids and normal scars. (C) The average number and proportion of nine cell types in each sample. Fibroblasts and endothelial cells account for a relatively high proportion. (D-E) Cell communications of each cell type. Fibroblasts are involved in the largest number of cell communications. (F) The left figure shows the number of interactions in keloids and normal scars, and the right figure shows the weight of interactions of each cell type in the two groups.

Abbreviations: KL: Keloid; NS: Normal skin; CC: Cell Communication

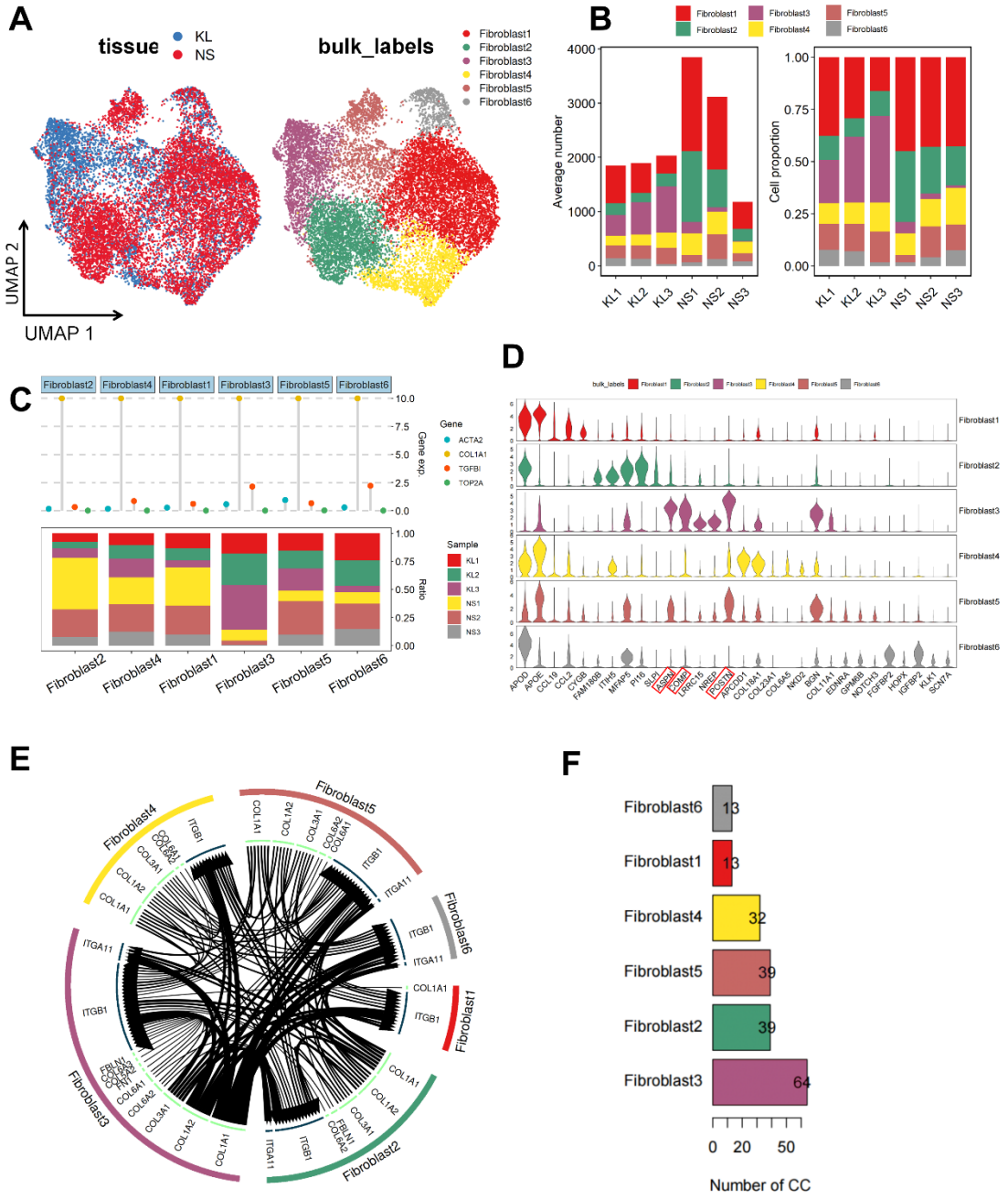


Figure 2 in revised manuscript: Identification of fibroblast subtypes and cell communication. (A) Six fibroblast subtypes were identified by UMAP dimensionality reduction analysis. (B) The histogram shows the average number and proportion of fibroblast subtypes in different samples. Fibroblast 3 predominates in keloids, whereas fibroblast 2 dominates normal scars. (C) The Cleveland dot plot shows 4 classical markers' expression (ACTA2, COL1A1, TGFBI, and TOP2A) in fibroblasts. The bar plot displays the proportion of each fibroblast cluster in 6 samples (3 keloids and 3 normal scars). (D) The expression of marker genes in each fibroblast subtype. ASPN,

COMP, and POSTN are highly expressed in fibroblast 3. (E-F) Cell communication between fibroblast subtypes. Fibroblast 3 takes part in a greater number of cell communications.

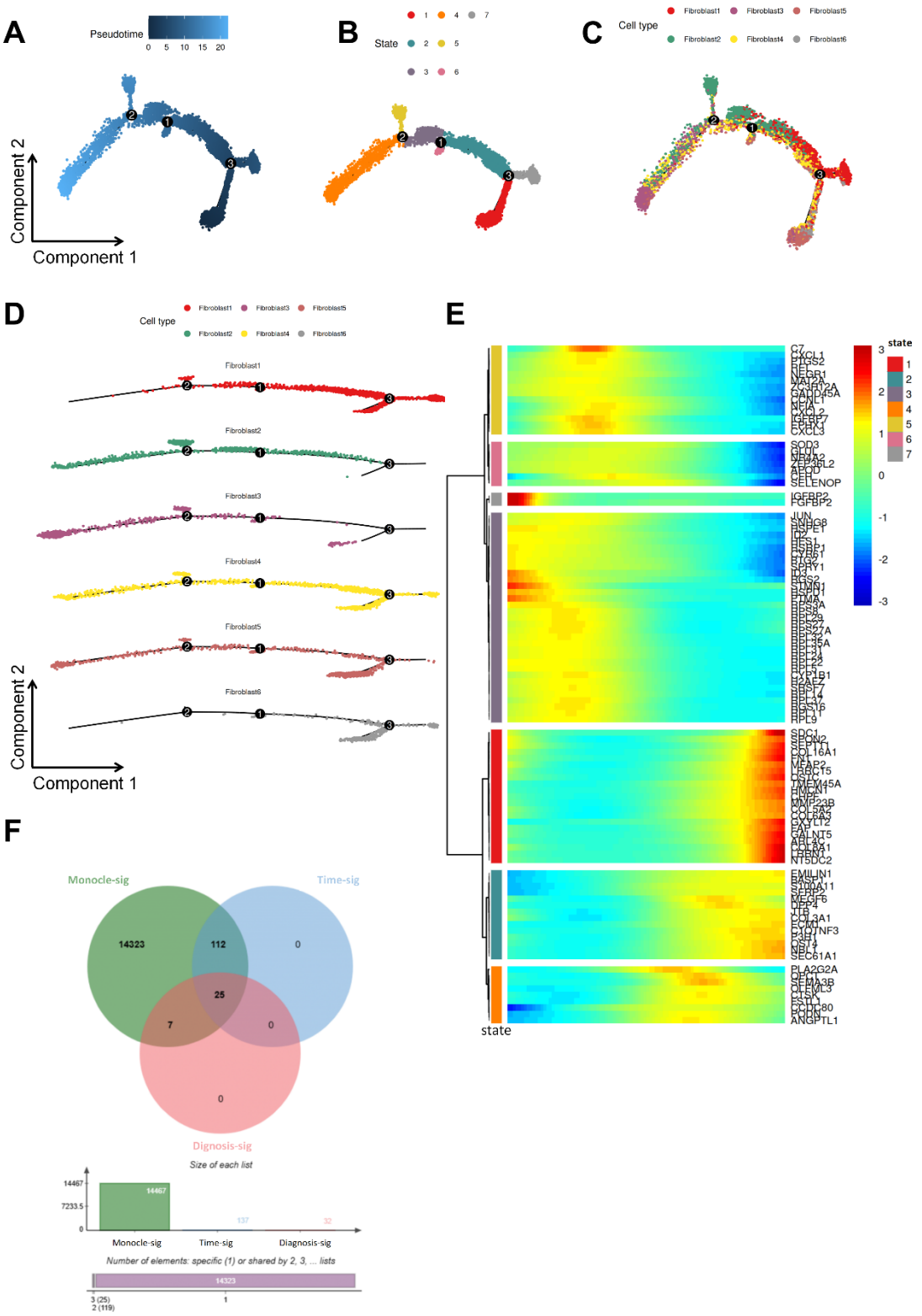


Figure 3 in revised manuscript: Pseudotime analysis and identification of DEFDRGs. (A) The pseudotime plot showed the potential fibroblast differentiation sequence. (B) Detection of 7 differentiation states. (C-D) Display of fibroblast subtypes on different differentiation trajectories. (E) The expression of the top 100 representative fate-related genes differentially expressed across. (F) Identification of DEFDRGs. DEFDRGs were obtained by taking the intersection of the fibroblast differentiation-related DEGs and the DEGs related to clinical diagnosis and time.

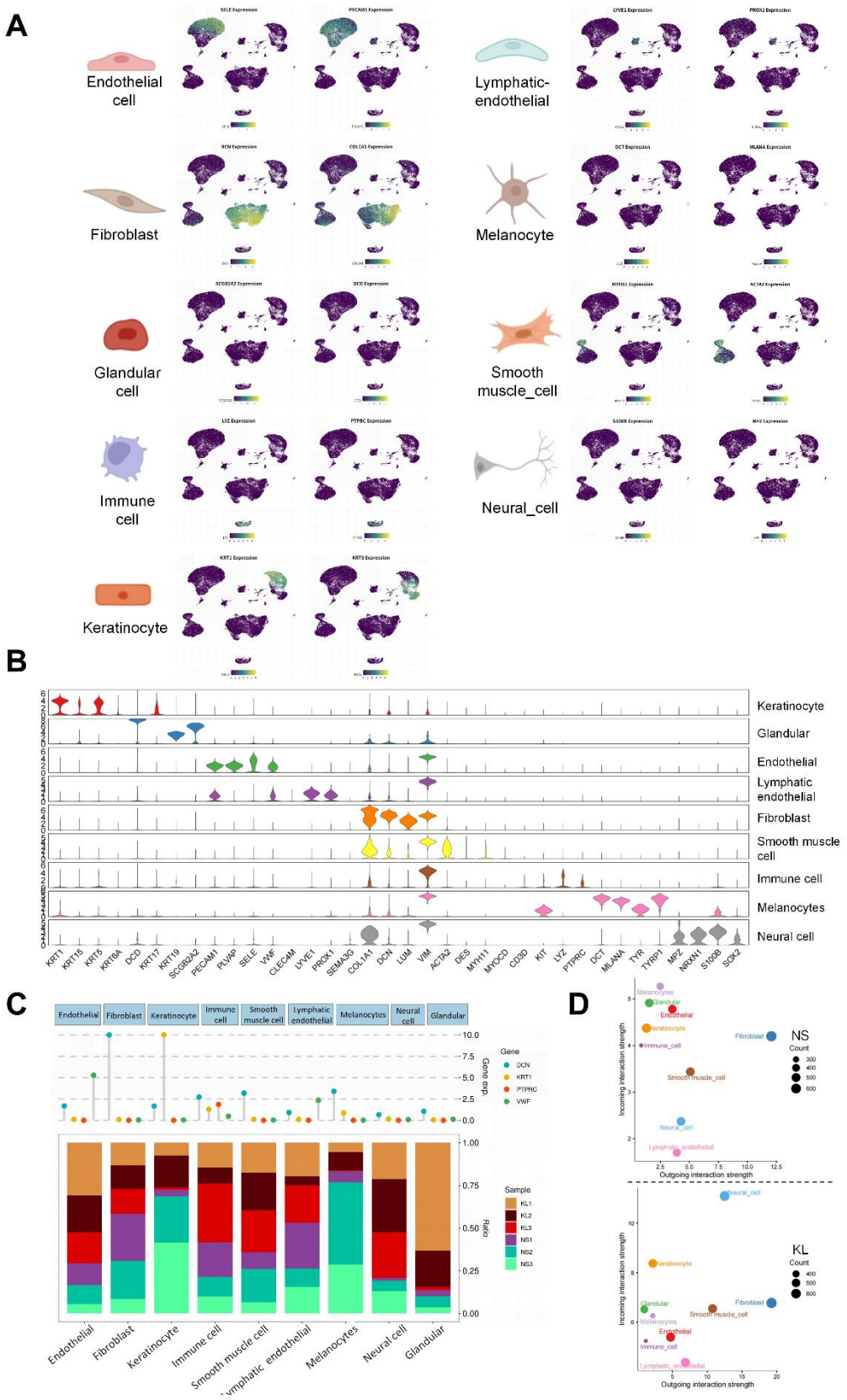


Figure S2 in revised manuscript: Integrated analysis of differential expression gene (DEG) and cell communication (related to Figure 1). (A) Feature plots of expression distribution of selected marker genes reported in the CellMarker database. (B) Violin

plots of the marker genes for each cell type are presented. (C) The Cleveland dot plot shows the expression of typical genes (DCN, KRT1, PTPRC, and VWF) in different cell types; the histogram shows the proportion of each cell type in different samples. (D) The incoming and outgoing interaction strength of different cells.

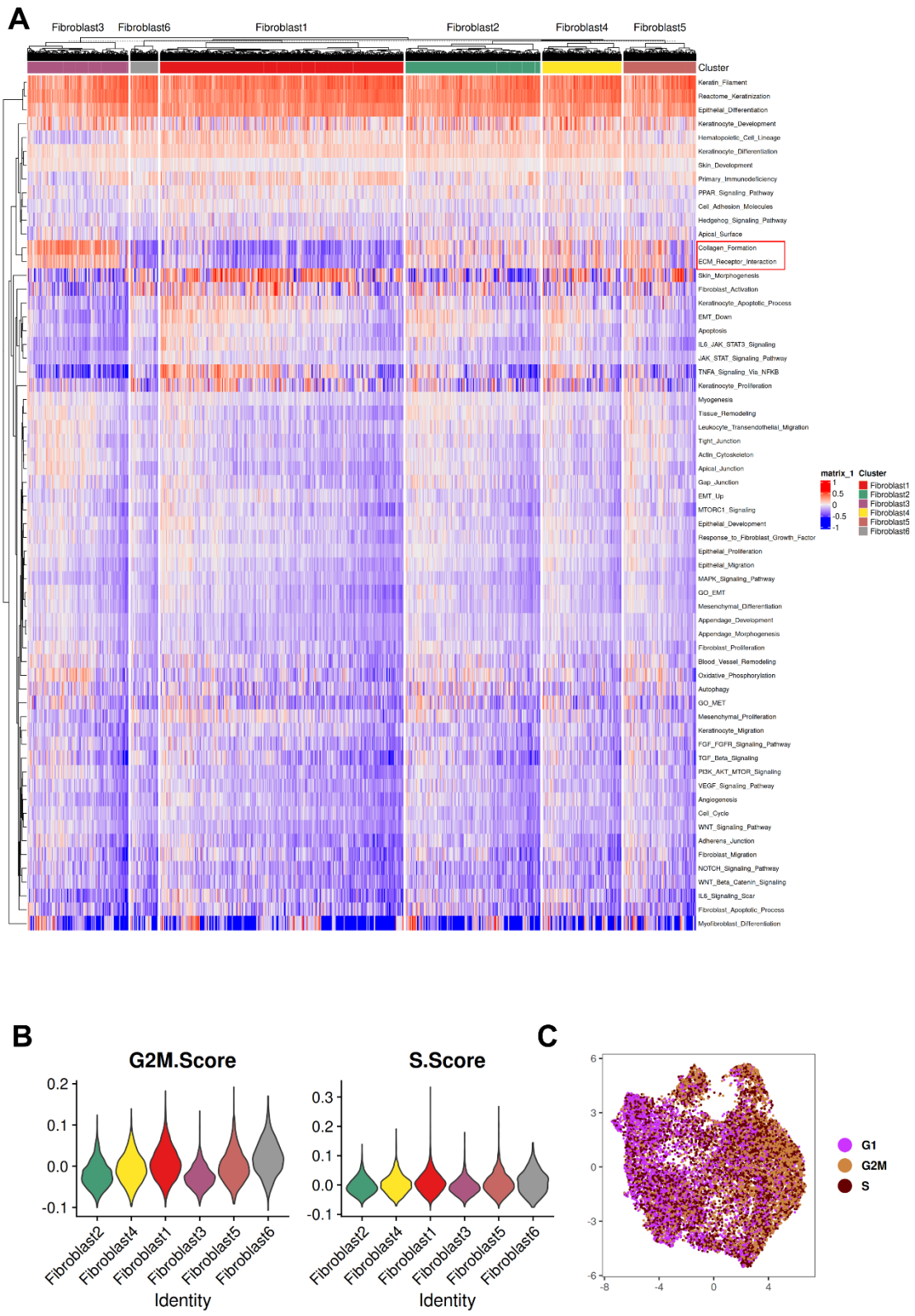


Figure S3 in revised manuscript: Analysis of signaling pathways enriched and cell cycle in fibroblast subtypes. (A) Signaling pathways in fibroblast clusters. Enriched

hallmark signaling pathways in fibroblast 3 were collagen formation and ECM receptor interaction. (B-C) The cell cycle phases among fibroblast subpopulations.

Comment 2: As shown in Figure S7, when Harmony was used for batch effect correction, no clusters expressing immune or endothelial markers were observed among the fibroblast subtypes. Additionally, the cluster represented by Fibroblast 6 in Figure 2B, which consists solely of KL1 samples, is also absent. This raises questions about whether batch effects were sufficiently corrected during data processing using CCA. Despite these discrepancies, the authors have not updated the main figure. This inconsistency requires further consideration and correction in the manuscript.

Response:

We sincerely appreciate your valuable advice. We have also noticed that after performing batch effect correction using the Harmony algorithm, no fibroblast subtypes expressing immune or endothelial cell markers were observed. Moreover, the fibroblast 6 subset, which was only present in the KL1 sample in the original Figure 2B, also disappeared. This might be attributed to the insufficient correction of batch effects during data processing with CCA.

Therefore, in light of your suggestions, we re-analyzed the scRNA-seq data in this paper using the Harmony algorithm, and rearranged the figures and revised the manuscript according to the new analysis results **(Figure 1, 2, and 3; Figure S2 and S3 in the revised manuscript)**. Notably, in the new analysis results, 25 DEFDRGs identified in the previous analysis were also verified in new analysis **(Figure 2F in the revised manuscript)**, which demonstrates the stability and reliability of the results.

The corresponding revisions in the manuscript are the same as the response to Comment 1.

Comment 3: The dataset “SRX3882450: RNA Seq of Keloid skin” focuses on wound healing in keloid-prone individuals. The RNA-seq data from keloid-prone individuals do not clearly establish whether the samples represent actual keloid lesions. Clinical photographs of day 0 and day 42 samples would provide critical context.

Response:

We sincerely apologize for any misunderstanding resulting from our inappropriate expression. The dataset "SRX3882450: RNA Seq of Keloid skin" presents the sequencing results of skin samples collected at different time points after injury from keloid-susceptible patients and normal healthy patients. It can reflect the differences in wound healing between keloid-affected individuals and normal healthy patients.

Indeed, as you pointed out, the clinical gross photos at DAY 0 and DAY 42 can more intuitively reflect the formation of keloids. Therefore, we contacted the original author John A. McGrath, a recent President of the European Society for Dermatological Research, via email, hoping to obtain the corresponding clinical photos. We were honored to receive a reply. However, the relevant clinical photos were unavailable, likely due to considerations for patient privacy protection. According to John A. McGrath's reply and the description in the study, the time point of DAY 42 was determined based on the patients' feedback regarding their concerns about keloid formation, which was based on clinical timings rather than considerations of the keloid -formation mechanism. The relevant email correspondence is as follows:

Thanks for your mail and kind comments. The timing of the second biopsy was arbitrary- we had asked the keloid prone individuals how soon after injury did they typically start to become aware that a keloid might be forming. Clinical timing rather than mechanism-based decision. The more critical time for keloid pathogenesis is likely to be much earlier and we have plans to do such earlier second biopsies in the future. The steroids had nothing to do with the omics - we just tried to do that as part of clinical care to try to reduce the risk of keloid forming following the research study.

Comment 4: Does the day 42 sample represent a keloid formation, or does it merely show changes in wound healing?.

Response:

We are truly grateful for your invaluable advice. Concerning whether DAY 42 indicates the formation of keloids or simply depicts the wound healing process, after a meticulous

review of the original article and in-depth email consultations with the original author, John A. McGrath, our understanding is as follows:

The overarching research approach of this study involved conducting biopsy samplings (DAY 0) on the buttocks of both keloid-prone patients and normal healthy individuals. Simultaneously, wounds were created, and then samples were re-collected from the exact same sites on DAY 42. This 42-day time point was determined according to the feedback from patients who were worried about keloid formation. Moreover, in line with ethical requirements, after the DAY 42 sampling, intralesional corticosteroid injections were administered to the wounds of keloid-prone patients to prevent keloid formation. However, despite these efforts, keloids still developed in 2 patients later. Had it not been for the local intralesional corticosteroid injections, keloids might well have emerged at the biopsy sites in the other 6 keloid-prone individuals as well. This indeed represents a limitation of this study. Evidently, even though keloids might not have fully formed by DAY 42, this time point could very well be a crucial stage in the development process of keloids. The original author, who has an in-depth understanding of keloids due to extensive research, also noted in the email reply, "The more critical time for keloid pathogenesis is likely to be much earlier, and we have plans to conduct such earlier second biopsies in the future." Clearly, the formation of keloids is a protracted process, and analyzing this data has the potential to uncover the underlying mechanisms of keloid formation. In addition, we have utilized multiple external datasets for verification **(Figure S5C-F in the revised manuscript)**, which further validates the effectiveness of our classification system.

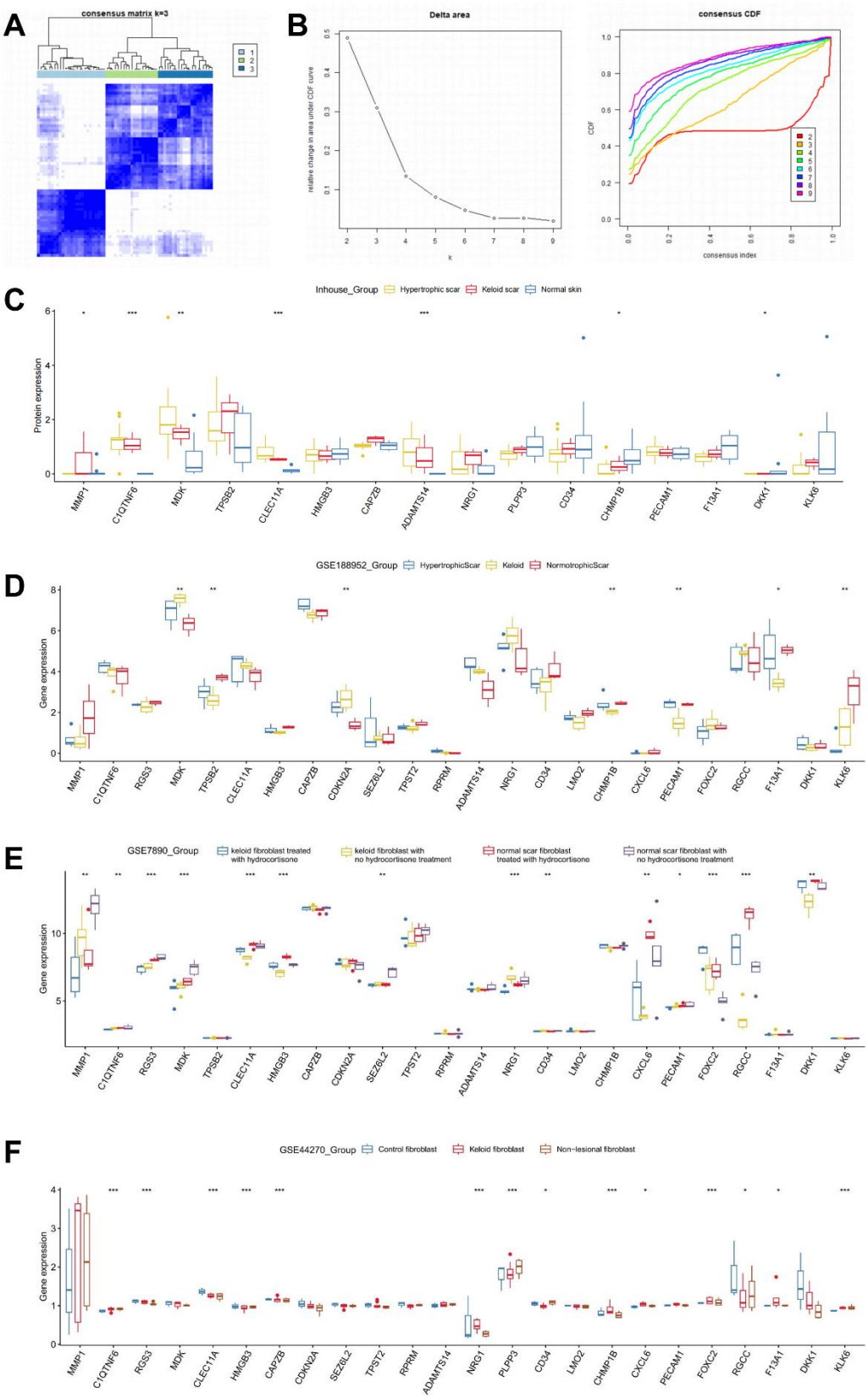


Figure S5 in the revised manuscript: Consensus clustering and verification in public databases (related to Figure 4). (A) Consensus clustering analysis based on 25 DEFDRGs. (B) The clustering effect is more specific when $k = 3$. (C) The differential

expression of proteins encoded by DEFDRGs in normal skin, hypertrophic scar, and keloid. (D-F) The verification of the differential expression of DEFDRGs using sequencing data in public databases. These three datasets found that many genes among these 25 DEFDRGs had statistical differences.

No.: COMMSBIO-24-1782B

Manuscript title: Uncovering a Novel Fibroblast Differentiation-Based Keloid Classification by Integration of Single-cell and Bulk RNA Sequencing

Communications Biology

25th May 2025

Dear reviewers

The interactive review started on: 25th March 2025

We are deeply grateful for your valuable time and constructive suggestions during the manuscript evaluation process. Your expert insights regarding key methodological considerations have significantly strengthened the scientific robustness and empirical validity of this work. In rigorous adherence to your recommendations, we have systematically addressed each comment, with all modifications clearly highlighted in the tracked-changes manuscript and corresponding figures.

Following rigorous incorporation of the reviewers' suggestions, we respectfully resubmit the revised manuscript for evaluation. For enhanced transparency: 1) Revisions from the previous round remain marked in red text, and 2) Current modifications addressing new feedback are highlighted in yellow. A comprehensive point-by-point response document detailing all revisions accompanies this submission.

We hope that the revision is acceptable. We look forward to hearing from you soon.

Sincerely yours,

25th May 2025

Reviewer #3

Comment 1: Where are the results of RNA velocity since authors mentioned this in the results content of paragraph 3.4 (Line 295)?

Response 1: We sincerely appreciate the reviewer's careful attention to detail and apologize for the oversight in our original manuscript. The lineage trajectories in this study were derived from Monocle 2 analysis rather than the RNA velocity, and the reference to RNA velocity has been removed from the revised manuscript.

Corresponding revisions in the manuscript are as follows：

Result section (Page 10, Line 294-296):

This computational framework reconstructed lineage trajectories through dimensionality reduction (DDRTree algorithm), revealing 7 distinct cellular states along fibroblast maturation paths (Figure 3A-C).

Comment 2: How do you set the starting point of fibroblast pseudotime trajectory? What is your rationale in Figure 3A?

Response 2: Thank you for this important question. The fibroblast pseudotime trajectory was inferred using Monocle 2. Monocle 2 employs DDRTree algorithm to learn a principal tree on a population of single cells by default, asserting that it describes the sequence of changes to global gene expression levels as a cell progresses through the biological process under study [1]. The starting point of the pseudotime trajectory is biologically interpreted as the initial state of the process. This assignment is rooted in the algorithm's reliance on dynamically changing gene expression profiles over pseudotime, where genes with time-dependent expression patterns are used to define the ordering of cells along the principal tree. Thus, the trajectory's origin is computationally determined by identifying the cellular state associated with the earliest detectable changes in these time-sensitive gene expression modules.

[1] Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017 Oct;14(10):979-982.

Comment 3: What is pseudotime direction in Figure 3E?

Response 3: Thank you for your valuable comments. Figure 3E presents the expression of Top

100 FDRGs. In this study, we focus more on the gene set related to fibroblast differentiation rather than the direction of differentiation itself. Moreover, the temporal order obtained from Monocle 2 analysis is a reference pseudotime. Therefore, we did not over - emphasize the pseudotime direction in Figure 3E. We apologize for any confusion this may have caused. According to your suggestion, we have added the pseudotime direction in Figure 3E.

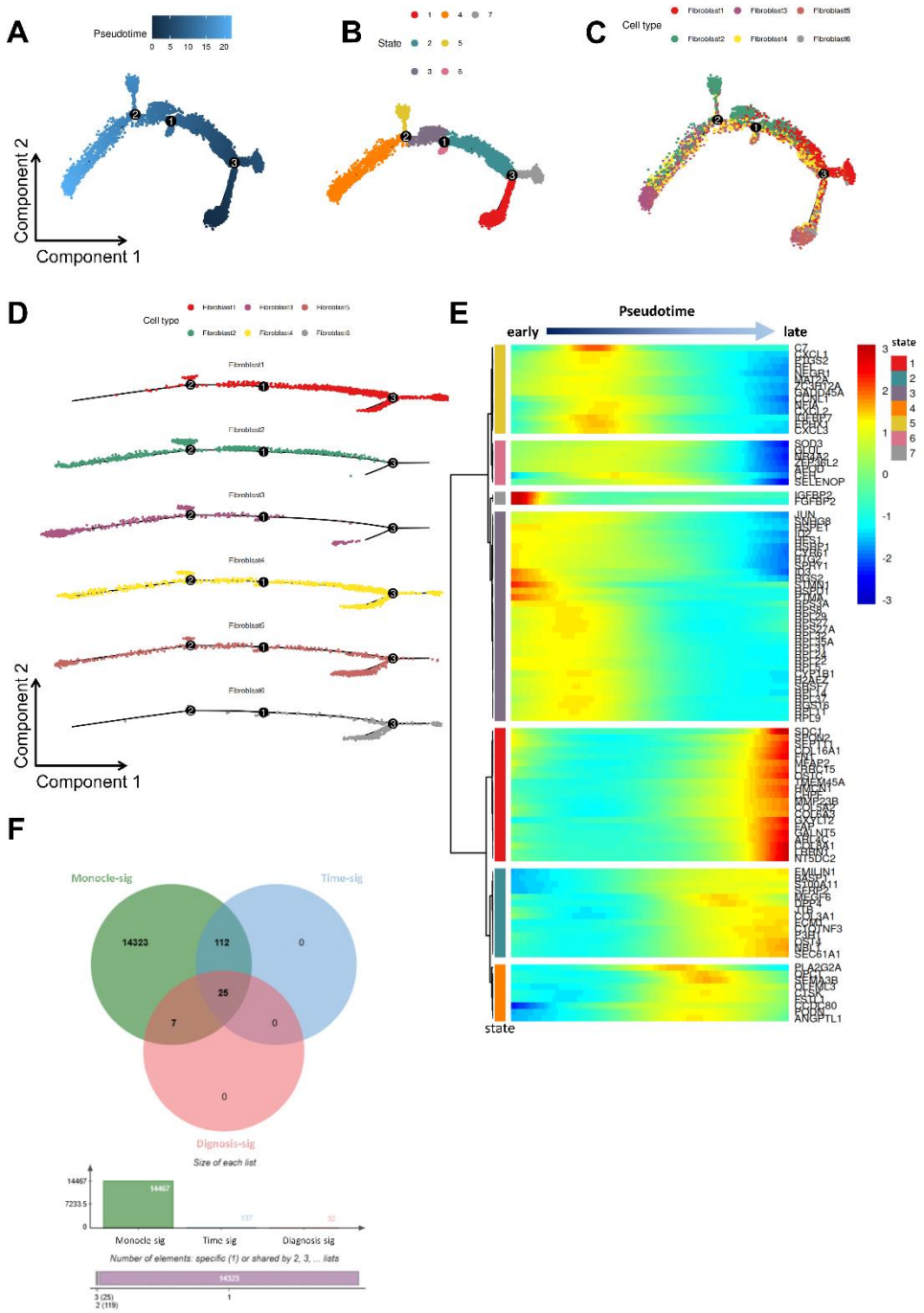


Figure 3 in revised manuscript.

Comment 4: **In Figure 7A, the images from MMP1 staining in LFDC, FOXC2 staining in MFDC, KLK6 and TPSB2 staining in HFDC are like a normal region, not a true keloid. Please change the representative images and quantify the results again.**

Response 4: Thank you for your valuable advice. We sincerely apologize that our staining results were insufficiently representative and for any inconvenience caused. We have updated the immunohistochemistry results in the manuscript and performed quantitative analysis of the new findings. The revised data are presented in **Figure 7 in revised manuscript**.

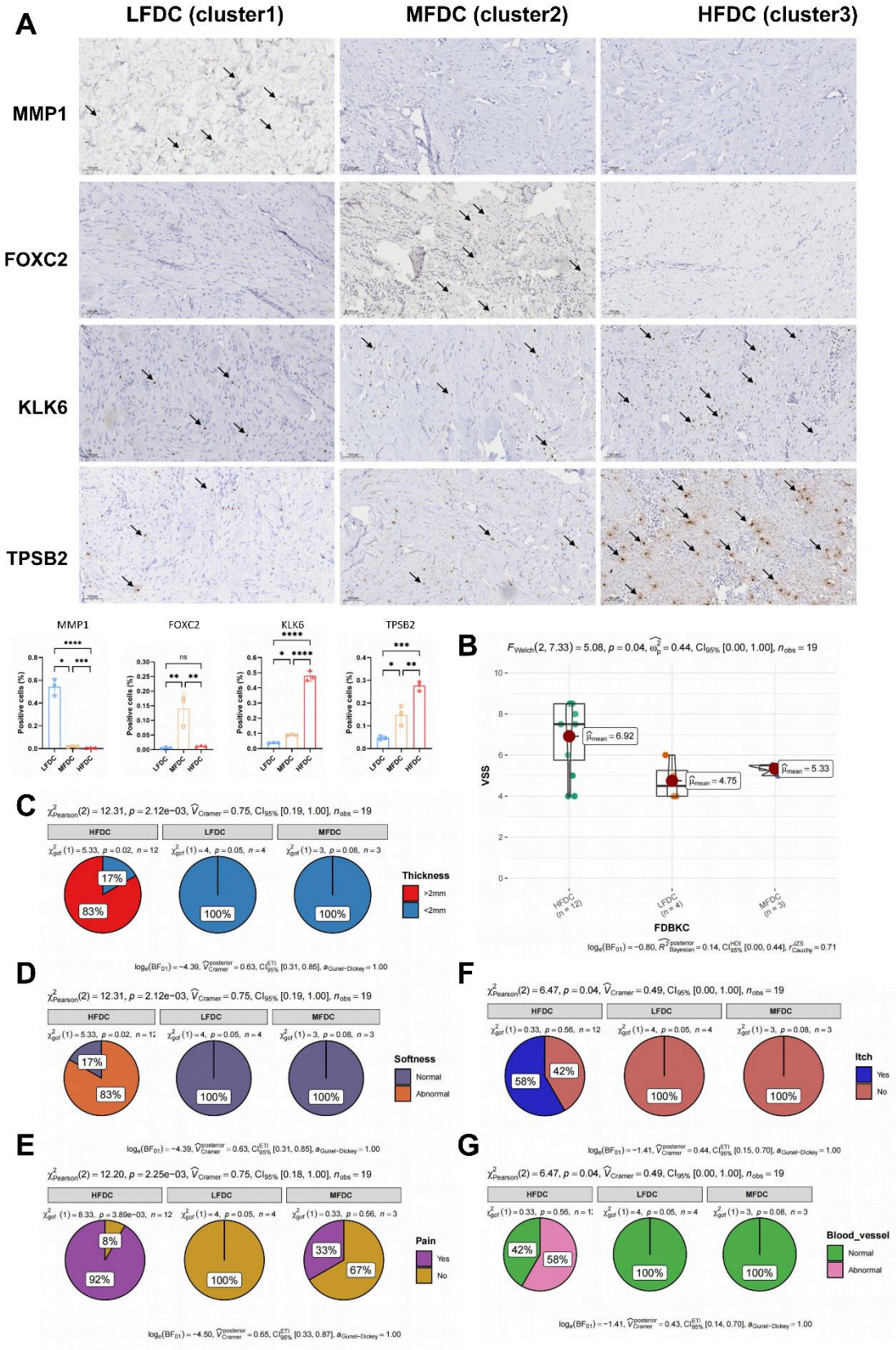


Figure 7 in revised manuscript.

Comment 5: PECAM1 is a very specific marker for endothelial cells. In addition, TPSB2 is a specific marker for mast cells in the skin scRNA-seq dataset. Even though authors used the software to remove doublet cells, sometimes free form mRNA can be captured within one droplet. So, when we identify some markers, we usually confirm whether the markers are specific for fibroblasts. If not, we check the references first that also showed similar results and try to explain the functions of those markers in fibroblast subgroups. However, there is almost no paper showing that PECAM1 and TPSB2 are expressed by fibroblasts. It's appreciated that authors used ICC in fibroblast cell line to verify that fibroblasts can also express a little PECAM1 protein. It's suggested that authors should use RNA FISH to demonstrate that PECAM1 and TPSB2 mRNA are also expressed in fibroblasts of keloid tissues. If authors can't verify this, please remove them from the manuscript.

Response 5: Thank you for your valuable feedback. The DEFDRGs we identified included PECAM1 and TPSB2. As you rightly pointed out, PECAM1 is a specific marker for endothelial cells, whereas TPSB2 serves as a specific marker for mast cells in the skin. To verify the reliability of our data analysis, we performed external validation using keloid-related datasets and fibroblast sequencing data extracted from keloid tissues from public databases. Notably, we detected significant expression of both TPSB2 and PECAM1 in fibroblasts **(Figure S5E-F in revised manuscript)**.

Additionally, we conducted a systematic literature review. As stated in the previous response, published studies have confirmed through single-cell sequencing data and *in vivo/in vitro* experiments the existence of fibroblast subpopulations expressing endothelial cell markers (e.g., PECAM1) [1]. Similarly, research has identified a distinct fibroblast subset expressing endothelial-associated genes in hindlimb ischemia models, which actively participates in angiogenesis through paracrine regulation [2]. These findings collectively support the biological plausibility of our observations regarding dual-marker expression patterns. Regarding TPSB2, current literature indeed lacks substantial evidence supporting its expression in fibroblasts. However, there is a significant correlation between TPSB2 and the pathogenesis of fibrotic diseases [3-4].

In addition to verifying the expression of PECAM1 protein in fibroblasts through cellular immunofluorescence co-staining, we also performed co-staining of TPSB2 with fibroblast markers COL1 and VIM. The immunofluorescence results confirmed that TPSB2 is expressed in some

fibroblasts, though at a relatively low frequency (**Figure 1 of the response letter**). The antibodies used were VIM (#5741, CST), COL1 (ab138492, abcam), and TPSB2 (K007207P, Solarbio).

Finally, according to your valuable suggestions, we performed RNA FISH on clinically obtained keloid samples to demonstrate that PECAM1 and TPSB2 mRNA are also expressed in fibroblasts of keloid tissues. The staining results are shown in **Figure 2 and 3 of the response letter**. In keloid samples, there are cells co-stained with TPSB2, COL1, and VIM probes, as well as cells co-stained with PECAM1, COL1, and VIM probes.

[1] Pal D, Ghatak S, Singh K, et al. Identification of a physiologic vasculogenic fibroblast state to achieve tissue repair. Nat Commun. 2023 Feb 28;14(1):1129.

[2] Meng S, Lv J, Chanda PK, et al. Reservoir of Fibroblasts Promotes Recovery From Limb Ischemia. Circulation. 2020 Oct 27;142(17):1647-1662.

[3] Xu X, Rivkind A, Pappo O, et al. Role of mast cells and myofibroblasts in human peritoneal adhesion formation. Ann Surg. 2002 Nov;236(5):593-601.

[4] Akers IA, Parsons M, Hill MR, et al. Mast cell tryptase stimulates human lung fibroblast proliferation via protease-activated receptor-2. Am J Physiol Lung Cell Mol Physiol. 2000 Jan;278(1):L193-201.

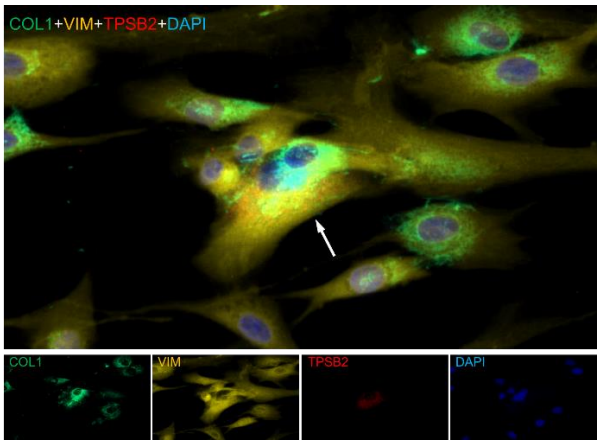


Figure 1 of the response letter. Immunofluorescence co-staining of COL1, VIM, TPSB2, and DAPI in human primary dermal fibroblasts. The arrow indicates the co-stained cells.

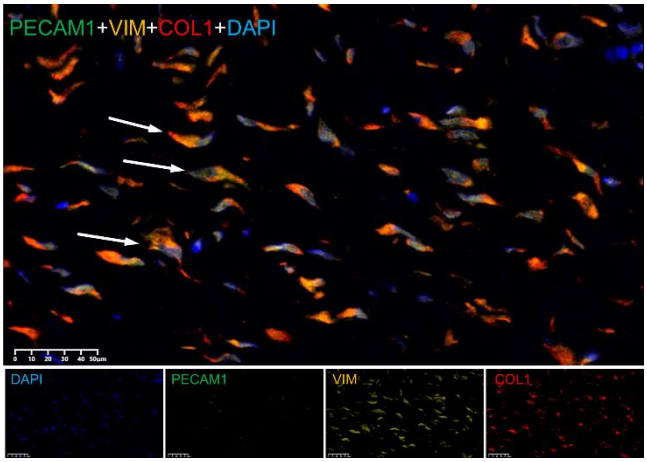


Figure 2 of the response letter. The RNA FISH results of COL1, VIM, PECAM1, and DAPI in human keloid samples. The arrow indicates the co-stained cells.

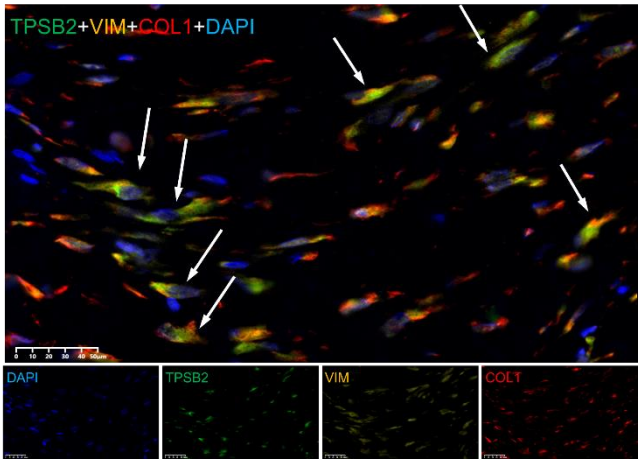


Figure 3 of the response letter. The RNA FISH results of COL1, VIM, TPSB2, and DAPI in human keloid samples. The arrow indicates the co-stained cells.

No.: COMMSBIO-24-1782C
Manuscript title: Uncovering a Novel Fibroblast Differentiation-Based Keloid Classification by Integration of Single-cell and Bulk RNA Sequencing
Communications Biology

Dear editors and reviewers,

First and foremost, we would like to extend our sincerest gratitude to the reviewers for their meticulous review of our manuscript and for providing thoughtful comments and constructive suggestions. Your insights have been invaluable in enhancing the quality and clarity of our work.

In response to your feedback, we have carefully revised the manuscript to address each point raised. We are pleased to re-submit the updated version for your consideration. In our detailed response letter, the reviewers' comments are listed in italicized font, followed by our point-by-point responses highlighted in blue for easy identification. To enhance the manuscript's quality, logical coherence, and readability, we conducted comprehensive linguistic refinement with assistance from native English-speaking colleagues. In addition, we have made modifications by cross-checking one by one in accordance with the formatting requirements of the journal, as specifically shown in the document *CommsBio_AIP_table_COMMSBIO-24-1782C*. All corrections and revisions within the manuscript itself have been highlighted in red to facilitate review.

We have made every effort to ensure that our revisions align with the high standards of the journal and adequately address the concerns raised. We sincerely hope that these changes meet with your approval and contribute positively to the manuscript's overall quality. Please do not hesitate to contact us if you have any further questions or require additional information about our work. We remain committed to cooperating fully with the editorial process and are available for any follow-up discussions.

Thank you once again for your time, valuable feedback, and the opportunity to improve our manuscript. We look forward to your positive response.

Sincerely yours,
Runzhi Huang, Zhaofan Xia, Shizhao Ji
16th June 2025

Reviewer #3

Comment 1: *Please incorporate the FISH data into the manuscript. Most of the questions have been addressed.*

Response: We greatly appreciate your recognition of our research and the valuable comments provided. As recommended, we have incorporated RNA FISH results into the revised manuscript to demonstrate that TPSB2 and PECAM1, beyond serving as markers for mast cells and endothelial cells, respectively, also exhibit expression in fibroblasts.

Modifications are shown as follows:

In Results section, Page 10, line 284-289:

Considering that *PECAM1* is primarily a marker gene for endothelial cells and *TPSB2* serves as a marker gene for mast cells, we performed RNA fluorescence in situ hybridization (RNA-FISH) staining (*PECAM1+COL1A1+VIM; TPSB2+COL1A1+VIM*) to further validate that keloid fibroblasts express these two genes. Staining results revealed co-stained cells in keloids, demonstrating that fibroblasts express *PECAM1* and *TPSB2* transcripts (Figure 6C).

In Methods section, Page 23, line 665-682:

RNA-FISH

For RNA-FISH on tissue sections, start with antigen retrieval using an EDTA buffer (pH 9.0), followed by three PBS washes. Block endogenous peroxidase with 3% hydrogen peroxide at room temperature in the dark for 25 minutes. Rehydrate sections in DEPC water and DEPC-PBS, then digest with 20 μg/mL protease K at 37°C for 20 minutes. After prehybridizing with prehybridization solution for 2h at 37°C, dilute probes in prehybridization buffer, denature at 85°C, chill on ice, and incubate at 42°C. Apply the probe mixture for overnight hybridization at 37°C in a humidified chamber. Perform sequential washes with SSC buffers, incubate with a biotin-conjugated secondary antibody for 1 hour, and conduct TSA signal amplification. Finally, counterstain nuclei with DAPI and incubate for 10 minutes in the dark at room temperature. Images were captured using a fluorescence microscope. The probes used in the experiment were designed and synthesized by Genomeditech (Shanghai, China).

The probe sequences were as follows (from 5' to 3'):

VIM: TTTGAGTGGGTATCAACCAGAGGGAGT;

COL1A1: TTTCCACACGTCTCGGTCATGGTACCT;

PECAM1: AGGAAGATTCCAGTTCGGGCTTGGAAA;

TPSB2: TTTTGAACAGCAGGGGCTGGCTCTCC.

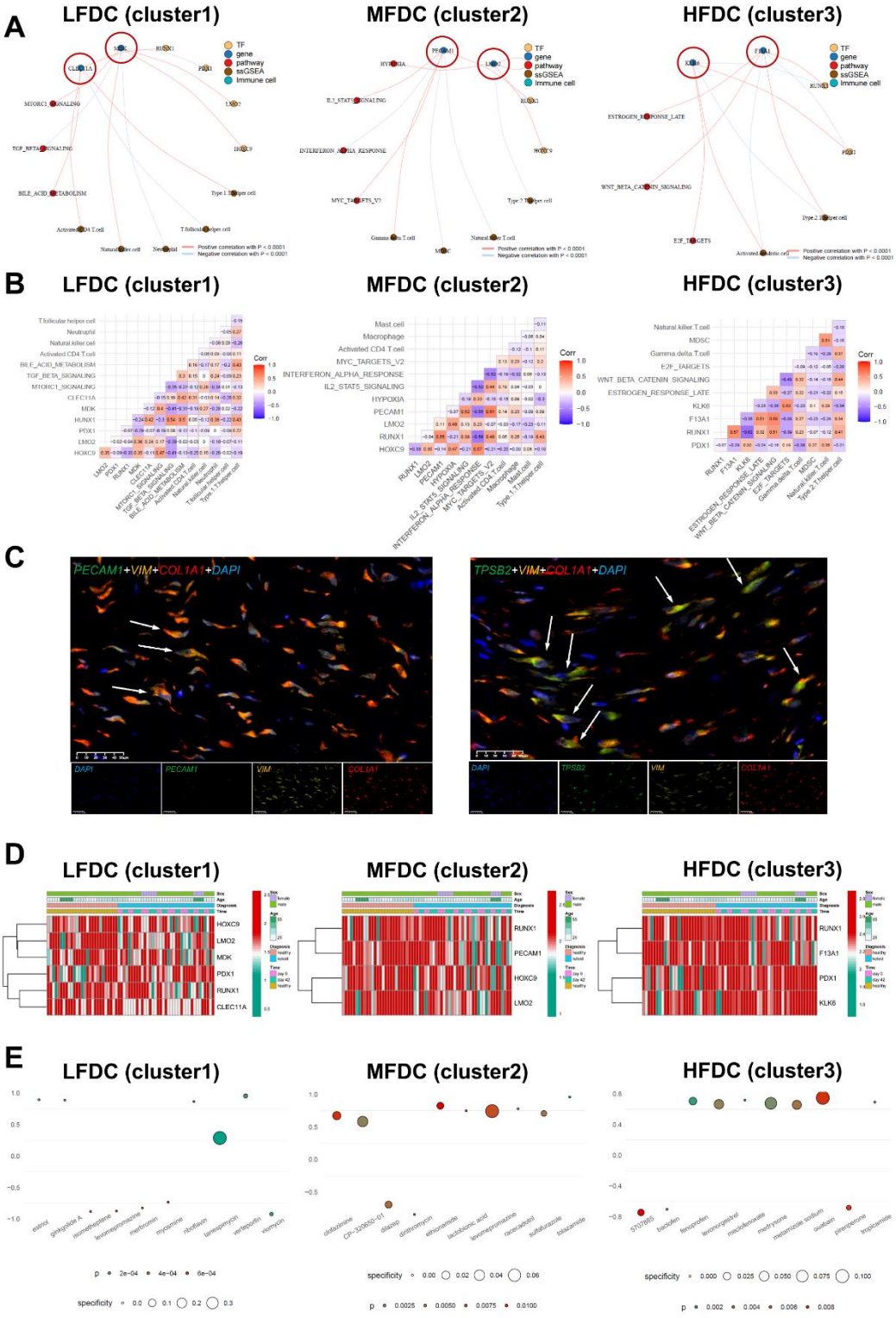


Figure 6. Regulatory network construction for keloid molecular subtypes and candidate drug identification. (A) Subtype-specific regulatory networks centered on DEFDRGs, integrating upstream DETFs, downstream signaling pathways, and immune cell/function regulatory potential across LFDC, MFDC, and HFDC. (B) Co-expression patterns and correlation heatmaps of network components (DEFDRGs, DETFs, pathways) in LFDC, MFDC, and HFDC. (C) RNA-FISH results of *COL1A1*, *VIM*,

TPSB2/PECAM1, and DAPI in human keloid samples. The white arrows indicate co-stained cells. Scale bars = 50 μm. (D) Expression profiles of core DEFDRGs and DETFs in LFDC, MFDC, and HFDC. (E) Candidate inhibitors predicted by Connectivity Map analysis in LFDC, MFDC, and HFDC. (LFDC: Low Fibroblast Differentiation Classification; MFDC: Moderate Fibroblast Differentiation Classification; HFDC: High Fibroblast Differentiation Classification; PCA: principal component analysis; DEFDRGs: Differentially Expressed Fibroblast-Differentiation-Related Genes; DETFs: Differentially Expressed Transcription Factors; RNA-FISH: RNA Fluorescence in Situ Hybridization)